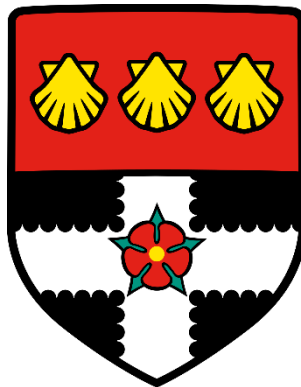# UNIVERSITY OF READING

Department of Geography & Environmental Science

Extending the predictability of flood hazard at the global scale

Rebecca Elizabeth Emerton

A thesis submitted for the degree of Doctor of Philosophy

November 2018

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Rebecca E. Emerton

Abstract

Flooding has the highest frequency of occurrence of all types of disaster arising from natural hazards across the globe. The anticipation and forecasting of floods is a key component in managing, preparing for and mitigating the effects of severe events, from local to international scales. This research aims to explore ways to extend the predictability of flood hazard at the global scale and provide earlier indications of potential flood events.

Two approaches for predicting river flow extremes on seasonal timescales are developed and tested; statistical forecasts based on the known influence of El Niño and La Niña on river flow and flooding at the global scale, and dynamical forecasts using numerical weather prediction systems. The statistical forecast development has shown that the likelihood of increased or decreased flood hazard during El Niño and La Niña events is much more complex than is often perceived and reported. The dynamical forecasts are shown to be more skilful than a long-term average climatology in many rivers worldwide, up to four months in advance in some cases. These approaches both have the potential to provide early warning information, and to support El Niño preparedness activities. As such, a comparison of the ability of the two forecasts to predict hydrological extremes during El Niño is undertaken, highlighting regions of the globe where each forecast is (or is not) skilful compared to a forecast of climatology, and the advantages and disadvantages of each approach.

Both of these new seasonal hydro-meteorological forecasts are openly available, with the dynamical forecasts produced operationally as part of the Global Flood Awareness System (GloFAS-Seasonal), supported by the Copernicus Emergency Management Service. This research has provided a step change in moving from forecasts that were previously only available for precipitation, to global-scale forecasts of hydrological variables at extended lead-times.

Acknowledgements

Contents

Chapter 1

Introduction

1.1 Motivation and Aims

Flooding has the highest frequency of occurrence of all types of disaster arising from natural hazards across the globe, accounting for 39% of all "natural disasters" since the year 2000 (Guha-Sapir et al., 2018). Floods can be caused by a variety of natural processes, and affect millions of people every year through displacement from homes, unsafe drinking water (sometimes leading to disease), destruction of infrastructure, and injury and loss of life. In 2017 alone, >57 million people were affected worldwide by the >120 disasters resulting from floods (Guha-Sapir et al., 2018). With an increasing global population and increasing populations living in flood-prone areas, the anticipation and forecasting of floods is a key component in managing, preparing for and mitigating the effects of severe events, from local to international scales.

Global overviews of upcoming flood events provide valuable information for organisations working at the global scale, across a range of water-related sectors from agriculture to humanitarian aid. Producing forecasts at the global scale has only become possible in recent years, due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations and land-surface hydrology modelling, and increased resources and computer power (Alfieri et al., 2012, 2013; Bierkens, 2015; Brown et al., 2012; ECMWF, 2018a). While several forecasting centres produce operational forecasts[1] of floods in the medium-range, that is, up to ~2 weeks ahead, earlier indications of potential flood events, many weeks or even months in advance, could provide crucial information for flood preparedness and disaster risk reduction. Indeed, the World Meteorological Organization (WMO, 2017) states that economic losses due to severe hydrometeorological events have increased, over the past fifty years, nearly 50 times, but that loss of life has decreased by a factor of 10^3. This significant decrease in loss of life is attributed to improved monitoring and forecasting of floods alongside more effective preparation and planning.

The aim of this research is to explore ways to extend the predictability of flood hazard at the global scale and provide earlier indications of potential flood events. Predictability is defined as

---

[1] The term operational here refers to real-time forecasts produced by a forecasting centre, that are a 24/7 supported service ensuring timely dissemination and ongoing provision of forecasts.

"the extent to which future states of a system may be predicted based on knowledge of current and past states of the system"; in other words, predictability is the degree to which a prediction can be made, qualitatively or quantitatively, or the longest achievable lead time for a forecast. (AMS, 2012). Broadly speaking, there are two key ways in which the inherent predictability of the atmosphere and land surface can be used to provide early warning information:

- Statistical analysis based on large-scale climate variability and teleconnections

- Seasonal forecasting using coupled ocean-atmosphere general circulation models[2]

While both of these have been studied and/or implemented for meteorological variables, forecasts of hydrological variables are often not considered or provided, particularly for large or global scales. For example, information on the likelihood of extreme precipitation driven by large-scale modes of climate variability is readily available, and is often used as a proxy for flooding due to the absence of the equivalent information for river flow. Additionally, while seasonal forecasts of meteorological variables including precipitation are produced at many operational forecasting centres, no such forecasts are available for river flow at the global scale. However, recent research has shown that the link between precipitation and flood magnitude is nonlinear (Stephens et al., 2015), and as such, precipitation may not be the best indicator of potential flood hazard (Coughlan de Perez et al., 2017). This thesis aims to combine both meteorological and hydrological aspects of flood predictability and forecasting, in order to explore both of the aforementioned avenues for extending flood predictability. This is done through the following specific objectives:

1. Analyse the link between El Niño Southern Oscillation (ENSO), the most dominant mode of large-scale climate variability, and river flow across the globe, using historical events to answer the question "what is the likelihood of flooding during El Niño?".

2. Develop and test seasonal forecasts of flood hazard for the global river network, by driving the hydrological component of the Global Flood Awareness System (GloFAS) with seasonal meteorological forecasts from the European Centre for Medium-Range Weather Forecasts' (ECMWF) coupled ocean-atmosphere general circulation model.

---

[2] It is noted that coupled ocean-atmosphere GCMs here refer to the dynamical models used to produce seasonal forecasts. In hydrology, this is often also referred to as numerical weather prediction (NWP), however NWP traditionally refers to atmosphere-only models used to produce short-range (up to 5 days ahead) weather forecasts that are not suitable for forecasting on seasonal timescales.

---

3. Assess the potential usefulness of both the statistical (1) and dynamical (2) approaches to extending flood predictability and providing early indications of flood hazard at the global scale, for decision-making purposes.

The results presented in this thesis will provide a hydrologically relevant, global scale analysis of flood hazard predictability, alongside providing the equivalent information for hydrological variables that exists for meteorology. This information has the potential to be used to inform decision-making across a range of water-related sectors, and aid flood preparedness and disaster risk reduction efforts.

1.2 Structure of the Thesis

This thesis is structured around four papers. To begin with, Chapter 2, the first paper presented in this thesis, provides a detailed overview of the current state of large-scale flood forecasting. Six operational large-scale flood forecasting systems are reviewed, and the challenges and future advances in global scale flood forecasting are discussed, including the possibility of extended-range forecasting out to seasonal timescales. Chapter 3 provides further background material relating to ENSO and its influence on weather and climate, including river flow, to support the introductory literature presented in Chapter 4.

Chapter 4 is the second of the papers presented in this thesis, addressing the first objective by using a new 110-year model reconstruction of river flow to evaluate the link between ENSO and river flow, and map the likelihood of increased or decreased flood hazard during El Niño and La Niña events.

The second objective of this thesis is addressed in Chapter 5, which presents the third paper. Chapter 5 introduces GloFAS-Seasonal, the first global scale seasonal hydro-meteorological forecasting system designed to provide early indications of high and low river flow for the global river network. GloFAS-Seasonal was developed and implemented operationally as part of this research. The paper provides technical detail regarding the development of the system, information on the forecast products available, and an initial evaluation of the skill of the forecasting system.

Chapter 6 works towards the third objective of this thesis, evaluating the potential usefulness of both the statistically-based historical probabilities of ENSO-driven flood hazard presented in Chapter 4, and the resource-intensive GloFAS-Seasonal forecasting system presented in

Chapter 5. This fourth paper uses river flow observations to assess and compare the ability of the two forecasts to predict high and low river flow during El Niño.

Chapter 7 summarises the findings and wider contribution of this thesis and outlines scope for further work.

The four papers presented in this thesis have been reformatted as chapters, and have not been modified. The published versions of Chapters 2, 4 and 5 are provided in the Appendix, alongside further co-authored publications related to this work. Chapter 6 was still in press at the time of completing this thesis. Author contribution statements are provided at the beginning of each relevant chapter.

Chapter 2

The Current State of Large Scale

Flood Forecasting

This chapter has been published as a review paper in Wiley Interdisciplinary Reviews (WIREs) Water with the following reference:

The contributions of the authors of this paper are as follows: R.E.E. conducted the literature review and wrote the paper with guidance from H.L.C., E.M.S. and F.P., with the exception of parts of section 2.3.4, written by A.W.W., and section 2.6.3, written by C.A.B. Section 2.5 was written in collaboration with F.P. Further information beyond that which was documented in the literature was provided by P.S. (EFAS & GloFAS), T.C.P. (BoM FFWS), A.H.W. (GLOFFIS), A.W.W. and J.D.B. (U.S. HEFS), N.H. and C.D. (E-HYPE). All authors commented on the manuscript. Overall, 80% of the writing was undertaken by R.E.E.

Abstract. Floods are the most frequent of natural disasters, affecting millions of people across the globe every year. The anticipation and forecasting of floods at the global scale is crucial to preparing for severe events and providing early awareness where local flood models and warning services may not exist. As numerical weather prediction models continue to improve, operational centres are increasingly using the meteorological output from these to drive hydrological models, creating hydro-meteorological systems capable of forecasting river flow and flood events at much longer lead times than has previously been possible. Furthermore, developments in, for example, modelling capabilities, data and resources in recent years have made it possible to produce global scale flood forecasting systems. In this paper, the current state of operational large scale flood forecasting is discussed, including probabilistic forecasting of floods using ensemble prediction systems. Six state-of-the-art operational large scale flood

---

---

forecasting systems are reviewed, describing similarities and differences in their approaches to forecasting floods at the global and continental scale. Currently, operational systems have the capability to produce coarse-scale discharge forecasts in the medium-range and disseminate forecasts and, in some cases, early warning products, in real time across the globe, in support of national forecasting capabilities. With improvements in seasonal weather forecasting, future advances may include more seamless hydrological forecasting at the global scale, alongside a move towards multi-model forecasts and grand ensemble techniques, responding to the requirement of developing multi-hazard early warning systems for disaster risk reduction.

2.1 Introduction

Flooding has the highest frequency of occurrence of all types of natural disaster across the globe, accounting for 39% of all natural disasters since 2000, with >94 million people affected by floods each year worldwide (Guha-Sapir et al., 2018) through displacement from homes, unsafe drinking water, destruction of infrastructure, injury and loss of life. With an increasing population living in flood-prone areas, the forecasting of floods is key to managing and preparing for imminent disaster.

Investment in building resilience is prioritised in the Sendai Framework for Disaster Risk Reduction (DRR) 2015-2030 (UNISDR, 2015), with one component of this being the development and use of multi-hazard early warning systems (WMO, 2017). The World Meteorological Organization (WMO) states that economic losses due to severe hydro-meteorological events have increased, over the past fifty years, nearly 50 times. However, the loss of life globally has decreased by a factor of 10^3. This significant decrease in loss of life is attributed to improved monitoring and forecasting of hydro-meteorological events alongside more effective preparation and planning. Four components are suggested by the WMO (WMO, 2017) for effective early warning systems; detection, monitoring and forecasting hazards, analyses of risks involved, dissemination of timely warnings and activation of emergency plans to prepare and respond.

The development of forecasting systems producing forecasts and warnings of severe hazards such as floods, droughts, storms, fires and tropical cyclones on a global scale are critical for disaster risk reduction and further decreases in loss of life. The Sendai Framework for Disaster Risk Reduction 2015-2030 (UNISDR, 2015) states that at global and regional levels it is important to "promote co-operation between academic, scientific and research entities and networks and the private sector to develop new products and services to help reduce

disaster risk, in particular those that would assist developing countries and their specific challenges", and forecasting systems such as those discussed here are essential in achieving this, particularly in providing forecasts for countries and regions where no other forecasts and early warnings are available.

The need for large scale flood forecasting systems can be broken down into three key factors:

i) To provide information on floodiness (Stephens et al., 2015) across areas larger than a catchment, for example to indicate where flooding during the rainy season will be worse than normal; information that is of high importance to humanitarian organisations (Braman et al., 2013).

ii) To provide forecasts in basins across the globe where currently there are no forecasts available, which is not a massive scale-up of resources. Large-scale forecasting is therefore cost-effective compared to focussing on developing and providing hydro-meteorological forecasts for single catchments, and greatly aids disaster risk reduction and flood early warning efforts globally.

iii) To support existing capabilities, for example by using ensemble forecasting techniques to enable probabilistic flood forecasts, or at longer lead-times for earlier warnings. Probabilistic and extended-range forecasting is computationally expensive, and in addition, many countries do not currently pay for access to these distributed meteorological forecast products and therefore are unable to produce any form of hydro-meteorological forecast.

This review outlines the developments which have led to forecasting floods on the global scale, the current state-of-the-art in operational large-scale (continental and global) flood forecasting, and future developments in global scale flood forecasting and early warning.

2.2 Advances in the Science and Techniques of Global Forecasting

Producing forecasts at the global scale has only become possible in recent years, due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations and land-surface hydrology modelling, and increased resources and computer power (Alfieri et al., 2012, 2013; Bierkens, 2015; Brown et al., 2012; ECMWF, 2018a). While several meteorological and hydrological forecasting centres now run operational flood forecasting models, many of these are for specific locations, river basins or countries (Alfieri et

al., 2012).

Table 1: Technical details of quantitative precipitation forecasts used in large-scale flood forecasting (adapted from Alfieri et al., 2012).

Product Type	Spatial Extent	Spatial Resolution	Temporal Resolution	Forecast Range	Ocean-Atmosphere Coupling	Uncertainty
Radar Nowcasting	~10,000 - 50,000km²	1-4km	5-60min	1-6h	No	Low
Ensemble Radar Nowcasting	~10,000 - 50,000km²	1-4km	5-60min	1-6h	No	
Radar-NWP Blending	Regional	~2km	15-60min	~6h	No	
Limited-Area NWP	Regional - Continental	2-25km	1-6h	1-3 days	Varies	
Ensemble Limited-Area NWP	Regional - Continental	2-25km	3-6h	~5-30 days	Varies	
Global NWP	Global	~15-100km	~3-6h	~5-30 days	Varies	
Sub-seasonal to Seasonal Forecasts	Continental - Global	~25-100km	~3-24h	~15-60 days	Yes	
Seasonal Forecasts	Global	~15-100km	~6-24h	Months	Yes	High

Global hydrological modelling is complex due to the geographical variation of rainfall-runoff processes and river regimes (Pappenberger et al., 2010), but large scale flood forecasting systems are now emerging with recent scientific and technological advances and increasing integration of hydrological and meteorological communities, allowing for uncertainty to be cascaded from the meteorological input to the river flow forecasts (Ramos et al., 2010).

In this section we analyse the key advances that have enabled the forecasting of floods at the global scale.

2.2.1 The Increasing Skill of Precipitation Forecasts

The skill of precipitation forecasts in global NWP models has increased significantly in recent years (Liu et al., 2013; Mittermaier et al., 2013; Novak et al., 2014) (e.g., gaining ~2days precipitation skill since 2000 (Richardson et al., 2012)). With skilful medium-range quantitative precipitation forecasts (QPFs) being produced by NWP models across the globe, it has become possible to produce skilful forecasts of river flow and flooding at large scales for the purpose of early warning (Bartholmes and Todini, 2005). While there exist many different definitions of a skilful forecast, this typically refers to correlation of the forecast with observations out to a certain lead time. Table 1 outlines the resolutions and forecast ranges of some of the main QPF products used in operational large-scale flood forecasting systems (Alfieri et al., 2012). Precipitation is challenging to forecast due to the chaotic nature of the atmosphere (Lorenz, 1969); whereby a small change in the initial conditions of the system can result in an unpredictable outcome. The underlying physical processes of precipitation generation are complex to model, and modelling deficiencies can lead to forecast inaccuracies, particularly at longer lead times (Cuo et al., 2011). In general, due to lack of observations, precipitation predictions are less skilful in the southern hemisphere, although the difference in the skill of forecasts between the hemispheres has reduced significantly since the introduction of satellite observations and data assimilation (Cuo et al., 2011; Simmons and Hollingsworth, 2002). Limited data are also an issue in much of the tropics, alongside difficulties associated with the simulation of convective precipitation (Krishnamurti et al., 1999). While QPF skill depends heavily on the region, season, intensity and storm type (Cuo et al., 2011), precipitation skill is generally good for rainfall generated by synoptic scale frontal weather systems (Olson et al., 1995). The intensity of precipitation tends to be one of the major problems in QPFs, with convective (Krishnamurti et al., 1999) and orographic enhancement (Arduino et al., 2005) processes tending to result in an under-prediction of intensity, alongside the tendency of most global models to over-predict the intensity of light precipitation (Haiden et al., 2014). Many NWP models struggle with displacement (Cuo et al., 2011; Ebert and McBride, 2000); while the areal extent, timing and intensity of precipitation may be correct, precipitation displacement can be extremely detrimental to forecasts of river flow and flooding.

With ongoing improvements to NWP models (resolution increases, new methods of simulating the physical processes and increasing computer power), (Mittermaier et al., 2013; Novak et al.,

2014; Richardson et al., 2012; Tang et al., 2013), alongside developments in model interoperability services, cloud services, and open data and models, precipitation forecasts have become more useful to hydrological applications.

2.2.2 Ensemble Flood Forecasting – Representing Uncertainty

Over the past two decades, NWP has moved from single-solution forecasts of the future state of the atmosphere, to probabilistic forecasts using ensemble prediction systems (EPS; Cloke and Pappenberger, 2009). Probabilistic forecasts allow the inherent uncertainties in NWP to be represented (Demeritt et al., 2007; Liu et al., 2013). In hydrological modelling, the four main sources of uncertainty are input data, evaluation data, model structure and model parameters (Kauffeldt, 2014; Pagano et al., 2013; Shaw et al., 2011; Wood and Lettenmaier, 2008). The relative importance of these uncertainties tends to vary according to catchment characteristics, event magnitude and lead time of the forecast (Cloke and Pappenberger, 2009; Ramos et al., 2010), but it is generally accepted that the greatest uncertainty in flood forecasting beyond 2-3 days lead time stems from the meteorological input (Cloke and Pappenberger, 2009; Kauffeldt, 2014).

The standard approach in NWP is to produce a single (deterministic) forecast from the initial state, whereas EPS recognise and represent the uncertainty in the initial conditions by perturbing them to produce several initial states (Buizza et al., 2005; Leutbecher and Palmer, 2008). The forecast model is run from each of the perturbed initial states, producing many varying, but valid and equally probable, forecast scenarios. In addition to sampling the error in the initial state, many centres also incorporate stochastic physics, which involves applying random perturbations of the parameterised physical processes (Buizza et al., 2007).

Predictions of river discharge are usually produced by providing the EPS as input to a hydrological model (Clark and Hay, 2004; Cloke et al., 2013b; Cloke and Pappenberger, 2009; Pagano et al., 2013). Prior to this, some pre-processing may be required (Cloke et al., 2013b; Pagano et al., 2013); scale corrections (downscaling or disaggregating) are made, as due to the irregular shape of catchments, the scale (temporal and spatial) does not usually correspond between the EPS and the hydrological model (Liu et al., 2013). Bias or spread corrections may also need to be made (Cloke and Pappenberger, 2009).

The use of EPS in flood forecasting allows probabilistic forecasts of flood events at much longer lead times than has previously been possible, and is useful in producing forecasts in catchments

Table 2: Operational large-scale flood forecasting systems.

Forecasting System	EFAS (European Flood Awareness System)	E-HYPE (European Hydrological Predictions for the Environment)	FFWS (Flood Forecasting & Warning Service)	HEFS (Hydrologic Ensemble Forecast Service)	GloFAS (Global Flood Awareness System)	GLOFFIS (Global Flood Forecasting Information System)
Domain	Continental (Europe)	Continental (Europe)	Continental (Australia)	Continental (USA)	Global	Global
No. Ensemble Members	65	1	≤4	23 Short to Medium Range, 1 Long Range	51	73
Forecast Range (Days)	15	10	10	Sub-Hourly to Several Years	45	15
Spatial Resolution	5km, Regular Grid	~15km, Irregular Grid, Varies by Basin	~10km	Varies by Basin	10km, Regular Grid	10km, 50km, Regular Grid
Forecast Frequency	12-Hourly	Daily	6-Hourly to 12-Hourly	Sub-Daily to Daily	Daily	6-Hourly
NWP Input	ECMWF ENS, ECMWF Deterministic, DWD Deterministic, COSMO-LEPS	ECMWF Deterministic	BoM ACCESS Global, Regional, City-Scale and Relocatable Deterministic Forecasts	RFC Deterministic, WPC Deterministic, GEFS, CFS, Historical Observations	ECMWF ENS	ECMWF ENS, GEFS, GFS, Historical Forcing
Rainfall-Runoff Model	Lisflood Europe	HYPE	GR4J (Daily), GR4H (Hourly), URBS	Suite of Models (see Figure 8)	HTESSEL	PCR-GLOBWB, W3RA
Routing Model	Lisflood Europe	HYPE	Muskingum Channel Routing	Suite of Models (see Figure 8)	Lisflood Global	Deltares wflow
River Network	JRC Dataset	HydroSHEDS, HYDRO1K	Catchment-SIM	Suite of Models (see Figure 8)	HydroSHEDS, HYDRO1K	PCR-GLOBWB, SRTM90m, HydroSHEDS
Organisation	JRC, ECMWF	SMHI	BoM	National Weather Service	JRC, ECMWF	Deltares
Website	www.efas.eu	e-hypeweb. smhi.se	www.bom.gov.au/ water/floods	water.weather.gov/ahps/forecasts.php	www.globalfloods.eu	
Corresponding Figure Number	2	5	6	8	10	12

Chapter 2. The current state of large scale flood forecasting

where no other input data is available. Cloke and Pappenberger (2009) give a detailed review of the benefits of ensemble over deterministic flood forecasts, particularly looking at advantages for issuing flood alerts and warnings. Probabilistic forecasts of upcoming events have been shown to provide greater skill than deterministic forecasts (Stephens and Cloke, 2014), and provide key information about the possibility of occurrence of an extreme event.

2.2.3 Operational Large Scale Flood Forecasting

There exist various large-scale hydrological models run by communities around the globe; Bierkens et al. (2015) give a detailed overview of the properties of 14 global scale and 4 continental scale models. Not all of these models are used operationally for the purpose of flood forecasting, and as such, a list of operational continental and global scale flood forecasting models, alongside key system information, is provided in table 2.

Figure 1 shows a simplified conceptual model for a large-scale flood forecasting system: the components required and the output generated within each component. The operational systems outlined in table 2 are the focus of this review, and each takes a different approach to the components of the conceptual model. In the following sections we benchmark the state of current science and technology in undertaking operational continental and global scale flood forecasting and early warning.

Figure 1: A conceptual large-scale hydro-meteorological flood forecasting system.

2.3 Continental Scale Flood Forecasting Systems

There are currently four operational continental scale flood forecasting systems, two for Europe; the European Flood Awareness System (EFAS) of the European Commission (EC), and the European HYdrological Predictions for the Environment (E-HYPE) model of

the Swedish Meteorological and Hydrological Institute (SMHI). The Bureau of Meteorology (BoM) run the Flood Forecasting and Warning Service (FFWS) for Australia, and the U.S. National Weather Service (NWS) run a model covering the Continental USA; the Hydrologic Ensemble Forecasting Service (HEFS). This section outlines the components of, and the forecast products produced by, each system.

2.3.1 The European Flood Awareness System

EFAS is an EC initiative developed by the Joint Research Centre (JRC) to increase preparedness for riverine floods across Europe. It was in development from 2002, tested from 2005-2010, and has been operational since 2012. After devastating, widespread flooding on the Elbe and Danube rivers in 2002, the EC began development of EFAS, with the aim of providing transnational, harmonised early warnings of flood events and hydrological information to national agencies, complementing local services (Thielen et al., 2009). Various consortia execute different aspects (e.g. computation and dissemination) of the EFAS operational suite.
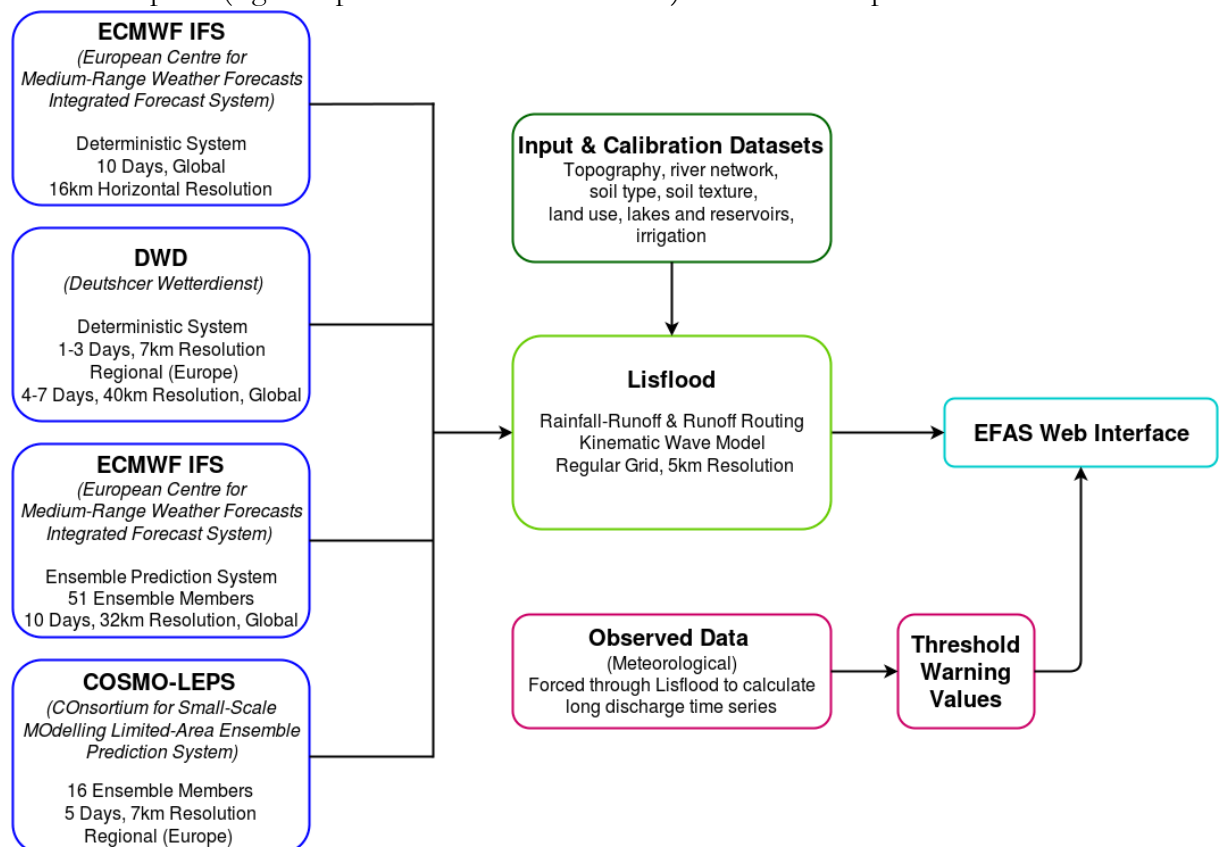
Figure 2: Components of the European Flood Awareness System (EFAS).

Model Components

Rather than using just one meteorological NWP forecast as input, EFAS uses four different forecasts; two ensemble forecasts and two deterministic. Figure 2 details the various

components of the EFAS suite, including key information regarding the NWP models. The precipitation, temperature and evaporation from each of the four forecasts are used as input to the Lisflood hydrological model, which is used as both the rainfall-runoff and the routing components shown in figure 1, and simulates canopy, surface and sub-surface processes such as snowmelt (including accounting for accelerated snowmelt during rainfall) and preferential (macropore) flow, soil and groundwater processes (Thielen et al., 2009).

Simulated ensemble hydrographs are produced by Lisflood, however these alone do not constitute a flood forecast; a decision-making element needs to be incorporated (Thielen et al., 2009). Due to the often limited number of discharge observations in many areas of the globe, these critical thresholds cannot be derived directly from observations. As such, meteorological data are run through Lisflood to calculate a 22-year timeseries of discharge, to provide a reference threshold for minor or major flooding at each grid cell.

Figure 3: The European Flood Awareness System (EFAS) showing (a) the main interface with high (red) and medium (yellow) reporting points, flood alerts (warning triangles) and probability (% likelihood) of exceeding 50mm of precipitation (green shading) during the forecast period (10 days), (inset a) the flood alert displayed when the alert point is clicked on, (b) the return period hydrograph with return period thresholds (1.5 years green, 2 years yellow, 5 years red, 20 years purple), (c) upstream snow melt forecast, (d) upstream precipitation forecast.

Forecast Visualisation

Alongside warnings for each forecast point, the EFAS interface (e.g. figure 3) provides ensemble hydrographs, which allow interpretation of the spread of the ensemble and the uncertainty in the forecast. Persistence diagrams showing information about the previous

four forecasts also give the user additional information on the forecast uncertainty, as NWP models should be able to pick up large scale synoptic weather systems which typically produce severe events, in advance, therefore showing a flood risk consistently in each forecast run (Thielen et al., 2009). The EFAS interface provides a map of Europe, with all points forecasting a flood event designated by a colour responding to the warning threshold; this allows an overview of forecast flood events across the continent. The information and visualisation within EFAS are designed to give clear, concise and unambiguous early warning results.

Warning Dissemination

Copernicus is the European Emergency Management Service, and EFAS is the operational flood early warning system designed to disseminate warnings for Europe under the Copernicus initiative. According to the World Meteorological Organization Executive Council (EC-LVII-Annex VII; WMO, 2005), National Meteorological and Hydrological Services (NMHS) constitute the single authoritative voice on weather warnings in their respective countries. Therefore, in order to respect the single voice principle also with regard to floods, EFAS real-time information is provided only to hydro-meteorological authorities signing a "Condition of Access" document.

Box 1: Example of an operational EFAS flood alert, sent to EFAS partners and national and regional services on 25th June 2015 via the EFAS dissemination centre (the Swedish Meteorological and Hydrological Institute, SMHI).

EFAS FLOOD ALERT REPORT

Dear Partner,

EFAS predicts a high probability of flooding for Norway - Otta and Lagen-Mjosa tributaries (Glomma basin) from Monday 29th June onwards.

According to the latest forecasts (2015-06-25 12 UTC) up to 100% EPS (VAREPS) are exceeding the high threshold (>5 year simulated return period) and up to 86% EPS (VAREPS) exceeding the severe threshold (>20 year simulated return period).

Compared to the VAREPS mean, the ECMWF deterministic forecast is comparable and the DWD deterministic forecast is lower.

The earliest flood peak is expected for Saturday 4th of July 2015.

Please monitor the event on the EFAS-IS interface (http://www.efas.eu)

Chapter 2. The current state of large scale flood forecasting

EFAS sends warning emails to these national authorities responsible for flood forecasting, designed to bring awareness of an upcoming flood event, with further details accessed through the interface. There are four types of warning emails provided; *Flood Alerts* are issued when a river basin has a probability of exceeding critical flood thresholds more than 2 days ahead, *Flood Watches* are issued when there is a probability of a river basin exceeding critical thresholds but the event does not satisfy the conditions for a *Flood Alert* (such as river basin size or warning lead time), and *Flash Flood Watches* are issued when there is a probability >60% of exceeding the flash flood high alert threshold. An example of an EFAS *Flood Alert* is given in Box 1. The 2 day lead time criteria is specified as the forecasting systems used by the national authorities have usually issued a national warning with a lead time of up to 2 days. Additionally, daily overviews are sent to the Emergency Response Coordination Centre (ERCC) of the EC, containing information on ongoing floods in Europe, as reported by the national services and EFAS warnings.

Forecast Verification

EFAS also undergoes forecast verification, with two methods used for this system. Firstly, the hits, false alarms and misses are assessed for each flood event, with events evaluated through feedback reports and news media. Secondly, skill scores are calculated and reported regularly through EFAS bulletins, available via the website (see table 2).

Operational Applications

EFAS is integrated in the daily forecasting procedures of many national hydrological services across Europe, providing operational early warnings and additional information which is used for decision making purposes at national and local scales. Additionally, EFAS is used by the ERCC to compile reports on the flood situation and outlook, and for the co-ordination of emergency response, at the continental scale.

2.3.2 The European HYdrological Predictions for the Environment Model

E-HYPE is a multipurpose model based on open data (table 3), which is used for various applications such as water management, research experiments and flood forecasting (Donnelly et al., 2016; SMHI, 2015). The E-HYPE Water in Europe Today (WET) tool (figure 4), compares the current hydrological situation with climatological data and past modelled events. The tool was originally designed to alert water managers to flow that is predicted to be outside of the normal range (based on the 75th and 25th percentiles), and

has evolved to provide information to many end users. Another setup of the HYPE model, EFAS-HYPE, uses further, restricted, datasets and is currently being tested as an additional model within EFAS. This section focusses on the river flow forecasts produced by the WET tool.

Table 3: Databases used within the flood forecasting systems. Due to the alternative set-up of the BoM FFWS (including event-based modelling, nowcasting and significant forecaster input; see section 2.3.3), this information was not available.

Data Type	Data Source					
					GLOFFIS	
	EFAS	E-HYPE	HEFS	GloFAS	PCRGLOB-WB	W3RA
Topography/ Routing	SRTM/CCM2	HydroSHEDS & HYDRO1K	NED & NHDPlus	HydroSHEDS & HYDRO1K	HydroSHEDS, HYDRO1K & NASA SRTM	HydroSHEDS, HYDRO1K & NASA SRTM
Land Cover	CORINE	CORINE and Globcover 2000	NLCD, MODIS, AVHRR	CORINE and Globcover 2000	GLCC, MIRCA	MODIS
Urban Areas	European Soil Data Centre (ESDAC)	Euroland SoilSealing 2009	n/a	Harmonized World Soil Database	GLCC	n/a
Lake Area & Spatial Distribution	GLWD (Global Lake and Wetland Database)	GLWD (Global Lake and Wetland Database)	NHDPlus	GLWD (Global Lake and Wetland Database)	GLWD, GRaND (Global Reservoir and Dams Database)	n/a
Lakes and Reservoirs	GLWD, GRaND (Global Reservoir and Dams Database)	GLWD, ERMOBST, FLAKE-Global, International Water Power & Dam, ILEC World Lake Database, LEGOS, SMHI	USGS & Federal state and local water management authorities (e.g. USACE, Reclamation)	GLWD, Global Reservoir and Dams Database GRAND	GLWD, FLAKE-Global, GRaND (Global Reservoir and Dams Database)	n/a
Soil Type	European Soil Data Centre (ESDAC)	Based on Land Use and Elevation	SSURGO	Harmonized World Soil Database	FAO DSW	n/a
Crop Types	n/a	CAPRI, MIRCA-2000	n/a	n/a	MIRCA	n/a
Irrigation	EIM (European Irrigation Map), GMIA (Global Map of Irrigation Areas)	EIM (European Irrigation Map), GMIA (Global Map of Irrigation Areas)	NHDPlus, Local water authorities	GMIA (Global map of Irrigation Areas)	MIRCA	n/a

Chapter 2. The current state of large scale flood forecasting

Model Components

In contrast to other systems, E-HYPE currently uses only deterministic NWP input to drive the hydrological model component, though ensemble forecasting is intended for future system developments. The HYPE model (Donnelly et al., 2016; Lindström et al., 2010) is a distributed rainfall-runoff model developed at SMHI, which divides catchments into subbasins rather than a regular grid. Each subbasin is further divided into classes based on land use, soil type and elevation (SMHI, 2015). Alongside processes such as snow accumulation and melting, evapotranspiration and groundwater recharge (Lindström et al., 2010), HYPE also takes into account anthropogenic influences including irrigation and hydropower (SMHI, 2015).

Forecast Visualisation

Within the WET tool, forecasts of river flow are compared to climatology, based on the ECMWF ERA-Interim reanalysis and evaluation datasets (figure 5) in order to produce an overview of river flow that is under or above the normal range. This information is displayed on a colour-coded map of the subbasins within the E-HYPE model (figure 4).
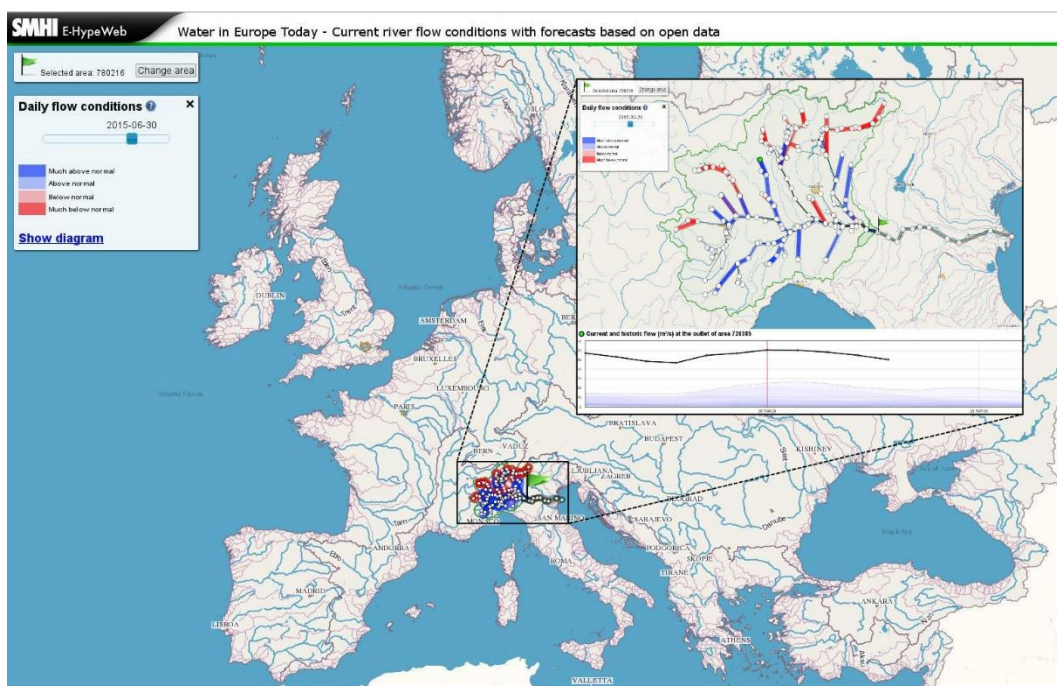
Figure 4: The Water in Europe Today (WET) tool interface with example forecast (inset) showing above-normal (blue shading) and below-normal (red shading) forecast river flow. The hydrograph shows current conditions and forecast river flow (black line) compared to climatology (blue shading). Forecasts are available at hypeweb.smhi.se/europehype/forecasts

Chapter 2. The current state of large scale flood forecasting
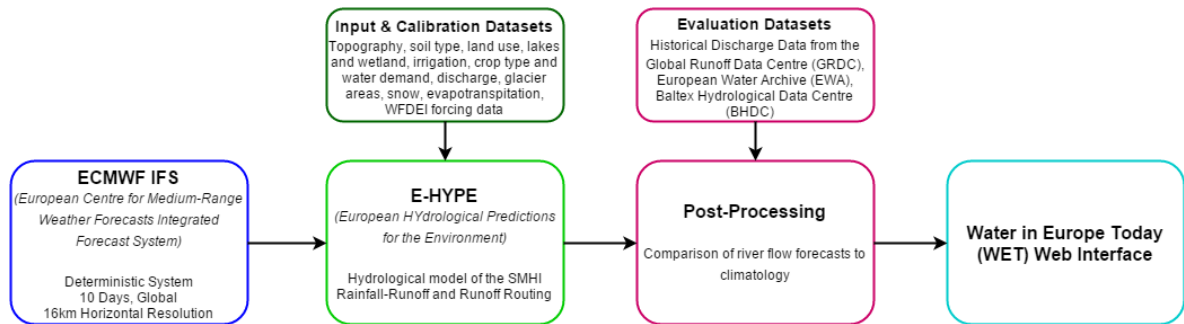
Figure 5: Components of the European Hydrological Predictions for the Environment (E-HYPE) Water in Europe Today (WET) tool.

Forecast Verification

Through the E-HYPE and WET interface, various model performance statistics are available. The model is verified against observed discharge from river gauges, and allows the user to quickly evaluate the performance of the model with regard to timing, variability and volume error for the point of interest or across a larger region. The overall model performance in terms of mean annual discharge is also presented. Donnelly et al. (2016) present a new method for evaluating the performance of a multi-basin model and results from this evaluation of the historical model indicated that the model is suitable for predictions in ungauged basins as it captures the spatial variability of flow. While the model performs well in terms of long-term means and seasonality, the performance is less effective in terms of daily variability, particularly in Mediterranean and mountainous areas, and in regions of most anthropogenic influence.

Operational Applications

E-HYPE is currently being used in several applications across Europe, such as seasonal flow forecasting for the EU EUPORIAS project which aims to help societies to deal with climate variability, and providing data for use in oceanography models and as part of the SWITCH-ON EU project. The WET tool is also used by various other smaller companies around Europe to provide water forecasts, for example soil-water forecasts for gardening companies.

2.3.3 The Australian Flood Forecasting and Warning Service

The Australian BoM has been producing flood forecasts operationally for several decades, with the technology and systems used to produce these forecasts continually evolving. More recently, the BoM has introduced short-term (up to 7 days ahead) continuous streamflow forecasting using deterministic NWP models, within the Hydrological Forecasting System

(HyFS) production environment (based on the Deltares Flood Early Warning System (FEWS) forecasting framework), alongside event-based hydrological modelling and nowcasting using radar rainfall estimates. The BoM services also rely on forecasters for the dissemination and communication of flood warnings and local information regarding river conditions.

Model Components

The NWP forecasts used to force the rainfall-runoff models are produced by the BoM's Australian Community Climate and Earth-System Simulator (ACCESS) NWP model. ACCESS has four components running at different spatial scales and resolutions (figure 6). In addition to the NWP model output, forecasters and hydrologists at the BoM can produce "What If" precipitation scenarios with which to force the hydrological models.

Alongside the semi-distributed GR (Génie Rural a 4 Parametres) hydrological models, event- based forecasting is used extensively; for this, local models are used in support of the continental scale system. The resulting river discharge estimations from both model versions are used, alongside observed data and statistical models, to produce automated graphical products such as maps, bulletins, warnings and alerts.

Role of the Forecaster

Whilst the other systems presented in this paper are almost entirely automated and model-based, the BoM system also relies on the input of expert meteorologists and hydrologists. In addition to producing "What If" scenarios to feed into the hydrological models, the forecasters are able to manually post-process the forecasts and observed data to produce further products and visualisations and assess the quality of the data and forecasts in real time. The forecasters are also able to produce additional warnings on the fly, for example if a reservoir is seen to fill, or their experience alerts them to an alternative possible scenario to those produced by the hydrological models. The hydrologists at the BoM are also responsible for dissemination and communication of the forecasts and warnings.

A further reason for the input of forecasters is due to the challenges of producing operational flood forecasts for a large continent with an unevenly distributed population. Metropolitan areas have a dense observation network for both rainfall and river discharge, however there are large areas of Australia that have no flowing rivers, such as in the Northern Territory where there is an average of one river gauge every 13,360km^2.

Figure 6: Components of the Australian Flood Forecasting and Warning Service (FFWS).

Warning Dissemination

The final products delivered to the end users include flood watches and warnings, and information on current river levels and precipitation, which are disseminated to various users at specified stages in the evolution of a flood event, through a dedicated web interface, email, fax and telephone. These are usually text forecasts, an example of which is given in box 2 for a minor flood event, written by the hydrologists based on the output of the HyFS, but can also include automated alerts and bulletins for certain users. Figure 7 shows the corresponding publicly available graphics for this flood event, while the BoM hydrologists also have access to more sophisticated graphical products produced by the automated component of the HyFS, such as ensemble hydrographs.

Figure 7: The BoM publicly available flood warnings showing (a) warnings and river conditions across Australia, (b) warnings and river conditions for a particular region, (c) current river levels at a specific warning point where flow is above the minor flood level.

Chapter 2. The current state of large scale flood forecasting

Box 2: Example of a flood warning written by hydrologists at the Bureau of Meteorology.

MINOR FLOOD WARNING FOR THE SNOWY RIVER

Issued at 9:58am EST on Wednesday 15 July 2015

River levels at Orbost are currently around the Minor Flood Level (4.2 metres) and rising. A peak of around 4.3-4.4 metres is expected during Wednesday afternoon [15/07/2015].

In the interests of community safety the SES suggests the following precautions:

Don't walk, ride or drive through floodwater,

Don't allow children to play in floodwater,

Stay away from waterways and stormwater drains, and

Keep well clear of fallen power lines

Current Emergency Information is available at http://www.ses.vic.gov.au

For emergency assistance call the SES on telephone number 132 500.

For life threatening emergencies, call 000 immediately.

The SES advises that rainfall run-off into waterways in recent fire affected areas may contain debris such as soil, ash, trees and rocks. People in fire affected areas should be alert to the potential for landslide and debris on roads.

Weather Forecast:

For the latest weather forecast see www.bom.gov.au/nsw/forecasts/

Next Issue:

The next warning will be issued by 10:00am Thursday [16/07/2015].

Latest River Heights:

Snowy R. at Basin Creek 4.33m falling 09:16 AM WED 15/07/15

Buchan R. at Buchan 1.65m falling 08:45 AM WED 15/07/15

Snowy R. at Jarrahmond 4.35m rising 09:00 AM WED 15/07/15

Snowy R. at Orbost 4.18m rising 09:00 AM WED 15/07/15

For latest rainfall and river level information see www.bom/gov.au/nsw/flood/

Chapter 2. The current state of large scale flood forecasting

Forecast Verification

Currently, the BoM uses a manual verification approach, sampling 10% of the warnings issued, based on specifications set out for each forecast point such as a minimum lead time of 6 hours, or a peak forecast accuracy of ± 0.5m. With updates to the FFWS, verification software will be introduced which will automatically compute statistics analysing the accuracy of the forecast river levels, peak and timing based on a comparison with observed river levels. The lead time provided for warnings will also be analysed and compared to the accuracy specifications, providing a measure of performance for a much greater sample of events, which will in turn drive further system improvement. Additionally, the HyFS continuous short-term forecasts are verified using a 15-day moving average climatology to calculate the mean absolute error skill score.

Operational Applications

At the BoM, the continuous short-term streamflow forecasts are used across Australia to provide an early indication of an upcoming flood event, in order to start making arrangements and decisions. These forecasts are then used as a "heads-up" to start running event-based models at the local scale to provide the official, public flood warnings. This is an excellent example of the use of large scale flood forecasting systems to enhance and supplement existing, local-scale forecasting capabilities.

2.3.4 The U.S. Hydrologic Ensemble Forecast Service

The HEFS is run by the NWS, and, for river basins across the U.S., provides "uncertainty-quantified forecast and verification products" (Demargne et al., 2014). From the late 1990s, NWS service assessments, alongside feedback from end users and the US National Academies (National Research Council, 2006) began to confirm the need for probabilistic river forecasts, for flood forecasting and water resources. In 2012, the HEFS began to run experimentally at several regional River Forecast Centres (RFCs), each of which forecasts streamflow for 100s of river locations, and is currently being rolled out operationally at all 13 RFCs.

The HEFS aims to produce ensemble streamflow forecasts which seamlessly span lead times from less than one hour up to several years, and which are spatially and temporally consistent, calibrated (i.e. unbiased with an accurate spread) and verified.

Model Components

The HEFS consists of five main components (Demargne et al., 2014), detailed in figure 8, and has been implemented to run as part of each RFCs configuration of the FEWS-based Community Hydrologic Prediction System (CHPS), which since 2010 has been the software platform used to run the traditional deterministic flood forecasts and long-range ESP forecasts. The system is designed to be driven with four meteorological forecast inputs, two of which (GEFS and CFSv2) are the output of NWP models; whereas the RFC forecasts and climatologies are created by meteorologists for the spatial units of the RFCs' watershed models using predictions from the NCEP Weather Prediction Center (WPC), local NWS Weather Forecast Offices (WFOs) and other sources (NOAA, 2012).

Each RFC may use different combinations of the 19 components within the Hydrological Processor (HP) suite, but the majority of RFC operations centre on a lumped implementation of the SAC-SMA (Burnash et al., 1973) and SNOW-17 (Anderson, 2006) models. The pre-processing step within the HEFS (MEFP, figure 8) creates an ensemble of seamless, hours-to-seasons, calibrated weather and climate forcings which are fed into the HP. Notably, through use of the MEFP and EnsPost pre- and post-processing components, both the uncertainties in the meteorological input and the hydrology are taken into account.

Forecast Visualisation

The graphics generator (figure 8) uses the resulting ensemble hydrographs to produce visualisations of the forecasts which can be communicated to a range of end users for the purpose of decision-making and warning dissemination. These final forecast products include spaghetti plots, exceedance probabilities in the form of bar graphs and probability distribution plots using comparisons with historical simulations (reanalysis datasets), and an expected value chart describing the ensemble distribution. Currently, graphics from the HEFS are operational at only a handful of RFCs and are currently being rolled out at the remaining RFCs. An example of an HEFS hydrograph for one river location, alongside the public web interface, is shown in figure 9. The forecast data associated with the graphical products is typically also available from the RFCs and many users can access the data directly to drive local decision support models.
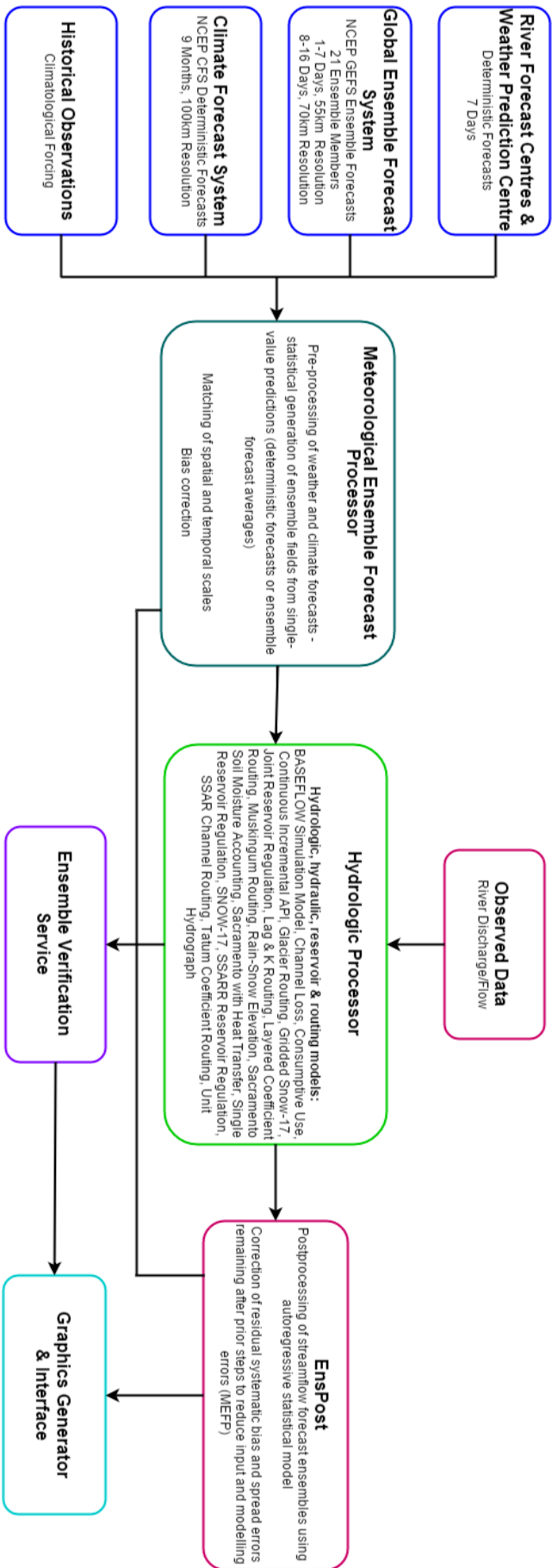
Figure 8: Components of the U.S. Hydrologic Ensemble Forecast System (HEFS).

Warning Dissemination

NWS product requirements are codified through NWS Directives (NOAA, 2015), and the RFCs generally issue products based on hydro-meteorological analyses and long-range predictions that are not time-critical, and inform non-hazard related user activities and decisions, such as the *Streamflow Guidance*. The NWS Weather Forecast Offices (WFOs), in contrast, issue the primary hazard-centred alerts related to flooding, including products such as a *Hydrologic Outlook* ("hydro-meteorological conditions that could cause flooding or impact water supply"), *Flood Watch* (flooding is likely), or *Flood Warning* (flooding is imminent or occurring). The WFO hydrologic products are based primarily on RFC analyses and predictions; for instance, an RFC forecast exceeding a flood threshold triggers a recommendation to the WFO to release a flood warning that is reviewed by the WFO forecaster. Protocols for linking the newer HEFS ensemble forecasts to alerts are still in development.

Forecast Verification

An additional component of the HEFS shown in figure 8 is the Ensemble Verification System (EVS), which produces statistics such as the bias in the forecast probabilities, the skill relative to a 'baseline' forecasting system and the ability to discriminate between events. EVS runs within HEFS and is also freely available as a standalone application. The verification statistics are provided as graphical and textual products. They are used to guide research and development of the HEFS and to improve the configuration of the HEFS for operational forecasting. Studies by Brown et al. (2014a, 2014b) found that the skill of the precipitation forecasts used for the HEFS are greatest at lead times of up to one week, for moderate precipitation, and in the wet season (December to March), with limitations in the summer season due to difficulties in forecasting convection. The studies also showed that the skill of the streamflow forecasts, for both the HEFS and traditional RFC deterministic forecasts, is substantially increased through use of the EnsPost component.

Operational Applications

The HEFS is currently being implemented by all thirteen NWS RFCs, with existing or proposed applications ranging from flood forecasting to river navigation, reservoir operation, and long-term planning and management of water resources. For example, reforecasts and operational forecasts from the HEFS are being used by the New York City Department of Environmental Protection (NYCDEP) to improve the management of water supply to NYC by optimizing the

quantity and quality of water stored in the NYC reservoirs while avoiding unnecessary infrastructure costs.

Figure 9: The U.S. Hydrologic Ensemble Forecast System (HEFS) overview map of locations forecasting floods, with colour representing flood severity. An ensemble hydrograph is shown for a flood event at one river location, including observed stage and flow (green), forecast stage and flow (purple) in terms of probabilities, and colours indicating the forecast severity based on flood stage data (minor flood – yellow, moderate flood – red, major flood – pink). Forecasts are available at water.weather.gov/ahps/forecasts.php

2.4 Global Scale Flood Forecasting Systems

At present, there are just two flood forecasting systems that are operational at the global scale: the Global Flood Awareness System (GloFAS) of the ECMWF and EC, and the Global Flood Forecasting and Information System (GLOFFIS) run by Deltares. There also exists a Global Flood Monitoring System (GFMS) developed by NASA (the National Aeronautics and Space Administration) and the University of Maryland, which uses satellite precipitation as input to a hydrological model to produce real-time global maps of flood events. Global flood monitoring is an important aspect of disaster risk reduction and has many potential applications across the globe; however the GFMS is not an operational hydro-meteorological flood forecasting system and as such is not discussed in detail in this review. The reader is referred to the GFMS website (NASA, 2015) and publications (Wu et al., 2014; Yilmaz et al., 2010) for further information on the GFMS. This section discusses the components of GloFAS and GLOFFIS, alongside the products and warnings provided to end users and verification techniques used to assess the performance of these systems.

2.4.1 The Global Flood Awareness System

GloFAS has been producing probabilistic flood forecasts with up to two weeks lead time in a pre-operational environment since 2011 (Alfieri et al., 2013); this environment enables continuous research, development and testing in order to produce an operational tool that is independent of administrative and political boundaries. GloFAS can provide downstream countries with early warnings and information on upstream river conditions alongside global overviews of upcoming flood events in large river basins, for decision makers ranging from water authorities and hydropower companies to civil protection and international humanitarian aid organisations.
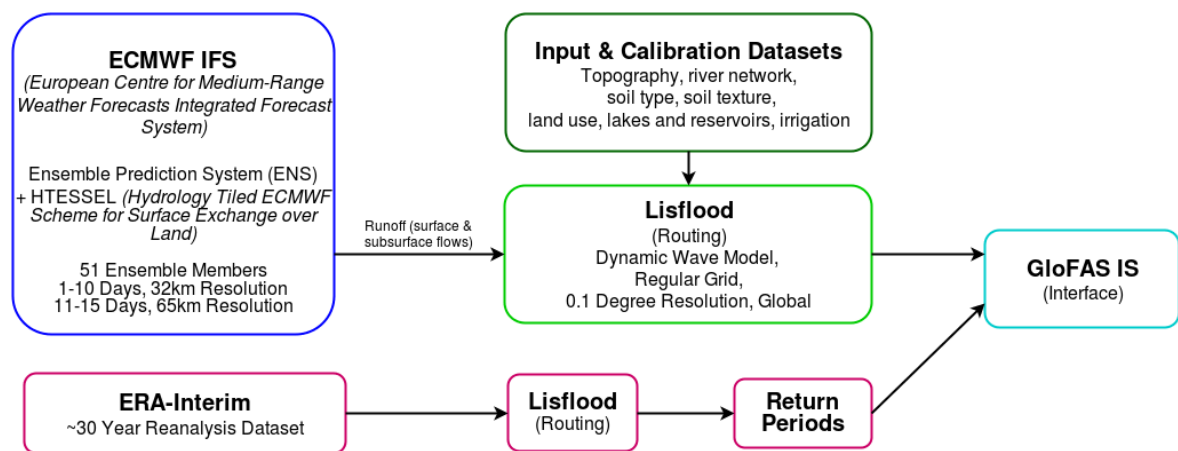
Figure 10: Components of the Global Flood Awareness System (GloFAS).

Model Components

In contrast to the other systems presented in this paper, GloFAS uses surface and sub-surface runoff forecasts produced by the NWP model rather than a separate rainfall-runoff component (figure 1). The Hydrology Tiled ECMWF Scheme for Surface Exchange over Land (HTESSEL) is contained within the IFS and is used as forcing for the Lisflood river routing model. Figure 10 details the components of GloFAS. Although Lisflood global (Van Der Knijff et al., 2010) is also a rainfall-runoff model, it is used here to simulate the routing processes and the groundwater processes, after re-sampling the runoff forecasts from the IFS to the 0.1° resolution of Lisflood. Additionally, GloFAS contains a loss function to account for water loss within the channel reaches in arid areas, which also simulates the river-aquifer and river-floodplain interaction and the influence of evaporation from large rivers.

Runoff from the ECMWF ERA-Interim reanalysis archive has also been run through Lisflood offline, producing a deterministic climatology of river flow which is used to compute return periods for the global river network.
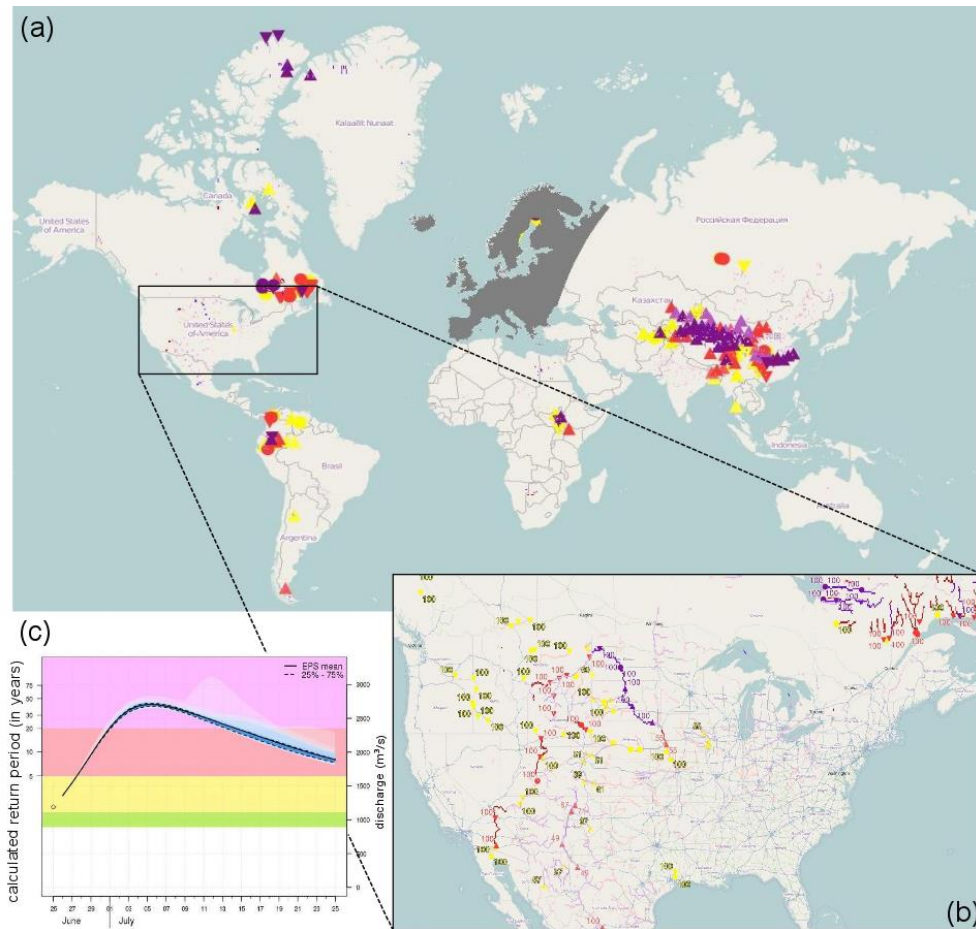
Figure 11: The Global Flood Awareness System (GloFAS) interface showing (a) a global overview of severe (purple) high (red) and medium (yellow) reporting points, (b) a more detailed view of warning points in the U.S.A., (c) the return period hydrograph with return period thresholds (1.5 – green, 2 – yellow, 5 – red, and 20 years – purple) for one point in the U.S.A. Forecasts are available at www.globalfloods.eu

Forecast Visualisation

Forecasts and warnings produced by GloFAS are provided through a password-protected interface (figure 11) where users can register to see a global overview of warning points, forecast precipitation accumulations, ensemble hydrographs including return period threshold exceedances and warnings, and persistence diagrams. The ECMWF and JRC do not directly disseminate flood warnings, as each country has national procedures to follow, but anyone is able to access and analyse the forecasts for decision-making purposes and research. It is noted that due to the forecast and warning responsibilities within Europe, all

countries for which EFAS produces forecasts are removed from the GloFAS interface as these are not publicly available.

Forecast Verification

Alfieri et al. (2013) analysed the performance of GloFAS, and found that forecasts were skilful at 58% of stations, which increased to 71% when model bias was removed. Evaluation of the early warning system found that the longest lead times, exceeding 25 days in some regions, are found in large river basins in South America, Africa and South Asia, while smaller basins have a maximum lead time of 20 days, and in some cases 10 days. The least skilful forecasts were for stations in arid and semi-arid regions, such as Australia, Mexico and the Sahel. Other discrepancies were found in relation to the modelling of snow accumulation and melting processes in HTESSEL and therefore the timing of the peak discharge during spring in snowmelt regions. Evaluation of GloFAS is updated regularly to reflect its continued and ongoing development.

Operational Applications

As of the 14th September 2015, GloFAS has 177 registered users from governmental or other public authorities (~28%), non-governmental organisations (NGOs, ~7%), the private sector (~10%), and from academic/training and/or research institutions (~55%). As with EFAS, GloFAS is used by national services to provide additional early flood information, and is used by, for example, civil protection and humanitarian aid organisations who benefit from a global overview of flood events and may have no other source of information for the region of interest. GloFAS is also used by the ERCC for the purpose of compiling reports on natural hazards and flood risk across the globe.

2.4.2 The Global Flood Forecasting Information System

The Global Flood Forecasting Information System (GLOFFIS) is a research-oriented operational system based on Delft-FEWS (Werner et al., 2013). GLOFFIS is one of three global systems run by Deltares in The Netherlands; also operational are a storm surge model, GLOSSIS (Deltares, 2018), and a water scarcity system GLOWASIS. These three systems belong to an open experimental Information and Communications Technology facility, IdLab, and are being used to test new ideas around interoperability, hydrologic predictability, big data and visualisation.
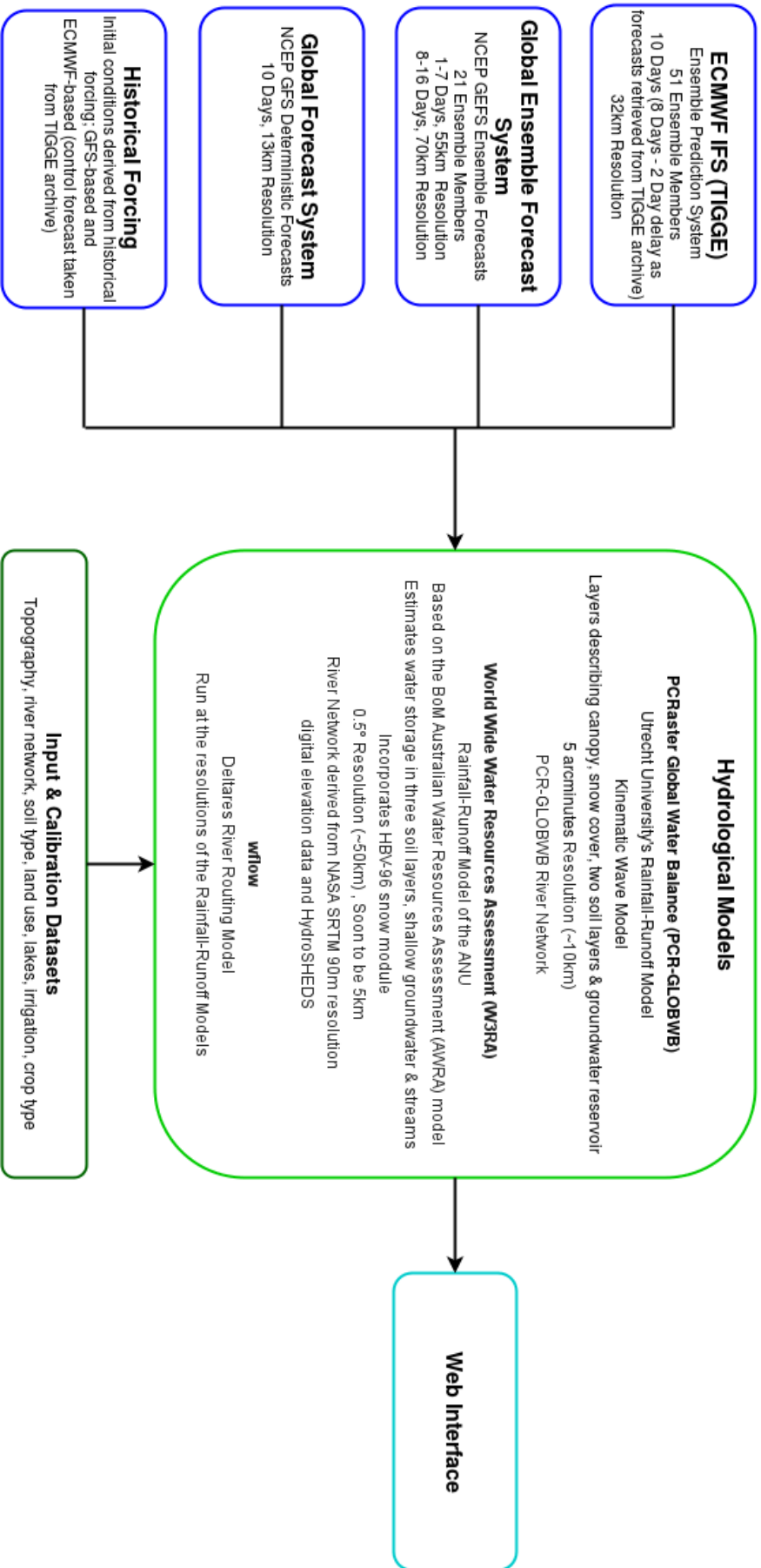
Figure 12: Components of the Global Flood Forecasting Information System (GLOFFIS).

Model Components

Similarly to the approaches taken by many of the continental-scale flood forecasting systems, GLOFFIS uses several meteorological inputs to drive the hydrological component of the system. The idea behind this is to validate, verify and inter-compare real-time rainfall (alongside temperature and potential evaporation) products as they become available. The initial conditions are derived from historical forcings based on both the GFS and the ECMWF control forecast (also extracted from the TIGGE archives), and a combination of FEWSNET (Africa) and Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Global Daily Precipitation, complimented by GFS temperature and potential evaporation. Each of the NWP inputs are fed into two hydrological models (with multiple initial conditions); PCR-GLOBWB and W3RA, which also incorporates the HBV-96 snow module, to account for snow processes.

The current components and resolution of GLOFFIS are detailed in Figure 12, with plans to update the resolution of the W3RA component to 0.05∘ (~5km) and implement an improved river network. In the future, the Japan Aerospace Exploration Agency (JAXA) Global Satellite Mapping of Precipitation (GSMaP) and the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) products will also be added as additional datasets from which to derive initial conditions.

Figure 13: Runoff output of the Global Flood Forecasting Information System (GLOFFIS) W3RA model in the Delft-FEWS forecast platform interface.

Chapter 2. The current state of large scale flood forecasting

Forecast Verification

Thorough statistical verification of GLOFFIS is underway using available open discharge and meteorological forecast data, alongside (real-time) eyeball verification. Real-time discharge data is being collected and can be accessed and compared with the simulated discharge within the Delft-FEWS GLOFFIS platform and reports generated by the system. The verification threshold levels are derived from long historical discharge records and historical simulations, similarly to the methods used in other continental and global scale forecasting systems.

Operational Applications

Although GLOFFIS is not yet fully implemented, it is being used internally at Deltares and by their customers, with discussions already underway between Deltares and other potential end users of the system. GLOFFIS is intended to be a research tool on predictability and interoperability first and foremost, but will be suitable for a variety of applications once fully operational.

2.5 The Grand Challenges of Global Scale Flood Forecasting

There are many challenges associated with global scale flood forecasting. These range from insufficient data, difficulties combining models and computer resource requirements, to the cost of running these models and methods of communicating forecasts efficiently. The challenges faced in operational flood forecasting are discussed in detail by Cloke and Pappenberger (2009), Hannah et al. (2011), Wood et al. (2011), Liu et al. (2012), Pappenberger et al. (2013, 2015a), Kauffeldt (2014), Pagano et al. (2014b) and Bierkens (2015); this section focusses on the current capabilities of the systems reviewed here, and discusses some of the grand challenges of global scale flood forecasting based on the current system limitations, alongside experiences and lessons learned from the development of these systems.

2.5.1 Current Capabilities

Large-scale flood forecasting has only become possible in recent years, and systems such as those outlined in this review are able to produce coarse-scale discharge forecasts at spatial scales covering entire continents or the globe using NWP products and other expertise, comparing these to observed and modelled historic events in order to produce forecasts of flood events in the medium-range, typically 7-15 days. Results from EFAS suggest that river flow and flood forecasts driven by meteorological forecasts are able to provide significant added value to the monitoring of European rivers (Alfieri et al., 2014a), whilst for GloFAS, results show

that the maximum added value is shown "(i) in medium-size river basins, (ii) in those with relatively fast response and (iii) in basins with no definite trend in the seasonal runoff", with lead times of up to 1 month possible in some large river basins (Alfieri et al., 2013). These systems are also capable of producing and disseminating basic forecast, and in some cases early warning, products in real time and are key in supplementing national and local flood forecasting capabilities while supporting global scale activities.

A recent study by Pappenberger et al. (2015a) provides evidence of the economic benefits of large-scale flood early warning systems, in addition to the clear benefits of forecasts and early warnings to populations at risk of flooding. The study demonstrates that the monetary benefit of EFAS is ∼400 Euros for every 1 Euro invested, indicating that large scale flood forecasting systems not only have the capability to provide early awareness of potential severe events, but also provide economic benefits through potential avoidance of flood damages.

2.5.2 Improving Data Availability

Grand Challenge: To access data of sufficient quality and length, assimilate new types of observations and meaningfully incorporate data of inhomogeneous quality.

One of the major challenges in large scale forecasting lies in the availability of input data of the quality that is required (Hannah et al., 2011), such as data required for estimation of the initial hydrologic state, geographical boundaries of river basins and large/global scale datasets of land use, soil data etc. For example, smaller-scale national flood forecasting systems are often able to assimilate or update discharge information in real time, whereas continental and global scale models are limited by the lack of availability of real-time, open data for this purpose.

Alongside the technical challenges associated with accessing and assimilating the data, there exist also non-technical data challenges. For example, there are difficulties with retrieving, quality controlling, formatting, archiving and redistributing the data collected (Pagano et al., 2014b) at centres across the globe. This often requires specialised training and staff, for example at the U.S. National Weather Service, much of the hydrologists' time is spent processing data and maintaining the infrastructure used to archive and distribute the data, and the stream measurements used in the BoM system are collected by several hundred entities and must be collated before processing.

More international and interdisciplinary data sharing (Hannah et al., 2011), through institutions such as the Global Runoff Data Centre (GRDC), and co-operation is essential in moving forward with global scale forecasting efforts, and would greatly increase the data available to forecasting centres not just for use in forcing these models, but for verification of the forecasts and continuous improvement of forecast accuracy. In order to work towards overcoming this challenge, it is important to contribute to open data policies and ensure that data availability is at the core of all related activities.

2.5.3 Model Parameterisation

Grand Challenge: To find regionalisation methods and ways to represent sub-grid scale uncertainty on the global scale.

Alongside the problems associated with the data required for forecasting flood events, there are further challenges involved in the parameterisation of models, and the use of a single model for all catchments across a continent or the globe. Wood et al. (2011) discuss the possibility that much higher resolution forecasting systems will soon be feasible, which would further provide detailed information regarding the storage, movement, and quality of water. In order to implement models of higher resolutions, there are other challenges that must also be addressed; these challenges lie in the parameterisation of processes at both current and future spatial resolutions, and the "lack of knowledge involved in evaluating and constraining the uncertainty in those parameters given current and future data availability" (Wanders et al., 2014).

This challenge could be addressed, for example, by developing scaling theories to represent effective parameterisation and associated uncertainties relevant to a global forecasting chain, and methods which can incorporate largely varying data and information availability.

2.5.4 Improving Precipitation and Evaporation Forecasts

Grand Challenge: To translate improved precipitation and evaporation forecasts into improved discharge forecasts.

There have been many improvements in NWP and precipitation forecasting thus far, which have enabled global flood forecasting, as discussed earlier in this review. Despite these improvements, there are still limitations in the NWP forecasts which affect the discharge and therefore flood forecasts. Some of these have been discussed, such as difficulties predicting convection (Krishnamurti et al., 1999) and orographic enhancement processes (Arduino et al., 2005). It is not only precipitation forecasts which need to be further improved, but other NWP

variables used in hydro-meteorological forecasting systems, such as evaporation. The challenge then lies in translating the continuous improvements made to the NWP forecasts into improved discharge forecasts.

Moving forward, it will be important to develop tools and methods, such as satellite measurements, to measure potential evaporation and precipitation on a global level with acceptable accuracy.

2.5.5 Incorporating Anthropogenic Influences

Grand Challenge: To understand which of the anthropogenic influences are having a significant impact on hydrological forecasting and therefore need to be included in global forecasting models.

The lack of knowledge of anthropogenic influences on runoff is a major challenge for large scale flood forecasting. These influences, for which there is currently no global database, include dams and their regulation, reservoirs, weirs, water extraction, irrigation and river re-routing; some of this activity also goes unreported and unregulated creating additional barriers to incorporating information on water management. One of the specific challenges noted by SMHI for Europe is the changes in processes modelled within these systems due to depleted aquifers.

It is also important for these systems to incorporate aspects of anthropogenic influence such as land use and urban areas. Many of the users of these systems require information on potential impacts of the forecast flood events, for example the number of people likely to be affected and how much agricultural land is threatened. The inclusion of more impact information is one of the current limitations and focusses for the development of EFAS and GloFAS. A further challenge exists in terms of the unevenly distributed global population, which results in sparse data networks in large, unpopulated regions and difficulties in the dissemination and communication of forecasts and warnings; this challenge is specifically mentioned by the BoM for Australia, but exists also at the global scale.

In order to account for anthropogenic influences in global flood forecasting systems, one solution would be to map all of these influences, and perform a sensitivity analysis to determine which are impacting the forecasts, so that the key anthropogenic influences can be incorporated into the models.

2.5.6 Resources and Costs

Grand Challenge: To quantify, understand and communicate the values and benefits derived from a global forecast whilst establishing a cost effective execution of these forecasts.

Thus far, the spatial resolution of global scale land surface models has largely been constrained by the computational resources required to run global weather models; currently, at best, ~20 km. The monetary costs of producing forecasts using large-scale prediction systems must also be taken into account. While the costs of running these systems are not generally published, the aforementioned study by Pappenberger et al. (2015a) states that the estimated cost of EFAS (across the four EFAS operational centres, see section 2.3.1) is 1.8 million Euros per year, with an estimated 20 million Euros in development costs over 10 years. In addition, with each improvement and update to a forecasting system, it also becomes necessary to re-run model climatologies, re-calculate thresholds and revise decision-making criteria, all of which can be technologically challenging and require significant computational time and resources (Pappenberger et al., 2010; Simmons and Hollingsworth, 2002).

As these systems develop, the resources required to run global flood forecasting systems will be reduced, whilst the technology used continues to improve. This will enable more centres to run global models at lower costs and with fewer time constraints in the future.

2.5.7 Effective Communication of Forecasts

Grand Challenge: To communicate uncertainties to a large range of user groups in countries across the globe, some of whom will not be known. Additionally, to embed these systems into national warning chains, whilst respecting sensitivities associated with the single voice principle (WMO, 2005).

A key challenge associated with global scale flood forecasting stems from the understanding and communication of flood forecasts. For instance, with the move towards ensemble flood forecasting, there is also a need for improved understanding of probabilistic forecasts. Ensemble forecasts produce large amounts of information, and it is vital that the most important information is conveyed appropriately for ease of use and correct interpretation of the forecasts, allowing for well-informed decisions and promoting a common understanding between end users.

One of the key challenges at present for EFAS is ensuring that the flood forecast and warning information is easily accessible to a broad range of users from countries across Europe, who

interpret the forecasts very differently. This challenge is amplified further when producing forecasts, as with GloFAS and GLOFFIS, for the entire globe and a spectrum of users ranging from experts in the fields of hydrology and meteorology, to those with no experience in using these types of products. GloFAS already has a range of partners and end users, from those who are interested in discharge forecasts for specific stations, to those who are interested purely in the impact of the floods. An additional consideration is that of the single voice principle, which states that national services constitute the single authoritative voice on weather warnings in their respective countries. As more systems are introduced with the capability to produce forecasts and warnings, the more difficult this principle becomes; in future it may be that many institutions are able to disseminate warnings and benefit from the wealth of available forecasts and information, and a new challenge of the systems will be to become the trusted source of information.

In order to effectively communicate forecasts and warnings, it is important to co-develop the forecast visualisations and warnings with a large range of users, and enable some flexibility for users to customise the interface. International and interdisciplinary cooperation is also key in moving forward with this challenge, as issuing forecasts and warnings can be challenging without the existence of a political agreement between upstream and downstream countries for the sharing of information related to floods (Hossain and Katiyar, 2006).

2.5.8 Forecast Evaluation and Intercomparison

Grand Challenge: To find new and novel methods to verify extremes, which are suitable for hydrological forecasting.

Many forecasting systems, including large-scale flood forecasting systems, are moving towards ensemble forecasting methods. While there are many benefits to using a probabilistic approach, a key challenge associated with ensemble flood forecasting is the evaluation of flood forecasts, due to the low frequency of occurrence of extreme floods alongside the lack of data from different flood events (Cloke and Pappenberger, 2009). The analysis of an ensemble's ability to fully represent the uncertainty is also complex and uncertain in itself.

This relates to a further grand challenge; that of implementing a Flood Forecasting Intercomparison Project to compare various aspects of these large-scale operational flood forecasting systems. This will be a valuable and important project moving forward, as these systems become more advanced and widely used for many applications, but is currently not undertaken due to the difficulties involved in comparing models of a variety of different scales,

with varying system set-ups and interfaces, and different objectives and end users. The computational resources required for such a project are also extensive.

To have effective forecast evaluation measures in place, it is important for institutions running these systems to facilitate access to the forecasts, in order that the forecasts can be evaluated by an unbiased, external entity.

2.6 The Future of Global Scale Flood Forecasting

Flood forecasting at the large (continental and global) scale is key to providing overviews and early warnings of flood events across the globe, including regions where no alternative local-scale flood forecasts are available. This section outlines aspects of the future of global scale flood forecasting, as we continue to work towards overcoming the grand challenges and move towards ever more valuable multi-hazard forecast and early warning systems.

2.6.1 Adaptive Modelling Strategies

Adaptive modelling strategies involve the idea of adjusting model predictions in real time if discrepancies are observed between the forecast and observations, where discharge measurements are available in real time. This allows the uncertainty in the forecasts to be further constrained. In meteorological applications, this is referred to as data assimilation and is used routinely in weather forecasts and NWP, whereas it is often referred to as updating in hydrology, and is not widely used at present in applications such as those discussed here (Shaw et al., 2011). Simple applications of updating require starting new forecasts using available observations (sequential data assimilation; Rakovec et al., 2012), whereas more complex updating involves the adjustment of current predictions to the observations when discrepancies occur, assimilating the new observed data into the model in real time (variational data assimilation). While data assimilation is not used extensively in flood forecasting systems to incorporate observations into the forecasts, this is likely to be increasingly incorporated in future to further improve the accuracy and lead time of large scale flood forecasts (Liu et al., 2012). An area of research which will be important in moving towards the incorporation of adaptive modelling strategies is the development of data assimilation toolboxes, allowing institutions to use and benefit from data assimilation tools which are otherwise incredibly complex. One example of this is OpenDA, "an open interface standard for a set of tools to quickly implement data assimilation and calibration for arbitrary numerical models" (Deltares, 2015).

2.6.2 Extended-Range Forecasting

Future advances in global scale operational flood forecasting are likely to include more long-range forecasting. There already exists an element of river-specific predictability in some large rivers where the movement of a flood wave downstream can take days or weeks, and a flood event is a relatively certain outcome once large amounts of precipitation are recorded upstream. Realistic initial conditions can be beneficial to seasonal prediction; for example, relatively large soil storage capacity leads to long memory of soil moisture, and the accuracy of soil moisture initial conditions may be key in long-range forecasting (Fennessy and Shukla, 1999). The same is true of snow cover and snow pack, particularly in climate zones where snow is the major water resource (Li et al., 2009).

Seasonal forecasts are currently used across a wide range of weather-sensitive sectors, with many operational weather forecasting centres producing seasonal forecasts, which provide "seasonal-mean estimates" of weather, such as whether the coming season will be wetter or drier than usual (Weisheimer and Palmer, 2014). Such forecasts have the potential to aid the forecasting of floods on seasonal timescales, providing crucial information for flood preparedness and mitigation (Yuan et al., 2015a). Seasonal hydrological forecasting has begun to emerge across the globe over the past decade, due to the ongoing development of coupled atmosphere-ocean-land general circulation models, while seasonal water supply forecasts have been used in the U.S. since the 1930s based on snow survey measurements, and later, precipitation data (Pagano et al., 2014a). Yuan et al. (2015a) highlight several questions related to the future of seasonal hydrological forecasting, from how to combine weather and climate models towards seamless hydrological forecasting, to how to improve the prediction of interannual variability of variables relevant to hydrological forecasting applications. Further to this, there also exists the challenge of the effective communication of seasonal flood forecasts and transfer of these forecasts into warnings and actions (Yuan et al., 2015a). The WMO S2S (Subseasonal to Seasonal) prediction project (WMO, 2015) aims to improve the understanding and forecast skill of the sub-seasonal and seasonal timescales, with a focus on extreme weather including floods, and will be key in moving towards extended-range flood forecasts.

2.6.3 Flash Flood Forecasting

Flash floods are associated with spatially and/or temporally intense precipitation and can have high societal impacts. For example 105 out of 139 countries list flash floods as being in the top

two of their most important hazards (WMO, 2006). Despite this there is currently no global flash flood forecasting system, but continental systems exist in Europe (as part of EFAS; Raynaud et al., 2015; Thielen et al., 2009); northern America (Gourley et al., 2012), southern Africa (Georgakakos et al., 2013) and Australia, alongside other national and basin scale systems around the globe (Hapuarachchi and Wang, 2008). These systems often take the form of one or a combination of empirical correlations, unit hydrographs and hydrological modelling driven by limited-area models (Hapuarachchi et al., 2011).

The challenge of creating a global flash flood forecasting system is that global NWP systems typically have a limited resolution of many of the fine spatial scale processes, such as convection, which are responsible for intense precipitation. Increasing the spatial resolution of global NWP systems may reduce this issue and allow for the implementation of a methodology such as that of Alfieri et al. (2014b), which utilises the surface runoff estimated from HTESSEL to forecast extreme runoff risk. An alternative could be to use forecasts of parameters which can be used to estimate the likelihood of intense sub-grid scale precipitation arising. For example, the ECMWF NWP model forecasts the CAPE (convective available potential energy) and CAPE-SHEAR parameters which show the atmospheric instability and the ability of supercell formation in the event of deep moisture convection, respectively (Tsonevsky, 2015).

With continuous improvements to NWP systems, new continental and global flash flood routines will be developed based on global NWP models (ECMWF, 2015). In addition to flash floods, future applications of global flood forecasting and multi-hazard early warning systems will begin to include other types of flooding, for example coastal storm surges.

2.6.4 Grand Ensemble Techniques

Recent advances in meteorological forecasting and NWP have moved towards multi-model forecasts and grand ensemble techniques. Programmes such as TIGGE (The Observing System Research and Predictability EXperiment (THORPEX) Interactive Grand Global Ensemble; ECMWF, 2006), have led to advances in ensemble forecasting, predictability and development of severe weather prediction products in meteorology. In hydrology, combining models for flood forecasting presents an additional challenge (e.g. due to different river networks and climatologies), but despite this, future applications of flood forecasting should move towards the establishment of grand ensemble techniques. In the future, increased access to monthly and subseasonal (for example, through the S2S project; WMO, 2015) forecasts from

multiple centres will enable us to push the limits of predictability through use of these grand ensemble techniques (Fan et al., 2015).

2.6.5 New Data Possibilities

Alongside the recent and future advances in forecasting systems, other technologies are constantly advancing and will have beneficial impacts on flood forecasting across the globe. For example, new satellites and earth observation technologies for flood observation are being adopted in hydrology to improve flood forecasts (García-Pintado et al., 2015; Khan et al., 2012). Garcia-Pintado et al. (2015) discuss several earth observation techniques which have the potential to improve flood detection and forecasting. Improved data from satellites may be able to provide more accurate topographical, land cover, land use, river network and river width information (Yamazaki et al., 2014); these are some of the most important data regarding river basin characteristics, and their accuracy is key to flood forecasting systems. Real time satellite observations of river width during flooding would also serve to improve both forecasts and warnings in real time, and verification of the forecasting systems post-event.

Alongside improved databases describing basin and river characteristics, observations of the data used as input to flood forecasting systems and in data assimilation techniques (Liu et al., 2012) could include snowpack extent, water levels (from altimetry), river discharge, river width, snow and soil moisture. Currently, continental and global scale observations of many of these variables are not available, but global coverage from satellites could prove extremely beneficial in large-scale flood forecasting applications, particularly in regions of poor data availability (Wanders et al., 2014).

2.7 Conclusions

Here, two global and four continental scale operational flood forecasting systems have been reviewed, outlining the current state-of-the-art in operational large-scale flood forecasting. Producing forecasts at the global scale has only become possible in recent years, with scientific and technological advances and increasing integration of hydrological and meteorological communities. Due to these recent advances, large-scale flood forecasting systems are able to produce coarse-scale discharge forecasts at spatial scales covering entire continents or the globe using NWP products and other expertise, comparing these to

observed and modelled historic events in order to produce medium-range forecasts of flood events.

Many countries are required to prepare for floods which originate outside of their borders. International and interdisciplinary collaboration is key in order to overcome many of the challenges involved in transboundary flood forecasting; large-scale forecasting systems have the potential to provide valuable added information about imminent flooding. So far, results from large-scale flood forecasting systems suggest that river flow and flood forecasts are able to provide significant added value to the monitoring of rivers across the globe (Alfieri et al., 2013; Pagano et al., 2014b). There remain many challenges for global scale flood forecasting, from lack of available data of the quality and scale required, to the effective communication of forecasts and warnings to varying end users and communities across the globe; ongoing research aims to overcome these challenges to further improve the accuracy and applicability of large-scale flood forecasting. The systems outlined in this paper are continuously evolving and are already proving to be key in supplementing national and local forecasting capabilities while supporting global-scale activities.

Chapter 3

El Niño Southern Oscillation

3.1 Introduction

In the previous chapter, the current state of large-scale flood forecasting was reviewed, alongside the grand challenges and future advances of large-scale flood forecasting, including extended-range forecasts out to seasonal timescales. Despite the chaotic nature of the atmosphere (Lorenz, 1963), which introduces a limit of predictability, seasonal predictions are possible as they rely on components that vary on longer timescales and are themselves somewhat predictable, such as the ocean and land surface. This "second type predictability" (Lorenz, 1993) for river flow and hydrological forecasting comes from the initial conditions of the land surface, including soil moisture and snow cover, and from large-scale modes of climate variability.

This thesis aims to explore ways in which we can extend the predictability of flood hazard beyond the capabilities of medium-range forecasting systems, and provide even earlier indications of potential flood events, many weeks or even months in advance. Section 1.1 introduced the two key ways to achieve this. Firstly, through statistical analysis based on large-scale modes of climate variability, and secondly, through seasonal forecasts of river flow produced using coupled ocean-atmosphere general circulation models. Both are explored in the following chapters of this thesis. While the latter was introduced in section 2.6.2, and is expanded on in Chapter 5, this chapter provides additional background information on the potential for predictability through climate variability, focussing on the El Niño Southern Oscillation (ENSO), the dominant mode of interannual climate variability (McPhaden et al., 2006).

The modulation of extreme events, such as flooding, by the large-scale circulation is often the origin of predictability of these events at subseasonal and seasonal timescales (Vitart, 2014). The link between large-scale atmospheric features and teleconnections has been shown, for other events such as tropical cyclones and extreme heat, to improve predictability and extend the lead time of forecasts of extreme events (Vitart, 2014). In addition to ENSO, there exist several other modes of climate variability that influence river flow regionally and can contribute to seasonal predictability of hydrological variables. These teleconnections include the North Atlantic Oscillation (NAO), the Indian Ocean Dipole (IOD) and the Pacific Decadal Oscillation

(PDO) (Yuan et al., 2015a, and references therein). This thesis focusses on ENSO as a potential source of flood hazard predictability due to its global influence on weather and climate.

3.2 ENSO Dynamics

ENSO is the largest signal of interannual climate variability (Rasmusson and Wallace, 1983; Kessler et al., 2015); it is a naturally occurring phenomenon in the tropical Pacific that impacts weather, climate and society worldwide. This phenomenon sees sea surface temperatures (SSTs) in the central and eastern equatorial Pacific fluctuate between unusually warm (El Niño) and unusually cool (La Niña) conditions, a cycle that occurs over a period of ~2-7 years. The 'Southern Oscillation' is the term given to the coinciding changes in atmospheric pressure between the east and west Pacific Ocean, which are themselves closely related to changes in the trade winds and represent the "atmospheric manifestation of the coupled ENSO phenomenon" (McPhaden et al., 2006). A schematic is provided in Figure 1.
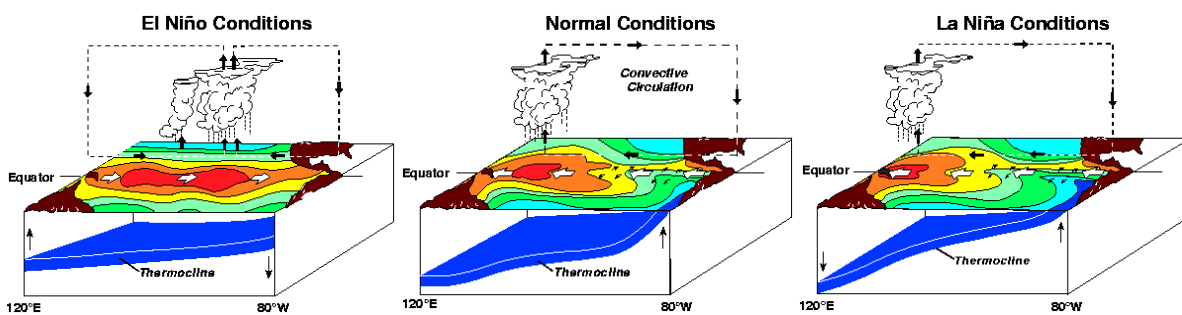
Figure 1: Schematic diagram of El Niño, normal (neutral) and La Niña conditions, indicating changes in SSTs across the Pacific Ocean (where red/orange indicates warmer SSTs, and blue/green cooler SSTs), and corresponding changes in the thermocline and convective circulation. Source: NOAA PMEL, 2018.

El Niño is driven by a positive feedback between the strength of the trade winds and the SSTs in the equatorial Pacific, known as the Bjerknes feedback (Bjerknes, 1966). The trade winds blow from east to west along the equator, and have the effect of pushing warm tropical Pacific waters to the west of the ocean, confining the Warm Pool to the western Pacific (McPhaden and Picaut, 1990) and preventing the upwelling of cold water along the equator and the west coast of South America. This contrast in SSTs between the east and west Pacific further reinforces the atmospheric pressure difference, which in turn drives the trade winds (McPhaden et al., 2006).

During an El Niño, the trade winds weaken, while pressure falls in the east Pacific and rises in the west Pacific. With the trade winds weakened, the warm water doesn't extend as far to the west, resulting in anomalously warm SSTs in the central and eastern Pacific, and the upwelling of cold water is cut off. This positive feedback continues, with the SST anomalies (SSTAs) reinforcing the weakening of the trade winds, and so on, until an El Niño develops. In order for the system to return to neutral conditions, or reverse to La Niña conditions, a negative feedback is required.

While the Bjerknes (positive) feedback drives El Niño, negative feedback is required to terminate an El Niño event. Both the atmosphere and ocean play a role in this, although the majority of the weakening results from the oceanic negative feedback and the delayed action of ocean wave dynamics, which allows the upwelling of cold water in the eastern Pacific to return. The thermodynamical heat flux feedback (Lloyd et al., 2009) is the main atmospheric negative feedback; the warm SSTs during an El Niño result in enhanced convection over the equatorial central and eastern Pacific. The weaker winds during an El Niño are strengthened due to the enhanced convection, and the increased cloud cover reduces incoming solar radiation resulting in a cooling of the SSTs.

The exact mechanism behind the oceanic negative feedback is debated in the scientific literature, with four key proposed theories ('oscillators'; Wang et al., 2016): the delayed oscillator (oceanic wave reflection at the western boundary, Battisti et al., 1989; Suarez and Schopf, 1988), the recharge oscillator (warm water is "discharged" to higher latitudes, Jin, 1997), the western Pacific oscillator (a western Pacific wind-forced Kelvin wave, Wang et al., 1999; Weisberg and Wang, 1997), and the advective-reflective oscillator (anomalous zonal advection, Picaut et al., 1997). It is also possible that more than one of these oscillators occurs, or that they work together to produce the negative feedback required to revert to neutral or La Niña conditions, and as such, the unified oscillator (Wang, 2001) accounts for the dynamics of all four oscillators. Wang et al. (2016) provide a detailed review of the various oscillator theories. Understanding the mechanisms driving ENSO is key in terms of predicting its evolution and therefore the expected impacts of each El Niño and La Niña event.

3.3 ENSO Diversity

ENSO events (El Niño and La Niña) vary in terms of their magnitude, timing and spatial pattern. These variations have been observed for many years, but research into identifying, describing and understanding the possible types of El Niño (while La Niña events also vary

somewhat in these characteristics, the interevent variations are much less distinct) gained momentum after the 2004 event, which exhibited unusual SSTAs and resulted in different impacts than the 'traditional' El Niño (Capotondi et al., 2015). This differing El Niño pattern was termed 'El Niño Modoki' (meaning similar, but different in Japanese) by T. Yamagata in 2004 (Ashok et al., 2007). While references are often made to the 'Modoki' and 'Canonical' El Niño types, they are more generally referred to as Central Pacific (CP) and Eastern Pacific (EP) El Niños respectively, referring to the location of the peak of the SSTAs (Capotondi et al., 2015; Kao and Yu, 2009; Takahashi et al., 2011). The peak SST warming typically occurs either in the central Pacific Ocean, in the Niño3.4 region (see Figure 2), or along the western coast of South America, in the Niño1+2 region. The difference in the spatial pattern of warming between CP and EP El Niño events can be seen in Figure 2.
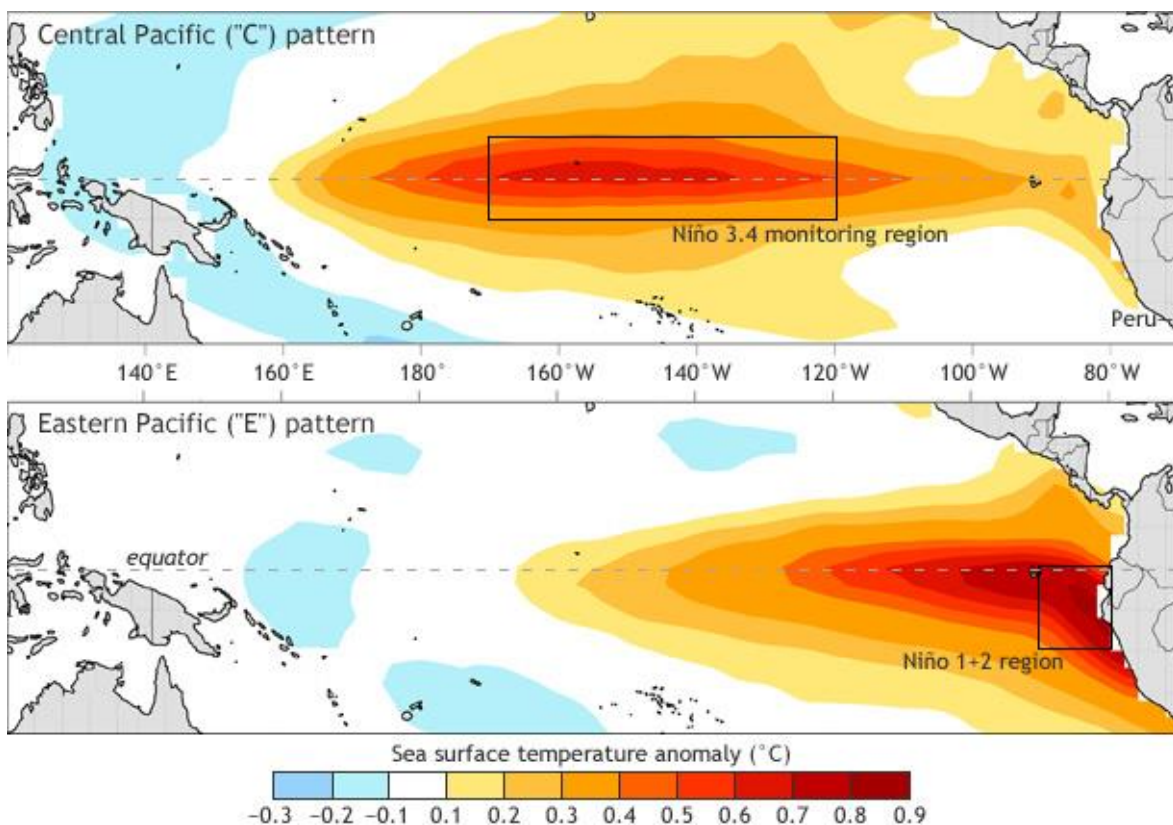
Figure 2: Spatial pattern of SSTAs during Central Pacific (CP, top) and Eastern Pacific (EP, bottom) El Niño events. The Niño3.4 and Niño1+2 regions are highlighted by the black boxes. Source: Takahashi, 2015.

While the CP and EP classifications provide a useful basis for differentiating between the majority of El Niño events, some events cannot be classified as either CP or EP, presenting a more mixed pattern. The classification can also depend on the index and methodology used

(Capotondi et al., 2015; Williams and Patricola, 2018), and variations have been observed within the temporal evolution of a single El Niño event (Karnauskas, 2013). As such, it is suggested that El Niño events exist on a continuum (Johnson, 2013). The origin of the varying types of El Niño, and lack of variation of La Niña, is still debated, but a recent study by Chen et al. (2015) suggests that "the asymmetry, irregularity and extremes of El Niño" result from westerly wind bursts (WWBs), with WWBs tending to be stronger and more frequent during larger magnitude El Niños. WWBs refer to short-lived bursts of westerly winds (opposing the easterly trade winds) in the Western Pacific for a period of several days. They are often connected to the Madden-Julian Oscillation, the dominant component of intraseasonal tropical climate variability (Zhang, 2005), and are known to have occurred with the onset of every El Niño observed during the past 50 years (Chen et al., 2015). WWBs trigger oceanic Kelvin waves, which propagate eastwards across the Pacific, depressing the thermocline in the eastern Pacific and therefore reducing the upwelling of cold water, leading to warm SSTAs (Zhang, 2005) and therefore initiating the development of an El Niño event.

Further to the variations in spatial SSTA pattern, there are also observed differences in the temporal evolution and magnitude of CP and EP El Niños. Onset of EP events occurs typically in (boreal) spring in the eastern Pacific, extending westward through summer and autumn, while CP events typically begin during the summer (Capotondi et al., 2015). Both types of event reach their peak during boreal winter. The magnitude of the SST warming can vary significantly from one event to the next, irrespective of the type of event, but the most extreme El Niños on record (those in 1982/83, 1997/98 and 2015/16) have all been EP events.

ENSO teleconnections can be significantly influenced by the pattern, magnitude and timing of the SSTAs, with some locations observing a different pattern, or even sign, of precipitation and temperature anomalies between CP and EP El Niños (Capotondi et al., 2015). As such, ENSO diversity is a key consideration in terms of predicting the impacts of El Niño events, including flood hazard.

3.4 Influence on Weather and Climate

As discussed in section 3.1, this thesis focusses on ENSO as a source of predictability, due to its influence on weather and climate patterns at the global scale. While there exists a myriad of studies examining the impact of ENSO in specific regions around the world, this section

provides a brief global overview of the impacts on precipitation, temperature and tropical cyclones.

Changes in the Walker Circulation and displacement of convection during ENSO events results in precipitation anomalies throughout the tropics. The Walker Circulation (Lau and Yang, 2015; Wang, 2004) refers to the large-scale atmospheric circulation along the equator; a schematic is shown in Figure 3. As mentioned in section 3.2, the trade winds blowing east to west along the equator have the effect of pushing warm water to the west, creating an SST gradient across the ocean basins. Air rises over the warmer water in the west of the basins and descends in the east, creating the various cells of the Walker Circulation. The more significant the warm pool, the stronger the upward motion and therefore convection and precipitation. During El Niño and La Niña, the changes in SST across the Pacific result in changes to the Walker Circulation.
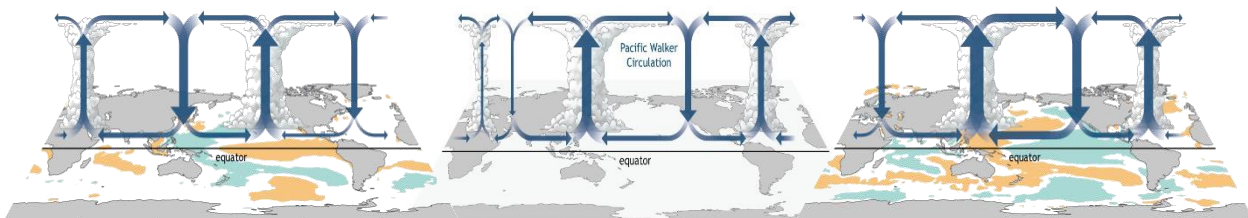
Figure 3: Schematic of the Walker Circulation during El Niño (left), neutral (centre) and La Niña (right) conditions, indicating the atmospheric circulation, areas of convection, and SSTAs (where blue indicates cooler SSTAs and orange warmer SSTAs). Adapted from Di Liberto (2014).

During La Niña, the circulation is amplified, whereas during El Niño, the location of upward motion over the Pacific moves further to the east, disrupting the pattern of circulation across the globe. In addition, the Hadley Circulation, which sees rising air moving away from the equator towards higher latitudes and descending in the subtropics, can be impacted by changes to the Walker Circulation during El Niño and La Niña, thus modifying midlatitude circulation patterns (Wang, 2002). These changes to the global atmospheric circulation therefore result in anomalous weather patterns across the globe.

The impact of El Niño and La Niña on global precipitation and temperature anomalies was widely established in the 1980s and 1990s, based on observations and satellite data (Bradley et al., 1987; Halpert and Ropelewski, 1992; Kiladis and Diaz, 1989; Lau and Sheu, 1988; Ropelewski and Halpert, 1987, 1989, 1996; Stoeckenius, 1981; Trenberth et al., 1998), with ENSO known to impact weather patterns on all seven continents, and in all ocean basins. One of the first studies to assess the impact using a gridded dataset with global coverage was that of

Dai and Wigley (2000). The study found that during an El Niño year, there is an increase in total annual precipitation by just ~0.2%; this is because ENSO events result in shifts in the location of rainfall belts, causing precipitation to fall in different locations than in a normal year. This is further emphasised by Goddard and Dilley (2005), in a study evaluating the impact of ENSO on climate anomalies and number of climate-related disasters, in light of the assumption that El Niño and La Niña result in more widespread climate anomalies "and therefore greater climate-related socioeconomic losses". They conclude that climate-related disasters do not increase during ENSO events, but climate anomalies are more predictable during El Niño and La Niña, and this added predictability could allow for communities from local to international scales to prepare for and mitigate the potential impacts. The extreme El Niño event in 1997/98 introduced a global interest in El Niño, both in the scientific communities and more widely in the general public, and extensive literature has since been published assessing local and regional impacts of ENSO events.
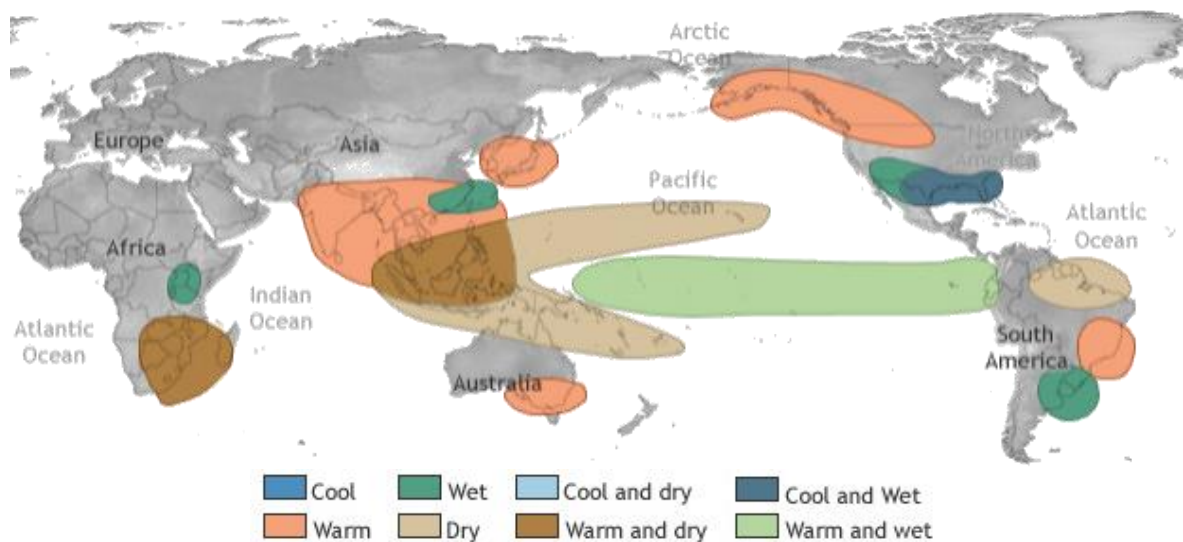
Figure 4: Typical impacts December – February during an El Niño. Adapted from NOAA, Climate.gov (2016c).

Typical ENSO impacts tend to be summarised using maps similar to the one shown in Figure 4. While these are an over-simplification of possible impacts, since there exists substantial uncertainty arising from uncertainty in the data, and from differences in the impacts from one ENSO event to the next (discussed further in Chapter 4), they provide a simple overview of global ENSO teleconnections.

Figure 4 indicates that northern Peru, parts of south-eastern South America, the southern USA, parts of eastern Africa, and China are likely to be wetter than usual during an El Niño, while

southern Africa, northern Australia, maritime south-east Asia and northern South America are likely to be drier than usual. The impacts also vary by season during an El Niño. The reverse tends to be true during a La Niña, with regions likely to be wet during an El Niño, more likely to be dry during a La Niña, and vice versa.

In addition to the global precipitation and temperature teleconnections, ENSO also impacts tropical cyclone activity in the North Atlantic, western North Pacific, southern Indian Ocean and Australian region (Bell et al., 2014; Camargo et al., 2007). Changes in the Walker Circulation influence upper-tropospheric westerlies over the North Atlantic; during an El Niño, these westerlies are increased, which increases vertical wind shear and suppresses tropical cyclone activity (Bell et al., 2014). In both the western North Pacific (Chan, 1985; Jien et al., 2015) and the southwest Pacific (Chand et al., 2013; Nicholls, 1979) during El Niño, tropical cyclone genesis shifts further east, due to the eastward shift of the SSTAs, while tropical cyclone activity is increased due to the warmer SSTs, increased relative humidity and low vertical wind shear.

3.5 Influence on River Flow and Flooding

The influence of ENSO on precipitation further impacts river flow and flooding at the global scale. Furthermore, Chiew and McMahon (2002) propose that the relationship between ENSO and river flow is likely to be more pronounced than between ENSO and precipitation, as rainfall variability is enhanced in runoff, and river flow also integrates information spatially. This, alongside the nonlinear relationship between precipitation and flood magnitude (Stephens et al., 2015), highlights the importance of considering hydrological variables in addition to the meteorology when considering the impacts of El Niño and La Niña.

Again, there exist various studies examining the relationship between ENSO and river flow at local and regional scales. At the global scale, Chiew and McMahon (2002) identified regions of the globe with a significant relationship between ENSO and river flow, which are similar to those where an impact on precipitation is observed, as would be expected. More recent studies have further analysed the link between ENSO and flooding at the global scale. For example, the first studies to assess the impact of ENSO on flooding, at the global scale (Ward et al., 2010, 2014a), found the influence on annual floods to be much greater than the influence on average river flows, with approximately one third of river basins around the globe impacted by ENSO. Ward et al. (2014b) also analysed both the positive and negative socio-economic impacts of El Niño and La Niña in terms of flood risk anomalies, showing "strong, complicated, and societally significant patterns" when looking at spatial variations rather than the global aggregations that

are often reported. Lee et al. (2018) also reinforce the importance of considering hydrological variables, highlighting the differences between ENSO-induced precipitation and streamflow anomalies in regions across the globe. The findings of these studies suggest that there is the possibility to provide probabilistic forecasts of ENSO-driven flood hazard at the global scale.

3.6 Predictability of River Flow and Flood Hazard

The known ENSO teleconnections discussed in sections 3.4 and 3.5 allow for the possibility to predict the likely impacts of ENSO events. While there are challenges associated with producing accurate predictions of ENSO events themselves, such as the spring predictability barrier (whereby forecasts of ENSO made before and during boreal Spring are less skilful; Duan and Wei, 2013; McPhaden, 2003; Wang-Chun Lai et al., 2018), the underlying decadal variability of ENSO (Barnston et al., 2012; Kirtman and Schopf, 1998) and ENSO diversity (see section 3.3), skilful predictions are possible with lead times of up to several months (Barnston et al., 2012). Once an ENSO event is forecast, statistical analyses of ENSO teleconnections can be used to predict the likely impacts should an El Niño or La Niña develop.

Statistical forecasts such as historical probabilities provide information about typical ENSO impacts based on historical evidence. The following chapter builds on the existing literature by mapping the historical probabilities of high and low river flow during El Niño and La Niña, which can be used to highlight regions of the globe that are most likely to be at risk of flooding, or drought, during an ENSO event. Historical probabilities of ENSO-driven precipitation and temperature anomalies, such as those produced by the International Research Institute for Climate and Society (IRI, 2018), are often used for El Niño preparedness activities. This thesis aims to provide the equivalent information for river flow as exists for meteorological variables, thus working towards extending the predictability of flood hazard at the global scale.

Chapter 4

Complex Picture for Likelihood of ENSO-Driven Flood Hazard

This chapter has been published in Nature Communications with the following reference:

The roles of the other authors of this paper in relation to the project are as follows: H. L. Cloke (supervisor: academic) E. M. Stephens (supervisor: academic), E. Zsoter (collaborator: ECMWF), S. J. Woolnough (supervisor: academic), F. Pappenberger (collaborator: ECMWF). R.E. conceived and posed the research question, carried out the analysis, wrote the paper, prepared the figures and submitted the paper. The study design and interpretation of the results was done in collaboration with H.L.C., E.M.S., F.P. and S.J.W, and E.Z. created the ERA-20CM-R dataset used in the study. The manuscript was written by R.E., with guidance and advice from H.L.C., E.M.S and F.P., and all authors commented on the manuscript. Overall, 95% of the research and 85% of the writing was undertaken by R.E.

Abstract. El Niño and La Niña events, the extremes of ENSO climate variability, influence river flow and flooding at the global scale. Estimates of the historical probability of extreme (high or low) precipitation are used to provide vital information on the likelihood of adverse impacts during extreme ENSO events. However, the nonlinearity between precipitation and flood magnitude motivates the need for estimation of historical probabilities using analysis of hydrological datasets. Here, this analysis is undertaken using the ERA-20CM-R river flow reconstruction for the 20th Century. Our results show that the likelihood of increased or decreased flood hazard during ENSO events is much more complex than is often perceived and reported; probabilities vary greatly across the globe, with large uncertainties inherent in the data and clear differences when comparing the hydrological analysis to precipitation.

4.1 Introduction

The El Niño Southern Oscillation (ENSO) is the most prominent pattern of interannual climate variability (McPhaden et al., 2006), and is known to influence river flow (Chiew and McMahon, 2002) and flooding (Ward et al., 2014a, 2014b, 2016) at the global scale. In the absence of hydrological analyses, products indicating the likelihood of extreme precipitation are often used as an early indicator of flooding during extreme ENSO events (IRI, 2018). However, the nonlinearity between precipitation and flood magnitude and frequency (Stephens et al., 2015) means that it is important to assess the impact of ENSO not just on precipitation, but on river flow and flooding. This is especially important as, as stated by Chiew and McMahon (2002), "it is likely that the streamflow-ENSO relationship is stronger than the rainfall-ENSO relationship because the variability in rainfall is enhanced in runoff and because streamflow integrates information spatially".

Here, a global scale hydrological analysis is performed to estimate the historical probability of increased or decreased flood hazard in any given month during El Niño / La Niña events, assessing the added benefit of directly analysing river flow over the use of precipitation as a proxy for flood hazard.

Historical probabilities provide useful information about typical ENSO impacts based on historical evidence (Bradley et al., 1987; Mason and Goddard, 2001) and are, as stated by Mason and Goddard (2001), "a better estimate of the future climate than the assumption that seasonal conditions will be the same as average". Nonetheless, there are some key considerations when using such information. One such consideration is that no two El Niño events are the same (Davey et al., 2014; Mason and Goddard, 2001); differences in the peak amplitude, temporal evolution and spatial pattern of warming are likely to affect the timing and magnitude of the resulting impact on river flow. There are many suggested ways to classify ENSO diversity (Capotondi et al., 2015), for example, El Niño events are often described as 'East Pacific' (EP) or 'Central Pacific' (CP), dependent on where the peak warming occurs. While this is an over-simplification of the complexity surrounding ENSO diversity, the location of the peak warming can alter the influence on river flow. An additional consideration is the influence of warming ocean temperatures on ENSO events and their related impacts. Recent studies (Cai et al., 2014, 2015a) suggest that projected changes in the Walker Circulation and associated weakening of equatorial Pacific ocean currents are expected to result in more frequent, and more extreme, El Niño and La Niña events (Cai et al., 2015a, 2015b).

In the past, studies have been limited to reanalysis datasets of no longer than ~40 years (Ward et al., 2014a, 2014b, 2016), in which there is a sample of ≤ 10 El Niño and ≤ 13 La Niña events, or observational data with inconsistent coverage, both spatially and temporally (Chiew and McMahon, 2002). We have created a 20th Century (1901-2010) model reconstruction of river flow in order to obtain a hydrological dataset with consistent global coverage over an extended time period. Research by Essou et al. (2016) indicates that global meteorological reanalysis datasets "have good potential to be used as proxies to observations" in order to force hydrological models, particularly in regions where few observations are available. This dataset was created by forcing a research version (described in section 4.4.1) of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013; Emerton et al., 2016) with the ERA-20CM (Hersbach et al., 2015) meteorological model reconstruction of the European Centre for Medium-Range Weather Forecasts (ECMWF) to produce a 10-member, 0.5° resolution reconstruction of river flow (from here on, ERA-20CM-R) containing 259,200 grid points covering the global river network (Supplementary Figure 1). Figure 1 depicts a time series of three key variables used in this study, alongside the timing of the 30 El Niño and 33 La Niña events identified in ERA-20CM-R (see section 4.4.2).
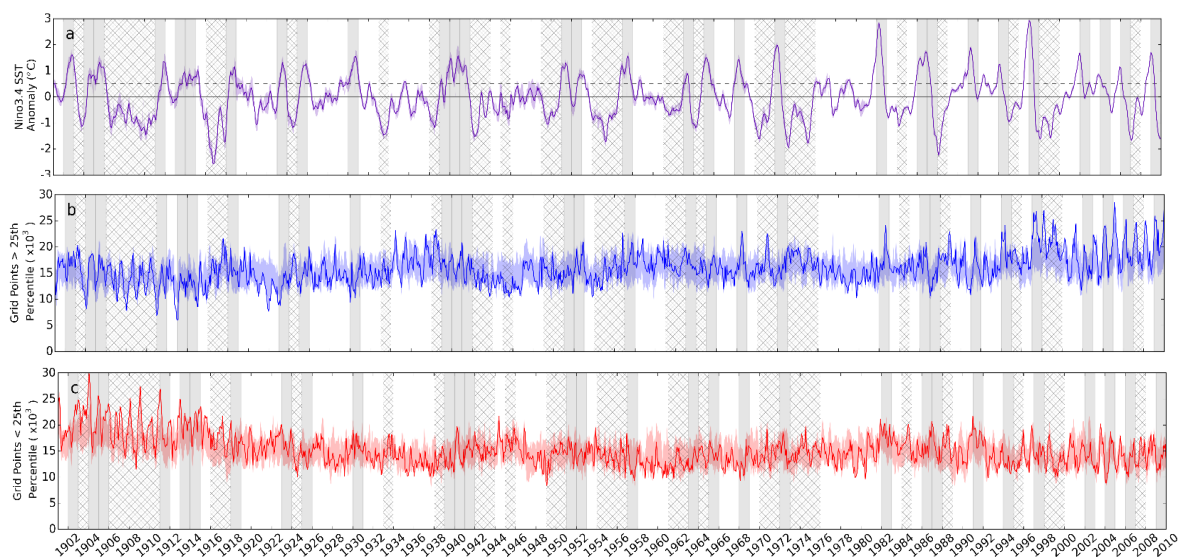
Figure 1: Time series of three key ERA-20CM-R variables and timing of El Niño and La Niña events. **(a)** 3-month running mean sea surface temperature anomaly in the Niño3.4 region (SSTA3.4), and number of grid points globally in which monthly mean river flow **(b)** exceeds the top 25th percentile and **(c)** falls below the lower 25th percentile. Solid lines show the mean of the 10 ensemble members, while shading indicates the spread of the members. The SSTA3.4 is used to identify El Niño and La Niña years in the dataset, highlighted here by the grey shaded and hatched bars, respectively.

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

Previous work by Ward et al. (2014b) has looked at the influence of El Niño on flood return periods, quantifying the percentage anomaly during El Niño years in comparison with climatology (defined as the long-term average of historical river conditions or meteorological parameters). To ensure accurate estimation of historical probabilities of ENSO-driven flood hazard, this analysis was replicated using the new ERA-20CM-R dataset and gives similar results (Supplementary Figure 2).

In this study, using a climatology of all years and all El Niño / La Niña years, we calculate the percentage of past El Niño / La Niña events during which the river flow fell in the upper [lower] quartile of climatology, defined here as "abnormally high [low] flow". Our results show that the likelihood of increased or decreased flood hazard during ENSO events is much more complex than is often perceived and reported; probabilities vary greatly across the globe, with large uncertainties inherent in the data and clear differences when comparing the hydrological analysis to precipitation.

4.2 Results

4.2.1 Historical Probabilities During El Niño

Figure 2a shows the historical probabilities for February during an El Niño, with the full set of El Niño and La Niña results presented in Supplementary Figures 7 and 8 respectively. El Niño events tend to span two calendar years, evolving in boreal spring and reaching their peak magnitude in winter of the same year, before decaying into the following spring/summer. Shortly after the peak, February sees some of the highest probabilities and extensive spatial coverage of regions influenced by El Niño (where >40% probability of abnormally high or low river flow represents a significant influence); 34.5% of the land surface indicates a significant increase in the probability of abnormally high or low river flow (19.2% for high, 15.3% for low) compared to any given year.

The influence of El Niño on river flow can be seen as early as June (see Supplementary Figures 7 and 8), shortly after ENSO tends to move into the warm phase, with some regions, mostly confined to the tropics, beginning to see up to a 50% probability of high or low river flow in the ensemble mean. In August and September, much of South America, south of the Amazon River, is somewhat likely (~40-60% probability) to observe higher than normal river flow however, in November, closer to the typical peak of El Niño events, a reversal to drier conditions across much of Brazil is observed. The southern USA has a high probability (up to

Figure 2: **Historical probability of increased or decreased flood hazard during one month of an El Niño. (a)** Probability of abnormally high (blue) or low (red) monthly mean river discharge. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology. **(b)** Uncertainty around the probability shown in (a), i.e. the difference between the minimum and maximum of the 10 ensemble members (%). The boxplot (b, inset) gives an example graphical representation of the uncertainty range at one grid point, marked on the map by an 'x', where the mean probability indicated in (a) is 63%. The range is given by the difference between the minimum and maximum of the 10 ensemble members; in this case 53% and 81%, giving a 28% range falling in the 20-40% bracket in (b). The month of February is chosen as, occurring shortly after the peak of an El Niño, it sees extensive spatial coverage of land areas influenced by El Niño.

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

70%) of high river flow from December onwards, while Mexico is another region that experiences a reversal in the influence of El Niño, from decreased flood hazard up until September/October, to increased flood hazard from November onwards. Other regions are much more consistent, such as Indonesia, which has a high certainty of abnormally low river flow throughout the evolution, peak and decay of El Niño. However, it is important to note that across the globe, the uncertainty around these probabilities can be high.

4.2.2 Evaluating the Uncertainty

Indeed, the historical probabilities themselves give an indication of the uncertainty in the response of the river flow to ENSO events. Here, the 10 ensemble members of ERA-20CM-R also allow interpretation of the uncertainty in the dataset, as each ensemble member represents an equally probable reconstruction of the river flow. In order to provide an indication of this uncertainty, Figure 2b shows the range of the probability around the mean probability shown in Figure 2a. The influence of El Niño is much more certain in some locations; for example, in coastal Ecuador/northern Peru, the probabilities vary by only 9%. These locations (darkest shading, 5-10% range) stand out in Figure 2b; these are the areas where there is potential to use such historical probabilities as an early indicator of increased or decreased flood hazard, as they tend to give high probabilities combined with small uncertainties. However, much of the globe shows a range of 20-40%, and some small regions, such as in northwest Spain and eastern Argentina, see a range up to 70% across the ensemble members. The implication is that while some regions see high probabilities of increased flood hazard, (e.g. up to 77% in northern Peru), across much of the globe the likelihood is much lower and much more uncertain than might be useful for decision-making purposes.

4.2.3 Importance of the Hydrology

Evaluating the historical probabilities of abnormally high or low precipitation, using the ERA-20CM precipitation dataset, confirms that there is additional information which can be gained from the hydrological analysis. For example, parts of northern Africa are likely to see high precipitation in February (Supplementary Figure 3a); however, the River Nile is likely to see dry river conditions (Figure 2a), indicating that the river is influenced more by upstream rather than local precipitation.

To further highlight the importance of considering the hydrological impacts, Figure 3 indicates regions, shown in pink [green], where the probability of high river flow is greater [smaller] than that of high precipitation. These differences suggest that the influence of El Niño is more

pronounced in the river flow in pink regions, and conversely, green highlights regions where the use of precipitation as a proxy for flood hazard results in an over-estimation of the probabilities. This could also indicate that the region is likely to experience a lagged influence of El Niño on river flow. The corresponding results for low flow are presented in Supplementary Figure 4.
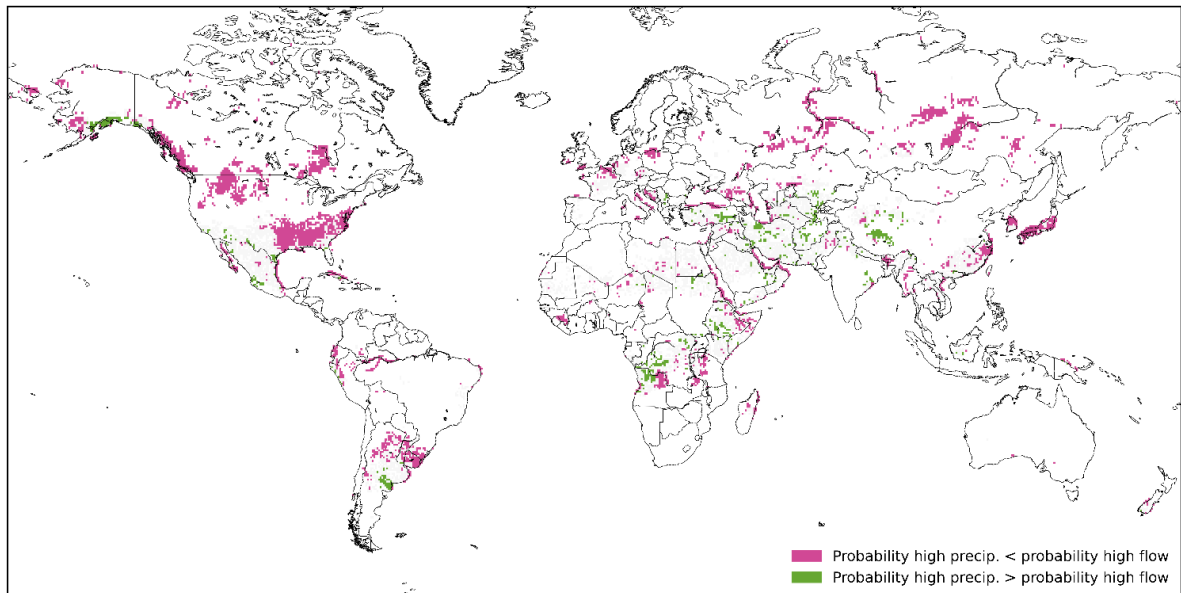
Figure 3: Comparison of historical probabilities based on precipitation and river flow. Regions where the difference in probability of abnormally high precipitation compared to probability of high river flow, in the month of February during an El Niño, is greater than 10% (based on the ensemble mean). Pink shading indicates that the probability of high precipitation is smaller than the probability of high river flow, while green shading indicates that probabilities are larger for precipitation.

4.2.4 Historical Probabilities During La Niña

El Niño events are often followed by a La Niña, the cool phase of ENSO. While La Niña events tend to be less widely discussed in the media, their influence on precipitation is often used as a proxy for flood hazard, as with El Niño. We have therefore extended this analysis to evaluate the probability of increased (or decreased) flood hazard during La Niña years. We find that many regions influenced by El Niño are likely to observe the opposite response during La Niña. Figure 4 shows these probabilities, again for February, during a La Niña event, with the full set of results shown in Supplementary Figure 8. It is evident that less of the land surface is significantly influenced by La Niña compared to El Niño during this month (22% of the land surface compared to 34.5%). Probabilities, while still significant, also tend to be lower than for
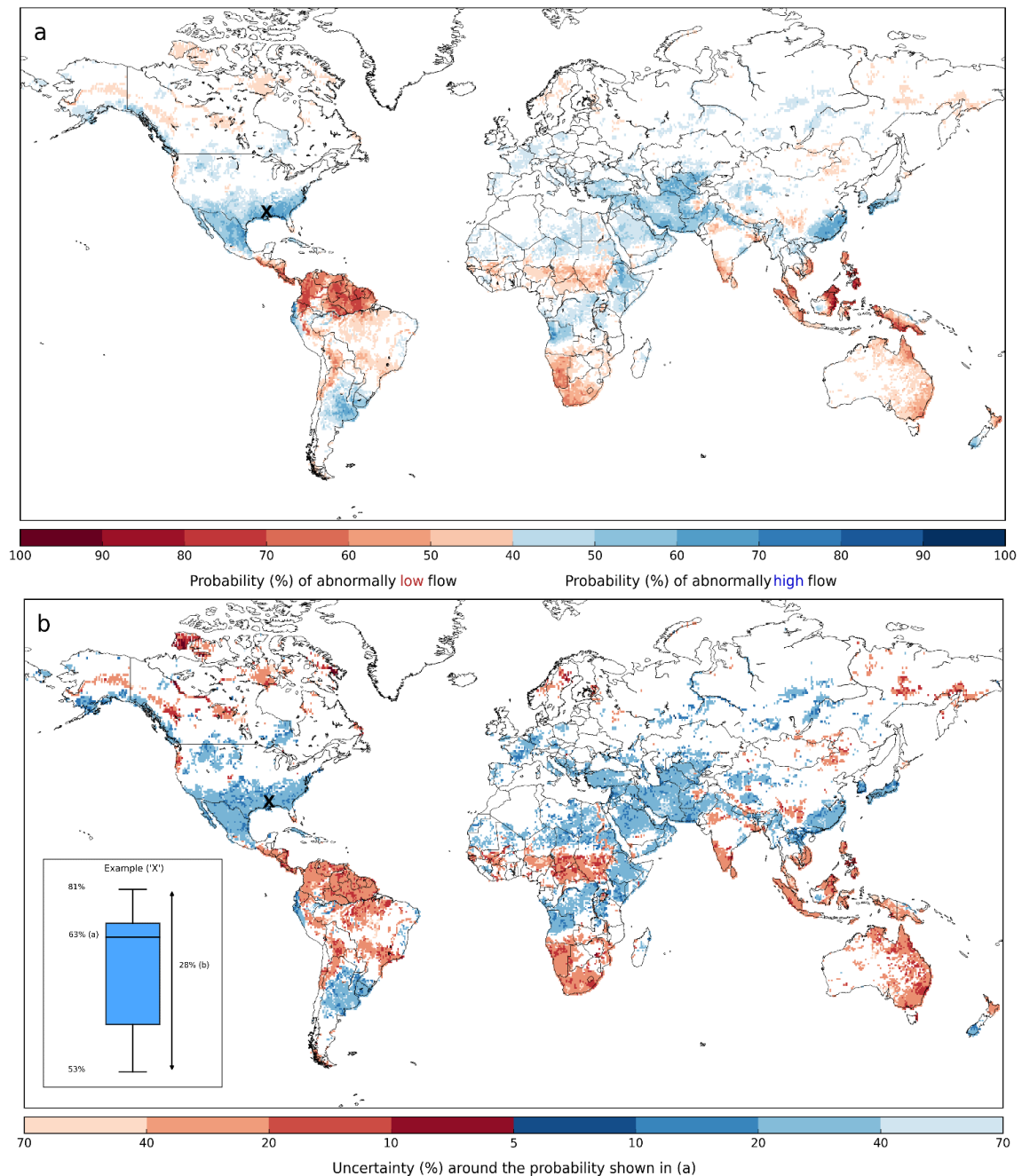
Figure 4: **Historical probability of increased or decreased flood hazard during one month of a La Niña. (a)** Probability of abnormally high (purple) or low (orange) monthly mean river discharge in the month of February during a La Niña. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology. **(b)** Uncertainty around the probability shown in (a), i.e. the difference between the maximum and minimum of the 10 ensemble members (%).

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

the same month during an El Niño; the highest probability of increased flood hazard shown in Figure 4a is 67%, and 69% for decreased flood hazard. Again, the uncertainty surrounding this mean probability is large (20-40% and in some areas >70%) across much of the globe; this can be seen in Figure 4b.

4.2.5 Maximum Probabilities During El Niño / La Niña

While the monthly maps of historical probabilities give an indicator of the probability of increased (or decreased) flood hazard and when this is likely to occur, it is perhaps useful to consider the event as a whole, as the peak conditions occur at different times across the globe.

Figure 5a [5b] shows the maximum probability of increased flood hazard during any month of an El Niño [La Niña] event; this provides an overview of whether a region is likely to experience a change in river conditions or not during or following the event. Figure 5 also indicates where the uncertainty surrounding the probabilities is high; this tends to be where the probability is lower, while regions with high probabilities also indicate higher certainty. This analysis further confirms that across much of the globe, such historical probabilities are much more uncertain than is often communicated. The corresponding results for decreased flood hazard are shown in Supplementary Figure 5.

4.2.6 Comparison with Observations

A comparison of the historical probabilities against observed datasets was also undertaken (see sections 4.4.5 - 4.4.6, and Supplementary Figure 6). While this proved challenging at the global scale due to a lack of consistent and extensive river flow records in regions of the world where ENSO events have the most influence, the evaluation suggests a potential over-estimation of the probabilities in both the precipitation and river flow reconstructions. This stresses that while these model reconstructions are currently the best available data for such research, there is a need for more extensive river flow observations in regions impacted by ENSO events.

Throughout the results, the complexity and uncertainty surrounding such historical probabilities is evident. Indeed, observations of flooding in February 2016, during the strong 2015-16 El Niño event, reflect this complex picture of ENSO-driven flood hazard. The expected flooding (based on the results shown in Figure 2a) in Peru, Bolivia, Argentina and Angola was observed (FloodList, 2018); yet in several other regions, such as Eastern China, Japan and parts of the Middle East, no flood events were recorded. Flooding also occurred in Indonesia despite a high likelihood of dry river conditions. In Kenya and Peru, two examples where flood preparedness

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

actions were taken ahead of El Niño, flooding was much less severe than expected (Muchangi, 2016; Red Cross Red Crescent Climate Centre, 2015). A recent Nature correspondence (Cohen, 2016) also highlighted the unexpected winter weather in the USA; California experienced heatwaves rather than prolonged rain events, while Seattle was expecting a worsening drought and instead endured the wettest winter on record (see also Supplementary Figure 7).
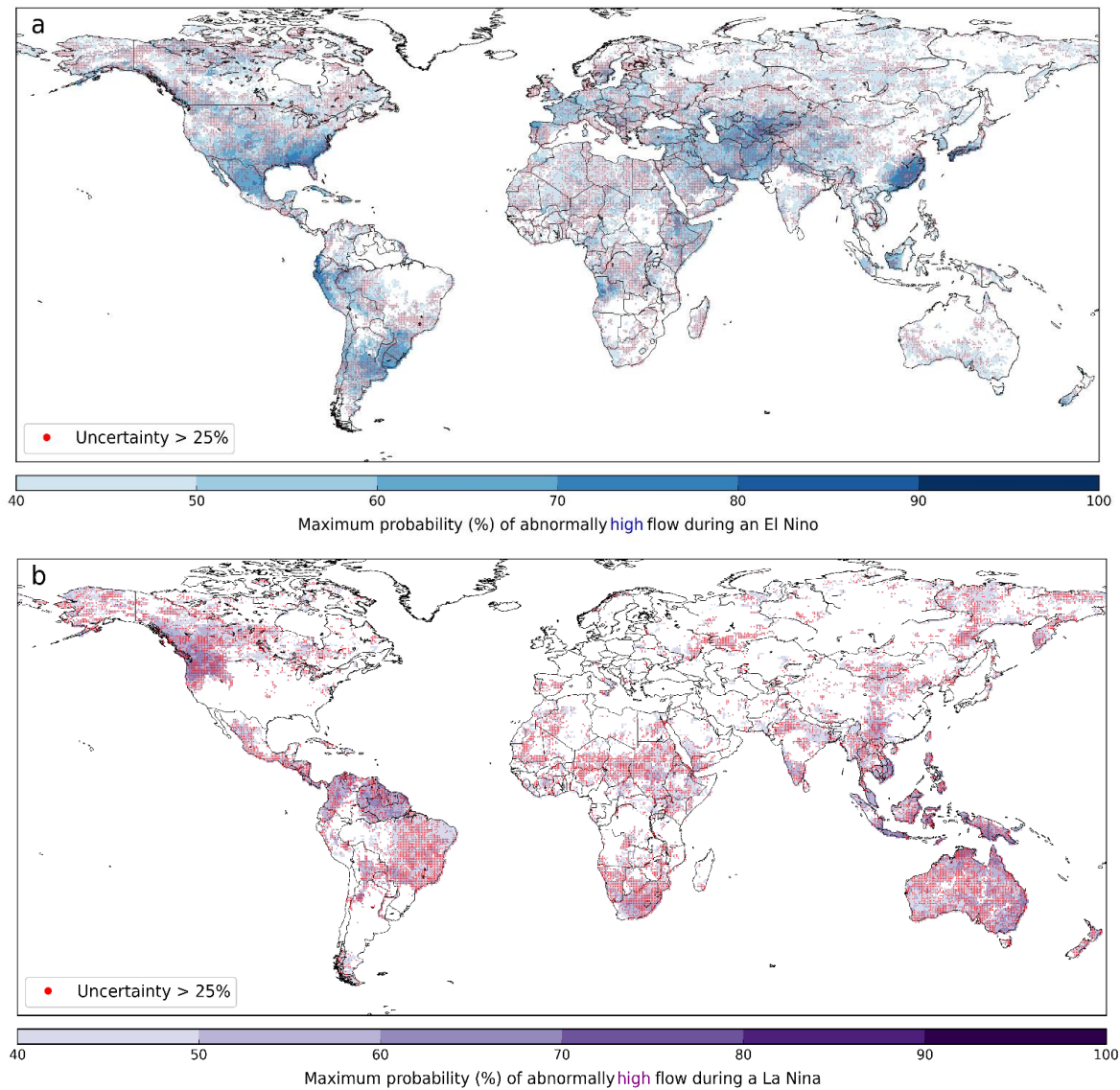
Figure 5: Maximum probability of abnormally high river flow in any month during (a) an El Niño event and (b) a La Niña event. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology during, or shortly after the decay of, an ENSO event. Stippling indicates where the uncertainty surrounding this probability is high, i.e. the range of the ensemble members exceeds 25% probability.

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

4.3 Conclusions

We have conducted a global hydrological analysis of ENSO as a predictor of flood hazard based on historical probability estimates using a new, extended-length model reconstruction of river flow. The importance of addressing the hydrology in addition to precipitation is evident in the differences between the probabilities of high river flow and precipitation, and in the ability to further evaluate areas likely to see a lagged influence of El Niño on river flow. We conclude that while it may seem possible to use historical probabilities to evaluate regions across the globe that are more likely to be at risk of flooding during an El Niño / La Niña, and indeed circle large areas of the globe under one banner of wetter or drier, the reality is much more complex. It is therefore important to undertake research that focusses on the region(s) of interest and consider the impact of ENSO diversity and other drivers of climate variability on the hydrology and flood hazard.

4.4 Methods

4.4.1 The New 20th Century River Flow Dataset

For this study, we have created a 20th Century (1901-2010) reconstruction of river discharge, in order to obtain a dataset with consistent global coverage over an extended time period. This was achieved by forcing an alternative setup of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013; Emerton et al., 2016) with the 10 ensemble members of the ERA-20CM (Hersbach et al., 2015) atmospheric model ensemble of the European Centre for Medium-Range Weather Forecasts (ECMWF) to produce a 10-member ensemble of river discharge for the global river network (ERA-20CM-R).

The operational set-up of GloFAS takes the runoff output from the ECMWF Integrated Forecast System (IFS) and runs this through the Lisflood hydrological routing model (Alfieri et al., 2013). Here, we instead use the Catchment-based Macro-scale Floodplain (CaMa-Flood; Yamazaki et al., 2011) routing model to create the river discharge reconstruction at 0.5° resolution from the gridded ERA-20CM runoff data. A map of the CaMa-Flood global river network is given in Supplementary Figure 1. We note here that the version of GloFAS used in this study is uncalibrated.

While the use of the ERA-20CM model reconstruction allows a consistent analysis at the global scale, and provides a much longer time period over which to study these extreme events, there are limitations that must be considered. ERA-20CM incorporates ENSO and 20th century

climate trends, and assimilates sea-surface temperature and sea ice cover (Hersbach et al., 2015). It does not, however, assimilate atmospheric observations. This is a drawback as the model reconstruction is able to provide a statistical estimate of the climate, but is not able to reproduce synoptic situations. We have therefore undertaken a comparison with the best available precipitation and river discharge observations for the 20th Century and are satisfied that ENSO teleconnections are well-represented in ERA-20CM(-R). Of course, there is further uncertainty introduced when going back as far as the early 20th Century when fewer observations were available; the 10 ensemble members go some way to representing this uncertainty and are a key benefit of this particular dataset, and thus are considered throughout this study.

4.4.2 Identifying the El Niño years

In order to conduct this analysis, we first identified the El Niño / La Niña years in the dataset. This was done using the definition that the U.S. National Oceanic and Atmospheric Administration (NOAA) use to declare El Niño [La Niña] conditions operationally (NOAA, 2016b). This definition states that the sea surface temperature (SST) anomaly must remain ≥ 0.5°C [≤ 0.5°C], in the Niño3.4 region in the central Pacific (5°S - 5°N, 170°W - 120°W), for at least five consecutive three-month periods. Here, we extracted the ERA-20CM SST data and calculated the three-month running mean SST anomalies for the Niño3.4 region, allowing identification of the 30 [33] years in which El Niño[La Niña] conditions were present from 1901 to 2010. These are listed in Supplementary Table 1, where the El Niño / La Niña year refers to the year in which the event evolves and typically also reaches its peak, as ENSO events often span two years, decaying into the following year. We note that while there is generally a good agreement between the ENSO events identified in ERA-20CM and those published by NOAA (NOAA, 2016a) for the same period, there are, however, some discrepancies. This is likely due to the different indices / definitions used to identify the ENSO events. For example, in 1977 and 1979, El Niño events are identified by NOAA, using the Multivariate ENSO Index (NOAA, 2016a), but these are not picked up in this study. In Figure 1, it is evident that the SST did exceed 0.5 °C in ERA-20CM, but this did not persist for long enough to be identified as an event. This is a limitation of the need to use one of the many varying methods of classifying and identifying ENSO events. This method was chosen as it is the most operationally relevant at the time of writing.

4.4.3 Historical Probability Estimation

For the results presented in this study, the 110-year ERA-20CM-R climatology was used to calculate the upper and lower 25th, 10th and 5th percentiles of river discharge for every grid box. The historical probability of abnormally high or low river flow in any given month was then estimated, through calculation of the percentage of the 30 [33] identified El Niño [La Niña] years in which the river discharge exceeded (high flow) or fell below (low flow) the three percentile thresholds, for each of the 10 ensemble members of ERA-20CM-R. The analysis presented in this paper is based on percentiles so as to avoid potential large errors caused by bias in the dataset compared to observations (discussed further below).

Maps of the resulting probabilities were produced based on the mean of the 10 ensemble members. As the number of ENSO events cover a substantial part of the 110-year period, there is a chance of picking up random effects. The maps produced therefore only display results where the probability is significantly greater than normal, i.e. ≥40%; an "event" (occurrence of abnormally high or low flow) with a probability of 40% during one month of an El Niño / La Niña has only a 5% chance of occurring by chance in that month, and thus represents a significant increase in the probability compared to the likelihood of occurring at random.

Additionally, the spread in the ensemble members is designed to reflect the uncertainty in the dataset, and can indicate a range of possible outcomes or probabilities. As such, we have further calculated the uncertainty around the mean probability for the whole globe, based on the range across the ensemble members. For each ensemble member, the range between the minimum and maximum ensemble members was calculated for every grid box individually. This allows us to interpret the uncertainty in the probability caused by uncertainty in the dataset.

El Niño / La Niña onset tends to occur in boreal spring/early summer and peak in winter (Trenberth, 1997), before decaying into the following spring. As such, the monthly analysis was undertaken for a period of two years; the year of onset, and the following year during which the El Niño / La Niña decays, in order to capture any lagged influence on river flow. Significant influence is shown in the results from June during the El Niño / La Niña year, to the following September (16 months). While it would seem advantageous to summarise the findings by season for simplicity, evaluation of the results shows that the patterns of influence across the globe can change dramatically, in some instances, from one month to the next. Summarising these maps into seasons may therefore result in a loss of information for some months.

4.4.4 Difference Between River Flow and Precipitation

A key aim of this paper was to evaluate the added benefit of the hydrological analysis over the use of precipitation as a proxy for flood hazard. To do this, the same method used to estimate the historical probabilities in the river flow reconstruction (ERA-20CM-R) was also applied to the ERA-20CM precipitation reconstruction. The horizontal resolution of the ERA-20CM precipitation data is ~125km, while the river flow data is at 0.5° (~55km) resolution. In order to compare these, the results from the precipitation data were remapped to the higher resolution of the river flow data using a simple nearest neighbour remapping algorithm. The difference between the historical precipitation probabilities and river flow probabilities was then calculated for the mean of the 10 ensemble members.

4.4.5 Comparison with Observations – Precipitation

In order to evaluate the results shown using the new ERA-20CM(-R) dataset, the same method for estimating historical probabilities was also applied to other, related datasets; the Global Precipitation Climatology Centre (GPCC) Full Data reanalysis (GPCC-FD; Schneider et al., 2015) at 0.5° resolution, and the Global Runoff Data Centre (GRDC) river discharge observations (BfG, 2017). Again, percentiles are used throughout to allow reliable comparison with observations despite potentially large bias in the model reconstruction values compared to observed values.

The GPCC-FD reanalysis is a global gridded precipitation dataset based on interpolated rain gauge data (Schneider et al., 2015). Comparing the ERA-20CM and GPCC-FD precipitation datasets indicates that the regions influenced by El Niño are well-represented by ERA-20CM (see Supplementary Figure 3b), and in line with well-known ENSO-sensitive regions, such as Australia, Indonesia, Argentina (the Rio de la Plata delta) and the southern USA – which have been shown to be well-represented in the GPCC-FD (Becker et al., 2013). However, the strength of this link appears to be over-estimated compared to observations, as the ERA-20CM data shows higher probabilities of abnormally high or low precipitation than the GPCC-FD. Some of this over-estimation may be caused by the use of the ensemble mean to produce the ERA-20CM maps, as averaging across the 10 ensemble members likely results in a reduction of the variance and we therefore pick up the forced part of the signal.

4.4.6 Comparison with Observations – River Discharge

As no gridded observational dataset of river discharge exists for the global river network, archived station data from the GRDC were used. Criteria for data suitability were chosen to

identify those stations which could be of use in this study. Firstly, only stations with at least a 75-year record of observations between 1901 and 2010 were included; these could be stations recording on a daily or monthly basis. Of these, any stations with more than 50% of the data missing were removed. In total, 1287 stations fit the criteria (232 monthly, 1055 daily), of which the majority have <30% of the data missing. Each of these stations were manually checked to ensure that they correspond to the correct river point (taking into account location and upstream area) on the model river network. A key limitation of using the GRDC observations for this study is that many of these stations lie in river basins outside of the tropics and subtropics - the regions which tend to be most strongly influenced by ENSO events. This highlights the need for more consistent global river flow observations, but in their absence, model reconstructions and reanalyses present the best available data for regional and global scale research based on historical evidence.

In order to compare the results based on observations with ERA-20CM-R, we produced a reliability diagram (Supplementary Figure 5) for the historical probability of abnormally high river flow, comparing the forecast (historical) probability of an event (in this case, river flow exceeding a given percentile) with the observed frequency of the event. This was achieved by first locating all grid points in the ERA-20CM-R dataset that contain a GRDC station that fit the criteria outlined above. For each percentage band (in 10% bins, as displayed on the maps shown in the results) of the "forecast", the observed frequency of river flow exceeding the upper 25th, 10th and 5th percentiles of the 110-year climatology was calculated for each GRDC station, before taking the mean across all stations, and all 16 months used in the analysis (June to the following September). This allows comparison of the predicted probability with the observed frequency. The reliability diagram (Supplementary Figure 5) and the discrepancy between forecasted and realised probabilities indicates that there is a potential over-estimation of the forecasted probabilities. There are limitations, however, in that we have very few, or no, observation stations with which to compare the results for the higher probabilities (Supplementary Figure 5, inset), particularly in regions that are most significantly influenced by El Niño / La Niña and where reliability may be better, such as the tropics. This suggests that such a reliability analysis may not be fully representative of the results. Additionally, the data records vary from station to station, therefore the number of El Niño / La Niña years included in the observational record of each station also varies.

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard

4.5 Supplementary Figures

Table S1: El Niño and La Niña years identified in the ERA-20CM SST data between 1901 and 2010.

ERA-20CM El Niño Years			ERA-20CM La Niña Years		
1902	1939	1982	1903	1942	1970
1904	1940	1986	1906	1943	1971
1905	1941	1987	1907	1945	1973
1911	1951	1991	1908	1949	1974
1913	1952	1994	1909	1950	1975
1914	1957	1997	1910	1954	1984
1918	1963	2002	1916	1955	1988
1923	1965	2004	1917	1956	1995
1925	1968	2006	1924	1961	1998
1930	1972	2009	1933	1962	1999
			1938	1964	2007
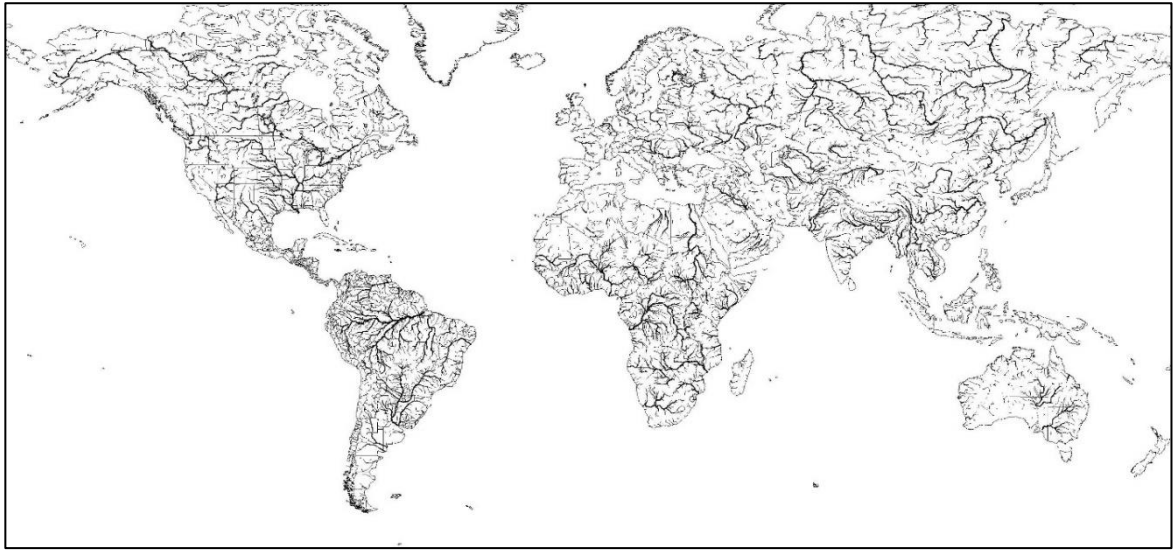
Figure S1: The CaMaFlood 0.5º global river network used in this study.
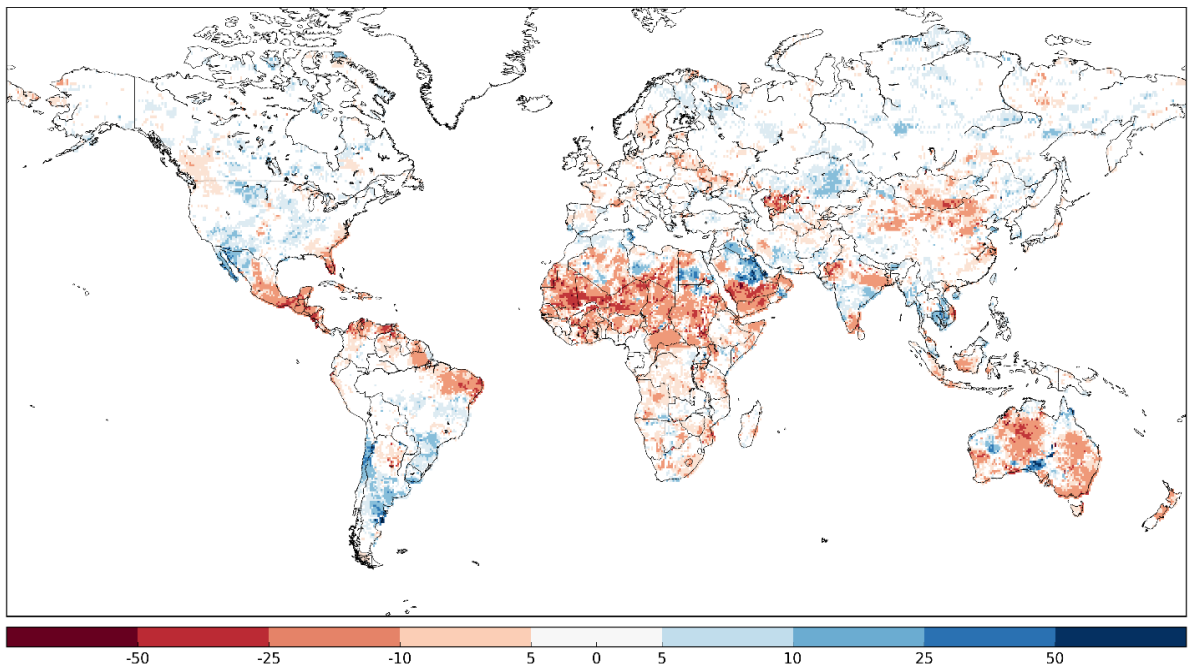
Figure S2: Percentage anomaly in the 100-year flood return period during El Niño. This replicates the analysis of Ward et al. (2014b) in order to ensure accurate estimation of the historical probabilities of ENSO-driven flood hazard using ERA-20CM-R.
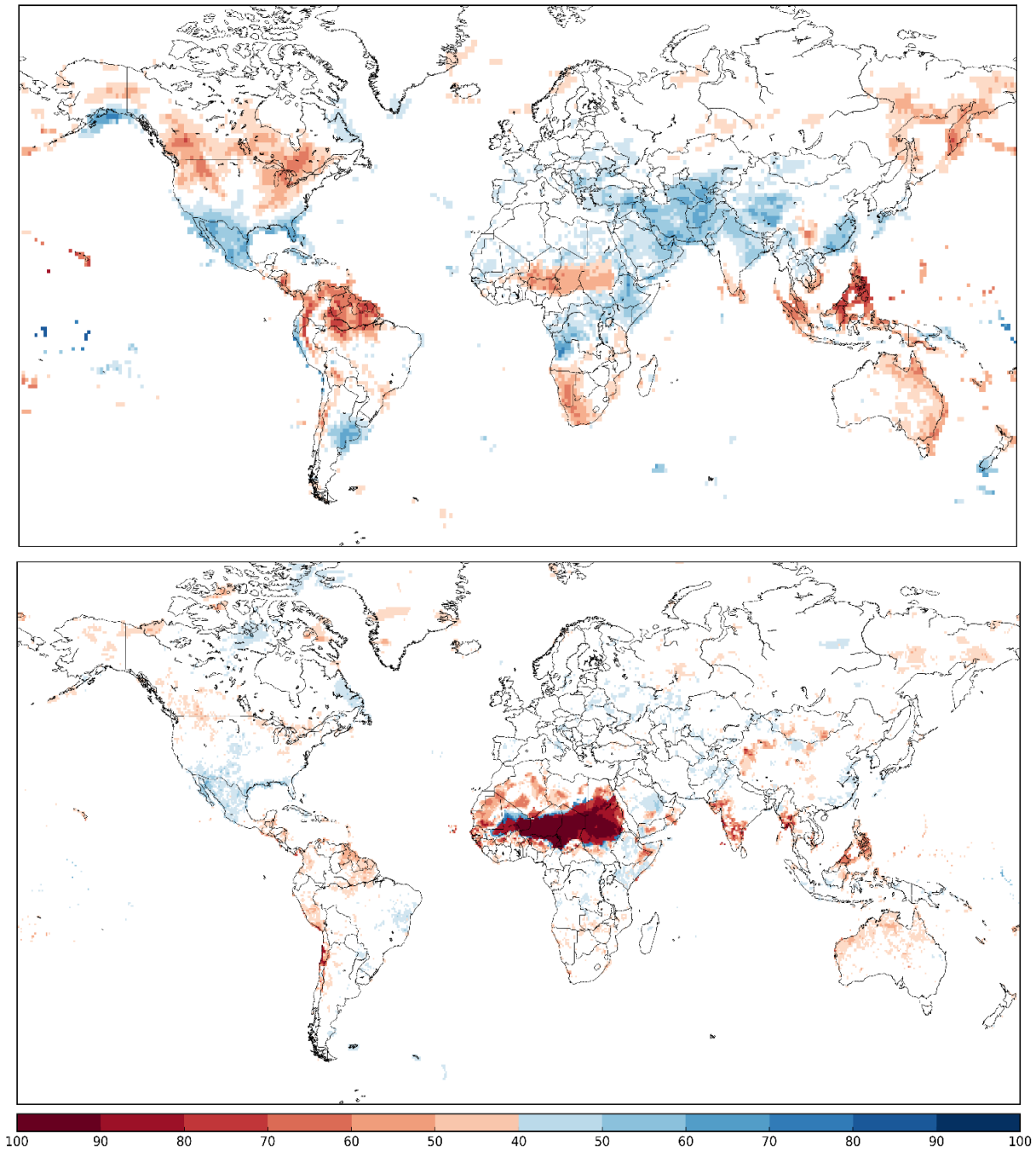
Figure S3: Probability of abnormally high (blue) or low (red) total monthly precipitation during the month of February during an El Niño, based on total monthly precipitation exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year (1901-2010) climatology. Using **(a)** the ERA-20CM dataset (based on the mean of the 10 ensemble members) and **(b)** the GPCC-FD gridded precipitation dataset based on interpolated gauge observations. The large area of 100% probability (red) across northern Africa in (b) is most likely a result of the interpolation used to produce the GPCC-FD dataset in a region with few available observations.
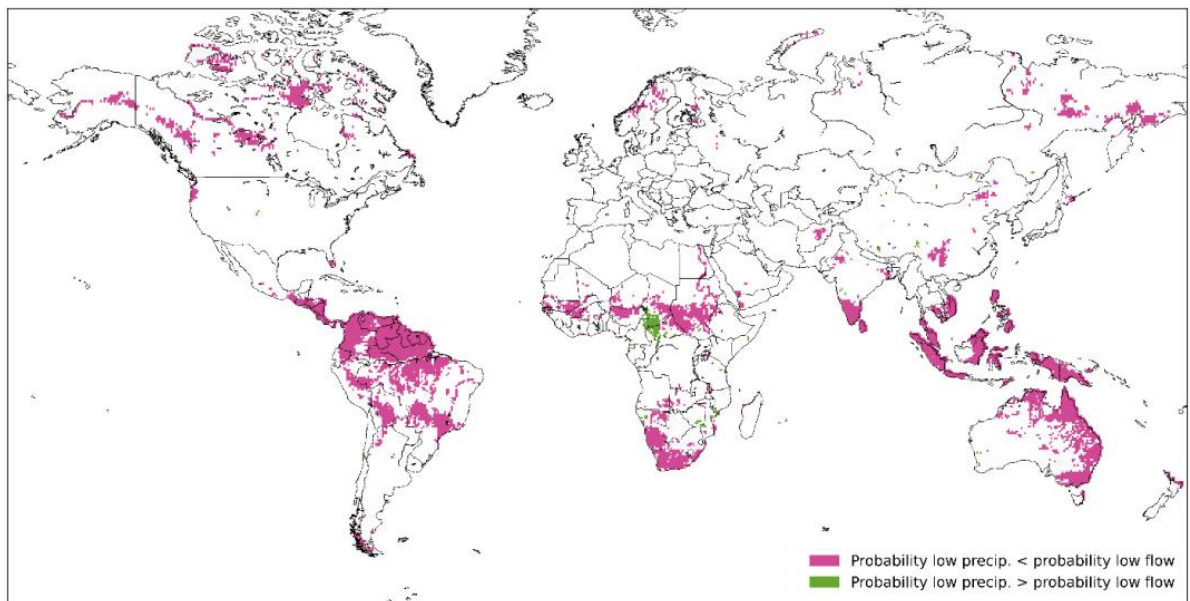
Figure S4: Regions where the difference in probability of abnormally low precipitation compared to probability of low river flow, in the month of February during an El Niño, is greater than 10% (based on the ensemble mean). Negative values (pink) indicate that the probability of low precipitation is smaller than the probability of low river flow, while positive values (green) indicate that probabilities are larger for precipitation.

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard
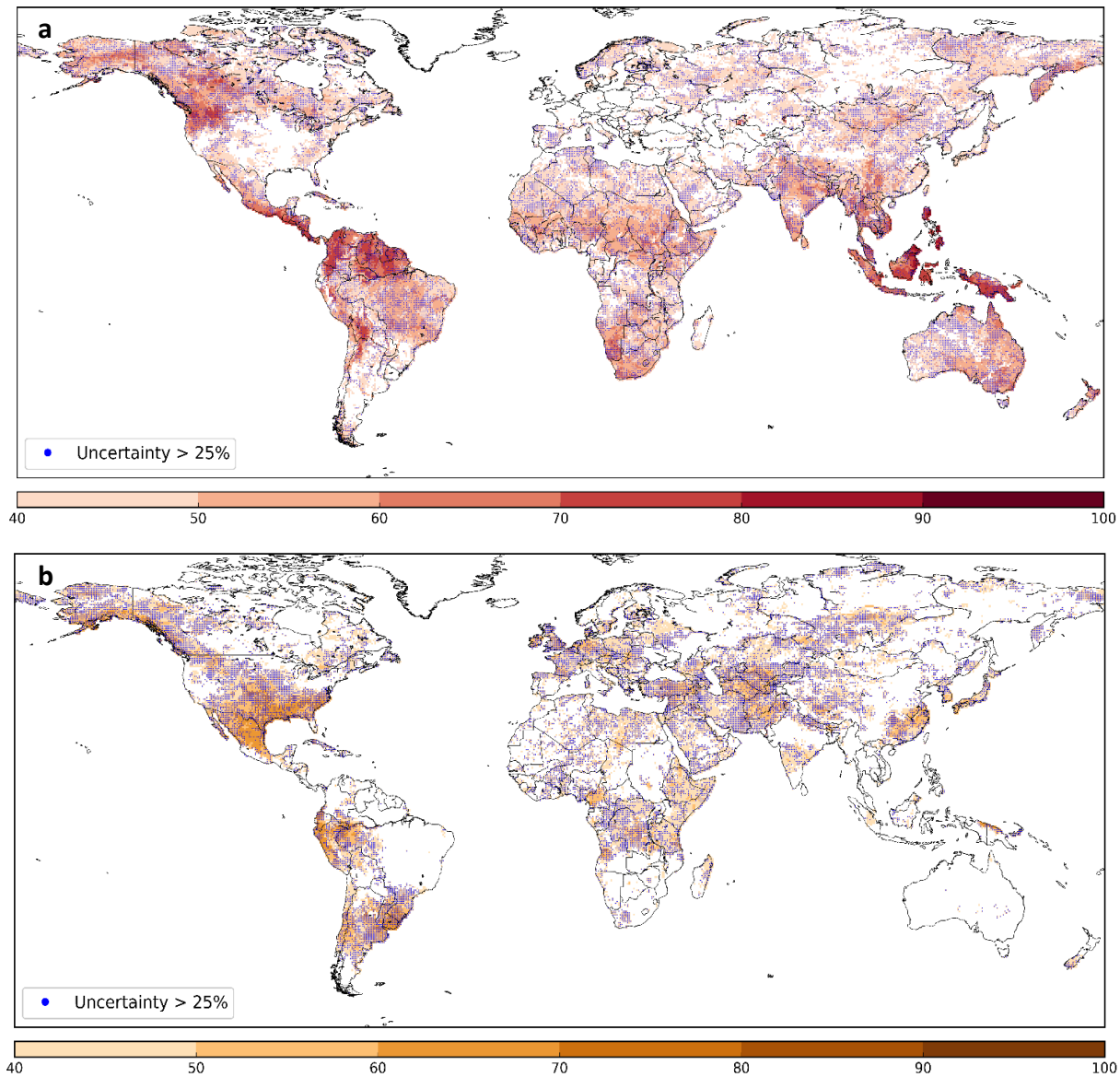
Figure S5: Maximum probability of abnormally low river flow in any month during **(a)** an El Niño event and **(b)** a La Niña event. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology during, or shortly after the decay of, an ENSO event. Stippling indicates where the uncertainty surrounding this probability is high, i.e. the range of the ensemble members exceeds 25% probability.
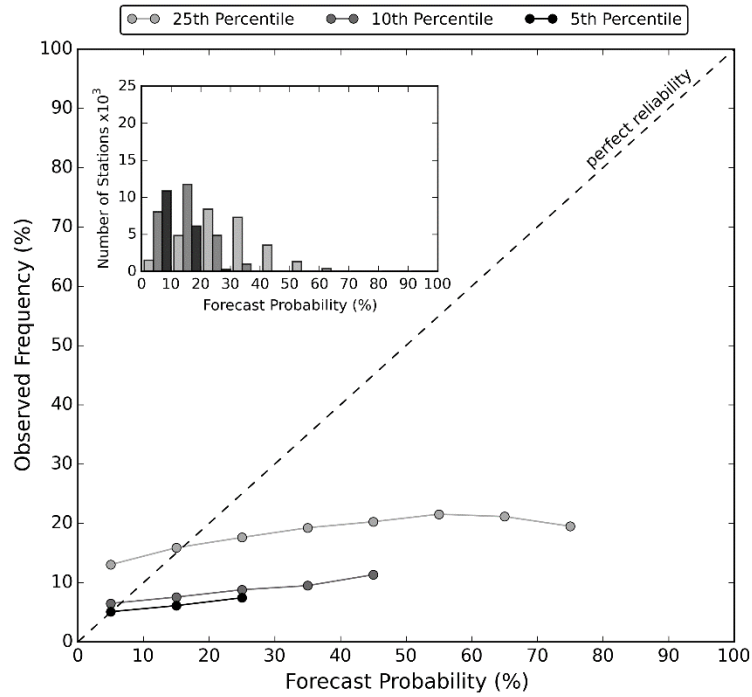
Figure S6: Reliability diagram comparing the forecast probability of abnormally high flow to the observed frequency in the GRDC observations. Results are included for exceedance of three river flow thresholds; the upper 25th, 10th and 5th percentiles. The results shown are an average across the 16 months from June during the El Niño year to the September following. Also shown is the number of available GRDC observation stations in each percentage band.
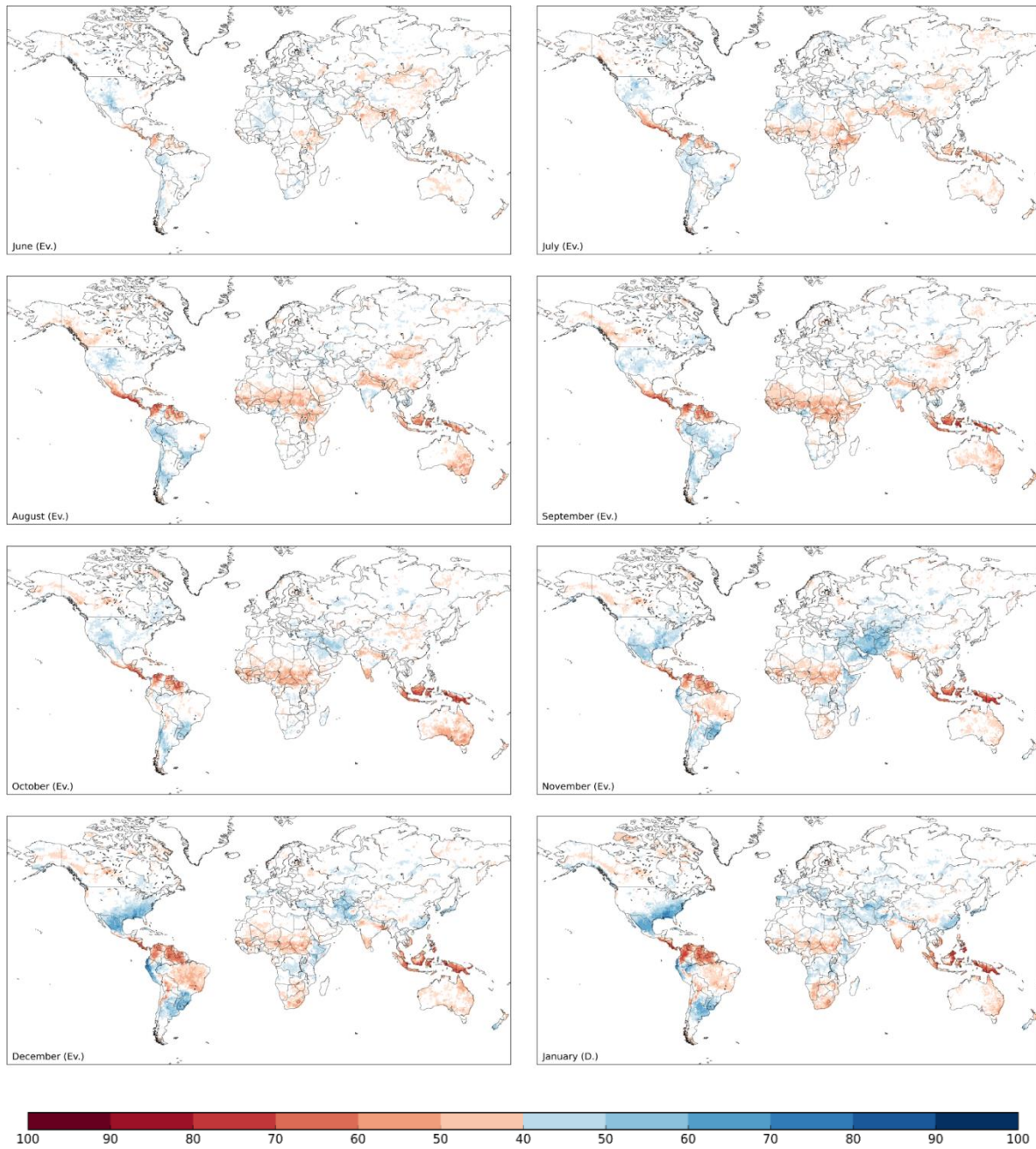
Figure S7 (continued on next page)
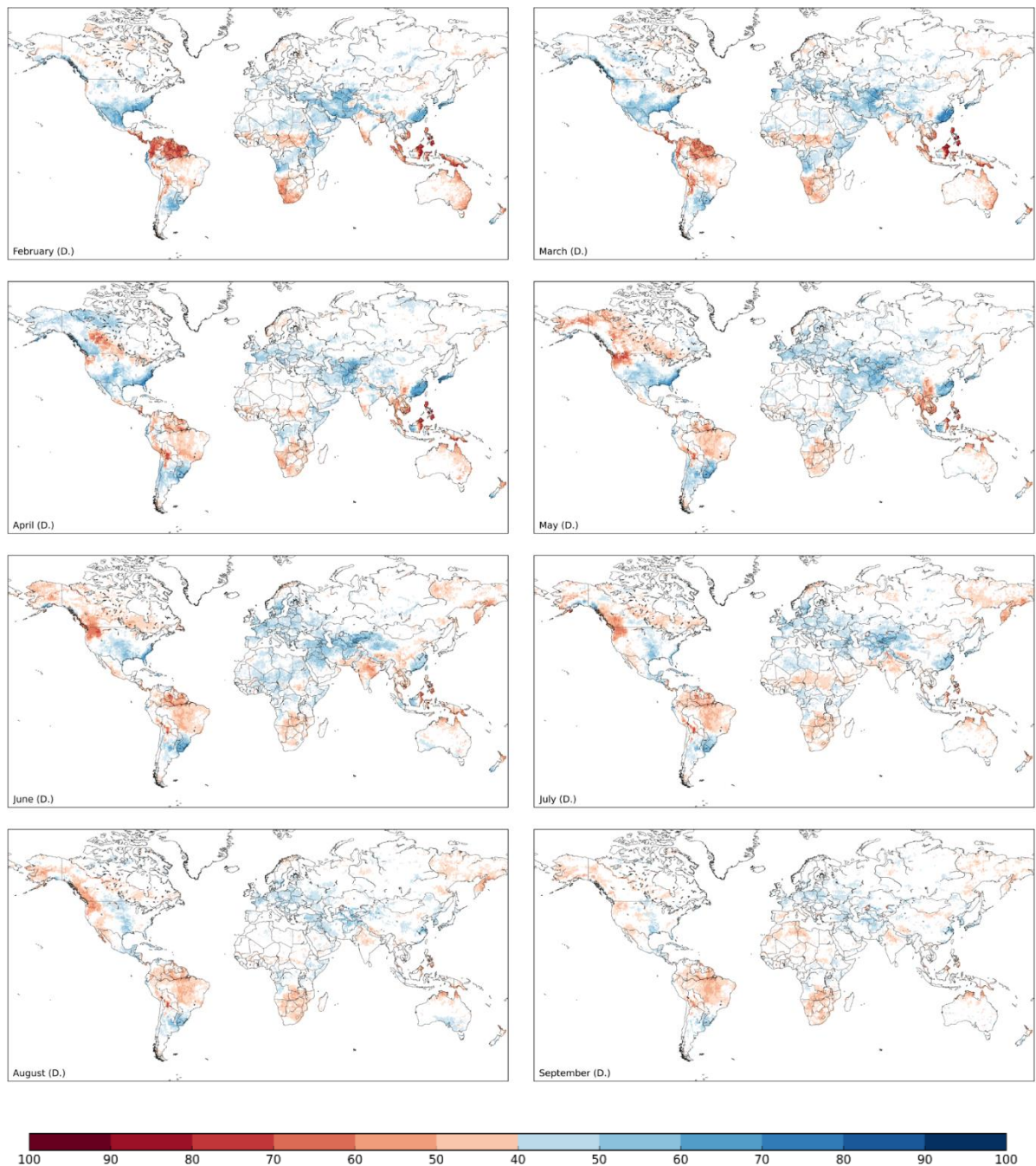
Figure S7: Probability of abnormally high (blue) or low (red) monthly mean river discharge during an El Niño. Each map shows the results for one month, based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year ERA-20CM-R river discharge climatology. "Ev." or "D." indicates whether this map corresponds to the year in which the event typically evolves and peaks ("Ev."), or the year in which the event is decaying ("D.").

Chapter 4. Complex picture for likelihood of ENSO-driven flood hazard
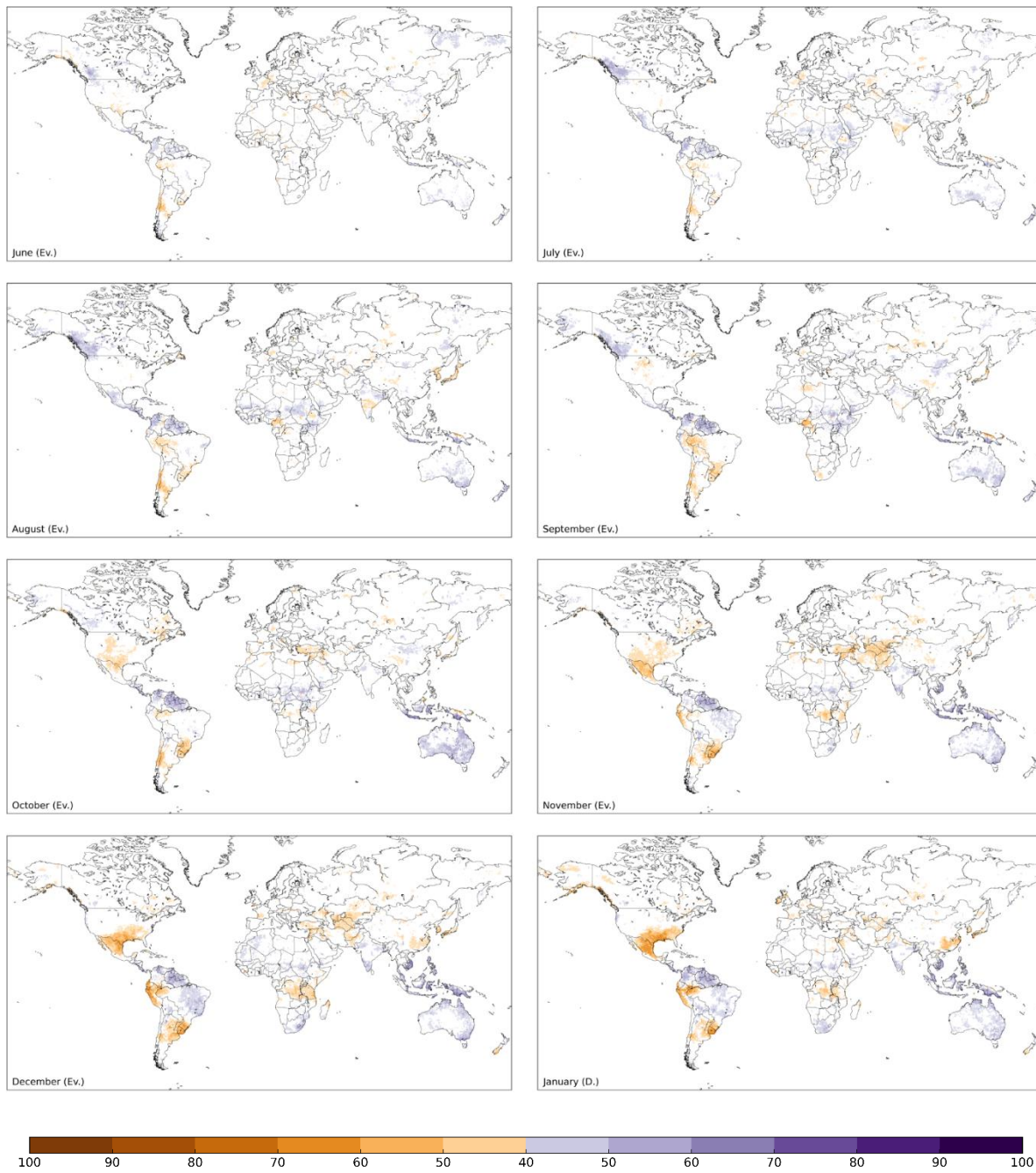
Figure S8 (continued on next page)
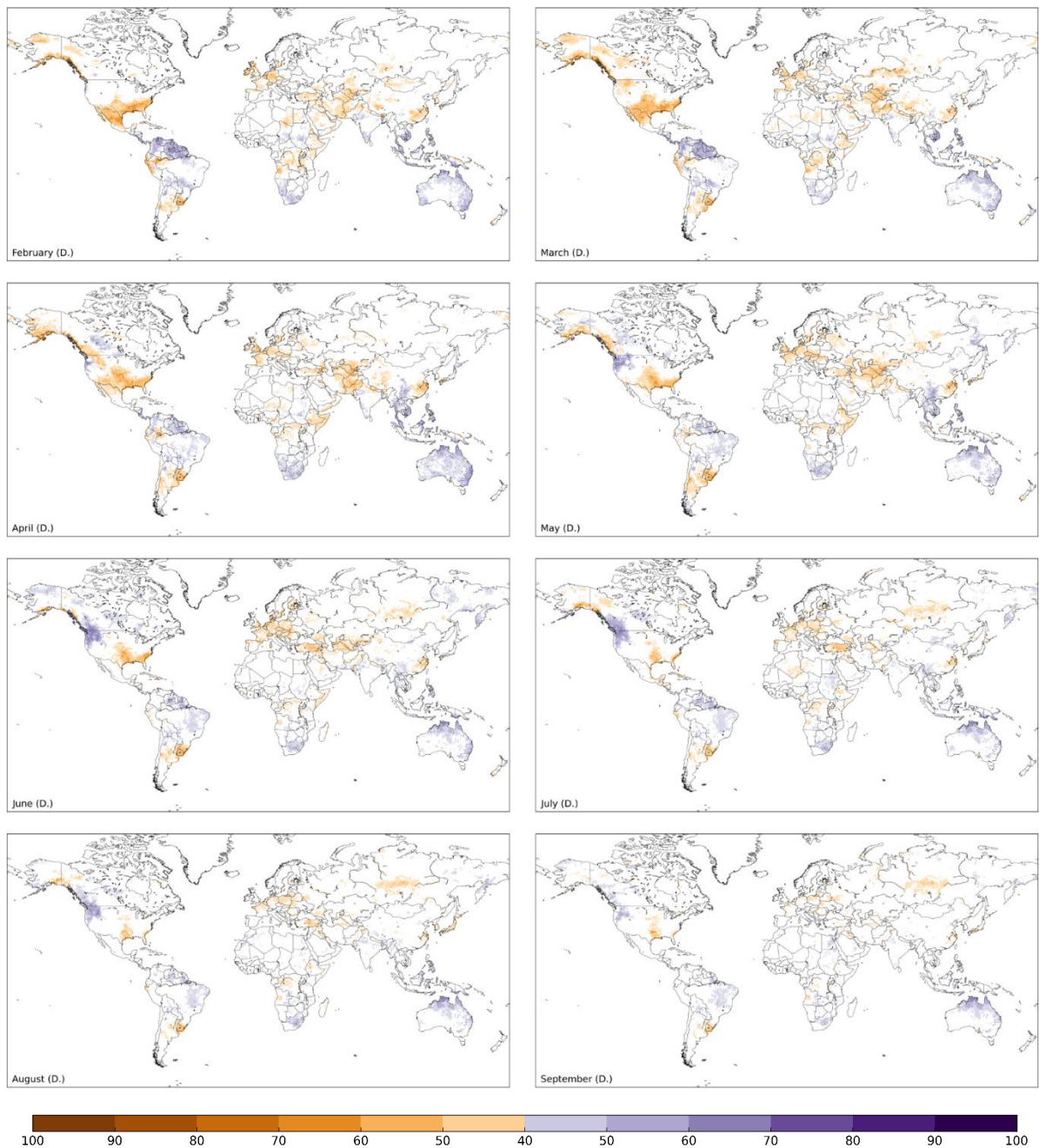
Figure S8: Probability of abnormally high (blue) or low (red) monthly mean river discharge during a La Niña. Each map shows the results for one month, based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year ERA-20CM-R river discharge climatology. "Ev." or "D." indicates whether this map corresponds to the year in which the event typically evolves and peaks ("Ev."), or the year in which the event is decaying ("D.").

As discussed in Chapter 1, there are two key ways in which we can use the inherent predictability of the atmosphere and land state to extend the predictability of flood hazard and provide early indications of potential flood events. In this chapter, we have presented a statistical analysis of the likelihood of increased or decreased flood hazard during El Niño and La Niña, based on the link between the most prominent mode of interannual climate variability, ENSO, and river flow. The following chapter explores the use of coupled ocean-atmosphere GCMs to provide seasonal forecasts of flood hazard for the global river network.

Chapter 5

Developing a Global Operational Seasonal Hydro-Meteorological Forecasting System

This chapter has been published in Geoscientific Model Development with the following reference:

The contributions of the authors of this paper are as follows: F.P. proposed the operational development of the forecasting system, R.E. wrote the new code required to produce and process the seasonal river flow forecasts, and to produce the new forecast products. R.E. and L.A. designed the new forecast products. E.Z. implemented the forecasts into operations at ECMWF, and produced the ERA5-R reanalysis and GloFAS-Seasonal reforecasts. D.M. provided technical support for the website and operational implementation. R.E. wrote the user information for the GloFAS website, designed and carried out the forecast evaluation and wrote the paper, with the exception of Section 5.2.4, written by D.M. All authors were involved in discussions throughout development of the system, and all authors commented on the manuscript. Overall, R.E. conducted 75% of the development of GloFAS-Seasonal, assisted in the operational implementation, conducted 100% of the forecast evaluation and undertook 90% of the writing.

Abstract. Global overviews of upcoming flood and drought events are key for many applications, including disaster risk reduction initiatives. Seasonal forecasts are designed to provide early indications of such events weeks, or even months, in advance, but seasonal forecasts for hydrological variables at large or global scales are few and far between. Here, we present the first operational global scale seasonal hydro-meteorological forecasting system: GloFAS-Seasonal. Developed as an extension of the Global Flood Awareness System

(GloFAS), GloFAS-Seasonal couples seasonal meteorological forecasts from ECMWF with a hydrological model, to provide openly available probabilistic forecasts of river flow out to 4 months ahead for the global river network. This system has potential benefits not only for disaster risk reduction through early awareness of floods and droughts, but also for water-related sectors such as agriculture and water resources management, in particular for regions where no other forecasting system exists. We describe the key hydro-meteorological components and computational framework of GloFAS-Seasonal, alongside the forecast products available, before discussing initial evaluation results and next steps.

5.1 Introduction

Seasonal meteorological forecasts simulate the evolution of the atmosphere over the coming months. They are designed to provide an early indication of the likelihood that a given variable, for example precipitation or temperature, will differ from normal conditions, weeks or months ahead. Will a particular region be warmer or cooler than normal during the next summer? Or will a river have higher or lower flow than normal next winter? Seasonal forecasts of river flow have the potential to benefit many water-related sectors, from agriculture and water resources management, to disaster risk reduction and humanitarian aid through earlier indications of floods or droughts.

Many operational forecasting centres produce long-range (seasonal) global forecasts of meteorological variables, such as precipitation (Weisheimer and Palmer, 2014). However, at present, operational seasonal forecasts of hydrological variables, particularly for large or global scales, are few and far between. A number of continental scale seasonal hydro-meteorological forecasting systems have begun to emerge around the globe over the past decade (Yuan et al., 2015a), using seasonal meteorological forecasts as input to hydrological models to produce forecasts of hydrological variables. These include the European Flood Awareness System (EFAS; Arnal et al., 2018; Cloke et al., 2013a), the European Service for Water Indicators in Climate Change Adaptation (SWICCA; Copernicus, 2018b), the Australian Government Bureau of Meteorology Seasonal Streamflow Forecasts (Bennett et al., 2017; BoM, 2018) and the USA's National Hydrologic Ensemble Forecast Service (HEFS; Demargne et al., 2014; Emerton et al., 2016). There are also various ongoing research efforts using seasonal hydro-meteorological forecasting systems for forecast applications and research purposes at regional (Bell et al., 2017; Bennett et al., 2016; Crochemore et al., 2016; Meißner et al., 2017; Mo et al., 2014; Prudhomme et al., 2017; Wood et al., 2002, 2005; Yuan et al., 2013) and global (Candogan Yossef et al., 2017;

Yuan et al., 2015b) scales. In addition to the ongoing research into improved seasonal hydro-meteorological forecasts at the global scale, an operational system providing consistent global scale seasonal forecasts of hydrological variables could be of great benefit in regions where no other forecasting system exists, and to organisations operating at the global scale (Coughlan De Perez et al., 2017).

Often, in the absence of hydrological forecasts, seasonal precipitation forecasts are used as a proxy for flooding. It has been shown that forecasts of seasonal total rainfall, the most oft-used seasonal precipitation forecasts, are not necessarily a good indicator of seasonal floodiness (Stephens et al., 2015), and other measures of rainfall patterns, or seasonal hydrological forecasts, would be better indicators of potential flood hazard (Coughlan De Perez et al., 2017).

While it seems a natural next step to produce global scale seasonal hydro-meteorological forecasts, this is not a simple task, not only due to the complexities of geographical variations in rainfall-runoff processes and river regimes across the globe, but also due to the computing resources required and huge volumes of data that must be efficiently processed and stored, and the challenge of effectively communicating forecasts for the entire globe. Indeed, global scale forecasting for medium-range timescales has only become possible in recent years due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations and land-surface hydrology modelling, and increased resources and computer power (Emerton et al., 2016). In addition to continued improvements in computing capabilities, the recent move towards the development of coupled atmosphere-ocean-land models means that it is now becoming possible to produce seasonal hydro-meteorological forecasts for the global river network.

Despite the chaotic nature of the atmosphere (Lorenz, 1963), which introduces a limit of predictability (generally accepted to be ~2 weeks), seasonal predictions are possible as they rely on components that vary on longer timescales and are themselves predictable to an extent. This "second type predictability" (Lorenz, 1993) for seasonal river flow forecasts comes from the initial conditions, and large-scale modes of climate variability. The most prominent pattern of climate variability is the El Niño Southern Oscillation (ENSO; McPhaden et al., 2006), which is known to affect river flow and flooding across the globe (Chiew and McMahon, 2002; Emerton et al., 2017; Guimarães Nobre et al., 2017; Ward et al., 2014a, 2014b, 2016). Other teleconnections also influence river flow in various regions of the globe, such as the North Atlantic Oscillation (NAO), Southern Oscillation (SOI), Indian Ocean Dipole (IOD) and

Pacific Decadal Oscillation (PDO), and contribute to the seasonal predictability of hydrologic variables (Yuan et al., 2015a). Coupled atmosphere-ocean-land models are key in representing these large-scale modes of variability in order to produce seasonal hydro-meteorological forecasts.

This motivates the development of an operational global scale seasonal hydro-meteorological forecasting system as an extension of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013), with openly available forecast products. GloFAS is developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission Joint Research Centre (JRC), and has been producing probabilistic flood forecasts out to 30 days for the entire globe since 2012. In 2016, work began, in collaboration with the University of Reading, to implement a seasonal outlook in GloFAS, aiming to provide forecasts of both high and low river flow for the global river network, up to several months in advance. On 10th November 2017, the first GloFAS seasonal river flow forecast was released. This paper introduces the modelling system, its implementation and the available forecast products, and provides an initial evaluation of the potential usefulness and reliability of the forecasts.

5.2 Implementation

The GloFAS seasonal outlooks are produced by driving a hydrological river routing model with meteorological forecasts from ECMWF. The forecasts are run operationally on the ECMWF computing facilities. This section provides an overview of the computing facilities, introduces the key hydro-meteorological components of the modelling platform (the meteorological forecast input, hydrological model and reference climatology), and describes the computational framework of GloFAS-Seasonal.

5.2.1 ECMWF High Performance Computing Facility

ECMWF's current High Performance Computing Facility (HPCF) has been in operation since June 2016, and is used for both forecast production and research activities. The HPCF comprises two identical Cray XC40 supercomputers, each of which is self-sufficient with their own storage, and each with equal access to the storage of the other. Each Cray XC40 consists of 20 cabinets of compute notes and 13 storage nodes. One compute node has 2 Intel Broadwell processors, each with 18 cores, giving 192 nodes (6912 cores) per cabinet. The Cray Aries interconnect is used to connect the processing power. The majority of the nodes of the HPCF are run using the high performance Cray Linux Environment, a stripped-down version of Linux,

as reducing the number of operating system tasks is critical for providing a highly scalable environment.

In terms of storage, each Cray XC40 has ~10PB of storage, and the Data Handling System (DHS) also comprises two main applications; the Meteorological Archive and Retrieval System (MARS), which stores and provides access to meteorological data collected or produced by ECMWF, and ECFS, which stores data that is not suitable for storing on MARS. The DHS holds over 210PB of primary data, and the archive increases by ~233TB per day. The reader is referred to the ECMWF website, *www.ecmwf.int*, for further information on the HPCF and DHS.

In addition to the Cray XC40s, the ECMWF computing facility also includes 4 Linux clusters consisting of 60 servers and 1PB of storage. The Linux clusters are currently used to run the river routing model used in GloFAS and to produce the forecast products, while the meteorological forcing and ERA5 reanalysis are produced on the HPCF. All data related to GloFAS-Seasonal are stored on the MARS and ECFS archives.

5.2.2 Hydro-Meteorological Components
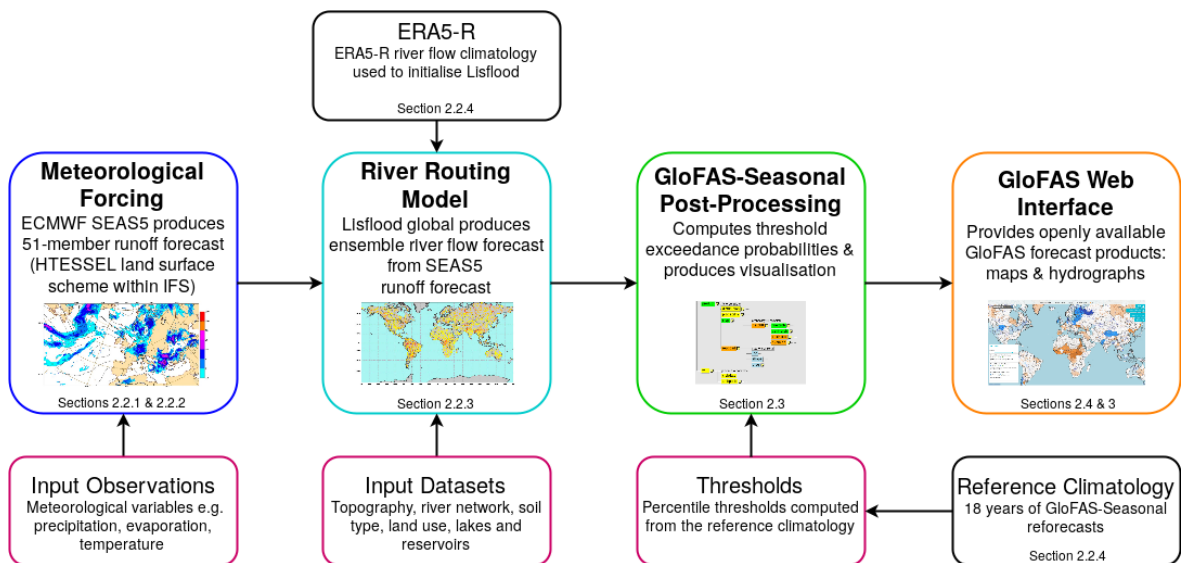
Figure 1: Flowchart depicting the key GloFAS-Seasonal forecasting system components.

5.2.2.1 Meteorological Forcing

The first model component of the seasonal outlook is the meteorological forecast input from the ECMWF Integrated Forecast System (IFS, cycle 43r1 ;ECMWF, 2018c). GloFAS-Seasonal makes use of SEAS5, which is the latest version of ECMWF's long-range ensemble forecasting system, made operational in November 2017 (ECMWF, 2017a; Stockdale et al., 2018). SEAS5

consists of 51 ensemble members (50 perturbed members and one unperturbed control member) and has a horizontal resolution of ~36km (T_{CO}319). The system, which comprises a data assimilation system and a global circulation model, is run once a month, producing forecasts out to 7 months ahead. Initial pre-implementation testing of SEAS5 has suggested that in comparison to the previous version (System 4), SEAS5 better simulates sea surface temperatures (SSTs) in the Pacific Ocean, leading to improved forecasts of El Niño Southern Oscillation (ENSO; Stockdale et al., 2018), which is closely linked to river flow across the globe and can provide added predictability.

SEAS5 is a configuration of the ECMWF IFS (cycle 43r1), including atmosphere-ocean coupling to the NEMO ocean model. SEAS5 is run operationally on the HPCF. Each ensemble member is a complex, HPC-intensive massively parallel code, written in Fortran (version F90). In addition, further complex scripting systems are required to control, prepare, run, post-process and archive all IFS forecasts. The data assimilation systems used to prepare the initial conditions for the forecasts also make use of Fortran and run on the HPCF. For further information, the reader is referred to the IFS documentation (ECMWF, 2018c).

5.2.2.2 Land Surface Component

Within the IFS, which includes SEAS5, the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land, HTESSEL (Balsamo et al., 2011), is used to compute the land surface response to atmospheric forcing. HTESSEL simulates the evolution of soil temperature, moisture content and snowpack conditions through the forecast horizon, to produce a corresponding forecast of surface and subsurface runoff. This component allows for each grid box to be divided into tiles, with up to 6 tiles per grid box (bare ground, low and high vegetation, intercepted water and shaded and exposed snow), describing the land surface. For a given precipitation, the scheme distributes the water as surface runoff and drainage, with dependencies on orography and soil texture. An interception layer accumulates precipitation until saturation is reached, with the remaining precipitation partitioned between surface runoff and infiltration. HTESSEL also accounts for frozen soil, redirecting the rainfall and snowmelt to surface runoff when the uppermost soil layer is frozen, and incorporates a snow scheme. Four soil layers are used to describe the vertical transfer of water and energy, with subsurface water fluxes determined by Darcy's law, and each layer has a sink to account for root extraction in vegetated areas. A detailed description of the hydrology of HTESSEL is provided by Balsamo et al., (2011).

HTESSEL comprises a Fortran library of ~20,000 lines of code, using both F77 and F90 Fortran versions, and is implemented modularly. While HTESSEL can be run on diverse architectures from a workstation PC to the HPCF, operationally, it is run on the HPCF.

5.2.2.3 River Routing Model

As HTESSEL does not simulate water fluxes through the river network, Lisflood (Van Der Knijff et al., 2010), driven by the surface and sub-surface runoff output from HTESSEL interpolated to the 0.1° (~10km) spatial resolution of Lisflood, is used to simulate the groundwater (subsurface water storage and transport) processes and routing of the water through the river network. The initial conditions, used to start the Lisflood model, are taken from the ERA5-R river flow reanalysis (see Section 5.2.2.4).

Lisflood is a spatially distributed hydrological model, including a 1-D channel routing model. Groundwater processes are modelled using two linear reservoirs, the upper zone representing a quick runoff component, including subsurface flow through soil macropores and fast groundwater, and the lower zone representing a slow groundwater component fed by percolation from the upper zone. The routing of surface runoff to the outlet of each grid cell, and the routing of runoff produced by every grid cell from surface, upper and lower groundwater zones through the river network, is done using a four-point implicit finite-difference solution of the kinematic wave equations (Chow et al., 1988). The river network used is that of HydroSHEDS (Lehner et al., 2008), again interpolated to a 0.1° spatial resolution, using the approach of Fekete et al. (2001). For a detailed account of the Lisflood model set-up within GloFAS, the reader is referred to Alfieri et al. (2013).

Lisflood is implemented using a combination of PCRaster GIS and Python, and is currently run operationally on the Linux cluster at ECMWF.

5.2.2.4 Generation of Reforecasts and Reference Climatology

In order to generate a reference climatology for GloFAS-Seasonal, the latest of ECMWF's reanalysis products, ERA5, was used. Reanalysis datasets combine historical observations of the atmosphere, ocean and land surface with a data assimilation system; using global models to "fill in the gaps" and produce consistent global best estimates of the atmosphere, ocean and land state. ERA5 represents the current state of the art in terms of reanalysis datasets, providing a much higher spatial and temporal resolution (30km, hourly) compared to ERA-Interim (79km, 3-hourly), and better representations of precipitation, evaporation and soil moisture (ECMWF,

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system
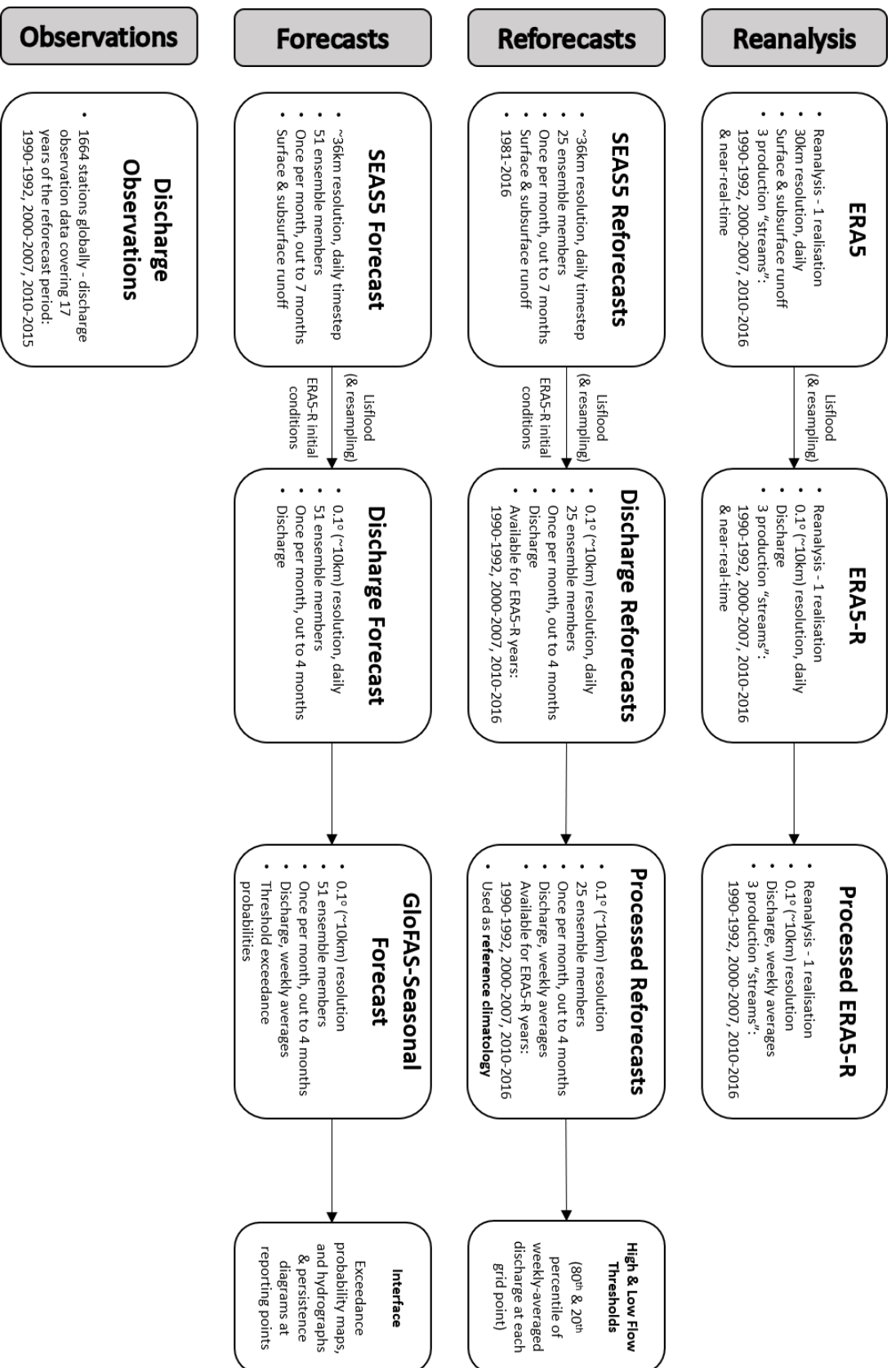
Figure 2: All datasets used and produced for GloFAS-Seasonal, including reanalysis, reforecasts, real-time forecasts and observations.

2017b). In order to produce a river flow reanalysis (ERA5-R) for the global river network, the ERA5 surface and subsurface runoff variables were interpolated to 0.1° (~10km) resolution and used as input to the Lisflood model (see Section 5.2.2.3). ERA5 is currently still in production, and while it will cover the period from 1950 to present when completed, the full dataset will not be available until 2019. ERA5 is being produced in three "streams" in parallel; at the time of producing the ERA5-R reanalysis, 18 years of ERA5 data were available across the three streams (1990-1992, 2000-2007 & 2010-2016). In addition to the historical climatology, ERA5 is also produced in near-real-time, with a delay of just ~3 days, allowing its use as initial conditions for the river routing component of the GloFAS-Seasonal forecasts. The ERA5-R reanalysis is thus updated every month prior to producing the forecast. Figure 2 provides an overview of all datasets used in and produced for the development of GloFAS-Seasonal.

Once the ERA5-R reanalysis was obtained, a set of GloFAS-Seasonal reforecasts was produced. From the 25-ensemble-member SEAS5 reforecasts produced by ECMWF, the surface and subsurface runoff variables were used to drive the Lisflood model, with initial conditions from ERA5-R. This generated 18 years of seasonal river flow reforecasts (one forecast per month out to 4 months lead time, with 25 ensemble members at 0.1° resolution). It is the weekly-averaged river flow from this reforecast dataset which is used as a reference climatology, including to calculate the high and low flow thresholds used in the real-time forecasts (described in Section 5.2.2.4).

5.2.3 GloFAS-Seasonal Computational Framework

The GloFAS-Seasonal real-time forecasts are implemented and run operationally on the ECMWF computing facilities using ecFlow (Bahra, 2011; ECMWF, 2012), an ECMWF work package used to run large numbers of programs with dependencies on each other and on time. An ecFlow suite is a collection of tasks and scheduling instructions, with a user interface allowing interaction and monitoring of the suite, the code behind it, and the output. The GloFAS-Seasonal suite is run once per month, and is used to retrieve the raw SEAS5 forecast data, run this through Lisflood and produce the final forecast products and visualisations using the newly developed GloFAS-Seasonal postprocessing code.

The GloFAS-Seasonal suite performs tasks (detailed below) such as retrieving data, running Lisflood, computing weekly averages and forecast probabilities from the raw Lisflood river flow forecast data, and producing maps and hydrographs for the interface. It is primarily written in Python (version 2.7), with some elements written in R (version 3.1) and shell scripts

incorporating Climate Data Operators (CDO). The code was developed and tested on OpenSUSE Leap 42 systems.
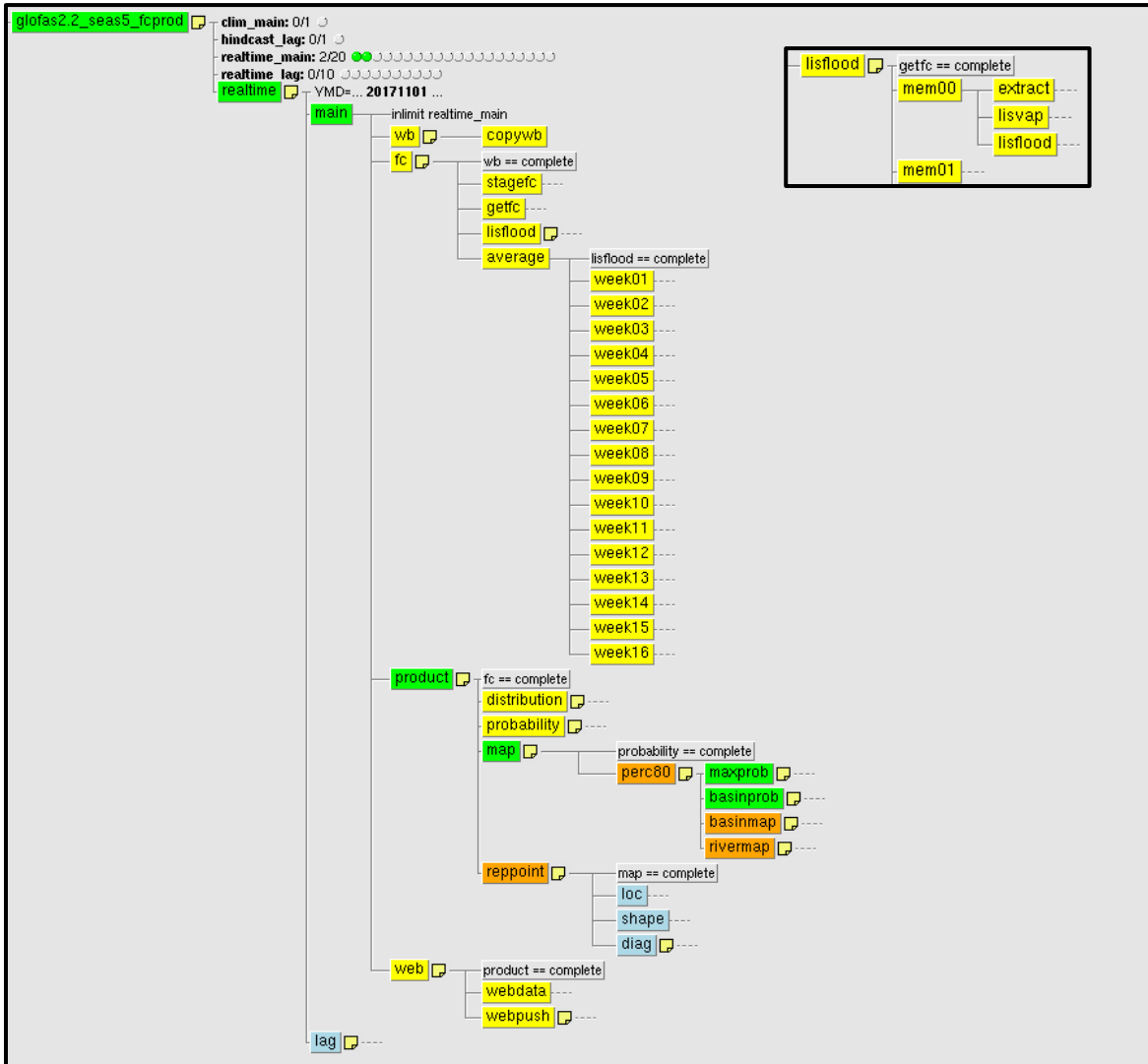
Figure 3: The GloFAS-Seasonal ecFlow suite. The inset image shows the subtasks within the lisflood task, for 1 of the 51 ensemble members. Colours indicate the status of each task, where yellow = complete, green = active, orange = suspended, pale blue = waiting, turquoise (not shown) = queued and red (not shown) = aborted / failed. Grey boxes indicate dependencies, for example "lisflood == complete" indicates that the lisflood task and all lisflood subtasks must have successfully completed in order for the average task to run.

When a new SEAS5 forecast becomes available (typically on the 5th of the month at 00:00UTC), the GloFAS-Seasonal ecFlow suite is automatically deployed. The structure of, and tasks within, the ecFlow suite are shown in Figure 3. Each 'task' represents one script from the GloFAS-Seasonal code. The suite first retrieves the latest raw SEAS5 forecast surface and sub-surface

variables for all 51 ensemble members (*stagefc* and *getfc* tasks), alongside the river flow reference climatology (see Section 5.2.2.4) for the corresponding month of the forecast (*copywb* task). The Lisflood river routing model (described in Section 5.2.2.3) is then run for each of the 51 ensemble members (*lisflood* task). Lisflood is initialised using the ERA5-R river flow reanalysis (see Section 5.2.2.4), and driven with the SEAS5 surface and sub-surface runoff forecast, to produce the 4-month ensemble river flow forecast at a daily time step, from which the weekly-averaged ensemble river flow forecast is obtained (*average* task). The weekly averages are computed for every Monday-Sunday, starting from the first Monday of each month, so that the weekly averages correspond from one forecast to the next. While SEAS5 provides forecasts out to 7 months ahead, the first version of GloFAS-Seasonal uses only the first 4 months. This is in order to reduce the data volumes required, and to allow assessment of the forecast skill out to 4 months ahead, before possible extension of the forecasts out to 7 months ahead in the future.

Once the weekly averaging is complete, the '*forecast product'* section of the suite is deployed, which post-processes the raw forecast output to produce the final forecast products displayed on the web interface. The code behind the *'forecast product'* section is provided in the supplementary material. For a full description of the forecast products, including examples, see Section 5.3. The suite computes the full forecast distribution (*distribution* task), followed by the probability of exceedance for each week of the forecast and for every grid point (*probability* task), based on the number of ensemble members exceeding the high flow threshold or falling below the low flow threshold. The high and low flow thresholds are defined as the 80th and 20th percentiles of the reference climatology, for the week of the year corresponding to the forecast week, so as to use thresholds based on time of year of the forecast. From these weekly exceedance probabilities, the maximum probability of exceedance across the 4-month forecast horizon is calculated for each grid point (*maxprob* task). Basin-averaged maximum probabilities are also produced (*basinprob* task), by calculating the mean maximum probability of exceedance across every grid point at which the upstream area exceeds $1500 km^2$ in each of the 306 major world river basins used in GloFAS-Seasonal (see Section 5.3.1). A minimum upstream area of $1500 km^2$ is chosen as the current resolution of the global model is such that reliable forecasts for very small rivers are not feasible. To put this in context, the upstream area of the River Kennet (a tributary of the River Thames) is $\sim 1000 km^2$, while the upstream area of the River Thames is $\sim 10,000 \ km^2$, the Mekong $\sim 800,000 \ km^2$, and the Amazon $\sim 6,000,000 \ km^2$.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system

These probabilities are used to produce the forecast visualisation for the web interface (Section 5.3). Firstly, the *map* task produces colour-coded maps of both the river network, again for grid points at which the upstream area exceeds 1500km^2, and the major world river basins. The *reppoint* task then produces an ensemble hydrograph and persistence diagrams for a subset of grid points (the 'reporting points') across the globe. Further details on the location of reporting points are given in Section 5.3.3. Finally, the *web* task collates and subsequently transfers all data required for the web interface.

This process, from the time a new SEAS5 forecast becomes available, takes ~4 hours on average to complete, with up to 10 tasks running in parallel (for example, running Lisflood for 10 ensemble members at the same time). It is possible to speed up this process by running more ensemble members in parallel, however, the speed is sufficient that it is not necessary to use further resources to produce the forecast more quickly. GloFAS-Seasonal forecast products are typically produced by the 5th of the month at 05:00UTC and made available via the web interface on the 10th of the month at 01:00UTC. This is the earliest that the GloFAS-Seasonal forecasts can be provided publicly, under the Copernicus license agreement. Data is automatically archived at ECMWF as the suite runs in real-time; ~285GB of data from each SEAS5 forecast are used as input for GloFAS-Seasonal. Each GloFAS-Seasonal forecast run produces an additional ~1.8TB of data, and makes use of the ~18TB reference climatology.

5.2.4 GloFAS Web Interface

The GloFAS website is based on a User-Centred Design (UCD), meaning that user needs are core to the design principles (ISO13407). The website uses Web 2.0 concepts such as simplicity, joy of use and usability, that are synonymous with engaging users. It is a Rich Internet Application (RIA), aiming to provide the same level of interactivity and responsiveness as desktop applications. The website is designed for those engaged in flood forecasting and water resources, as users can browse various aspects of the current forecast or past forecasts in a simple and intuitive way, with spatially distributed information. Map layers containing different information, e.g. flood probabilities for different flood severities, precipitation forecasts, seasonal outlooks, etc. can be activated, and the user can also choose to overlay other information such as land use, urban areas or flood hazard maps. The interface consists of three principal modules: MapServer, GloFAS Web Map Service Time and the Forecast Viewer. These are outlined below.

5.2.4.1 MapServer

MapServer (Open Source Geospatial Foundation, 2016) is an open source development environment for building spatially-enabled internet applications, developed by the University of Minnesota. MapServer has built-in functionality to support industry standard data formats and spatial databases, which is significant to this project, and the support of popular Open Geospatial Consortium (OGC) standards including WMS. In order to exploit the potential of asynchronous data transfer between server and client, the GloFAS raster data has to be divided into a grid of adequate dimensions and an optimal scale sequence.

5.2.4.2 GloFAS Web Map Service Time

The OpenGIS Web Map Service (WMS) is a standard protocol for serving geo-referenced map images over the internet. A Web Map Service Time (WMS-T) is a web service that produces maps in several raster formats or in vector format that may come simultaneously from multiple remote and heterogeneous sources. A WMS server can provide support to temporal requests (WMS-T), by providing a TIME parameter with a time value in the request.

The WMS Specification (OGC, 2015) describes three HTTP requests; *GetCapabilities*, *GetMap* and *GetFeatureInfo*. *GetCapabilities* returns an XML document describing the map layers available and the server's capabilities (i.e. the image formats, projections, and geographic bounds of the server). *GetMap* returns a raster map image. The request arguments, such as the layer id and image format should match those listed as available in the GetCapabilities return document. *GetFeatureInfo* is optional, and is designed to provide WMS clients with more information about features in the map images that were returned by earlier GetMap requests. The response should contain data relating to the features nearest to an image coordinate specified in the GetFeatureInfo request. The structure of the data returned is not defined in the specification and is left up to the WMS server implementation. The GloFAS WMS-T (GloFAS, 2018b) can be freely used, allowing access to the GloFAS layers in any GIS environment, such as QGIS (QGIS Development Team, 2017) or ArcMAP (Environmental Systems Research Institute, 2018). The user manual for the GloFAS WMS-T is available via the GloFAS website (GloFAS, 2018a).

5.2.4.3 Forecast Viewer

The GloFAS forecast viewer is based on the Model View Controller (MVC) architectural pattern used in software engineering. The pattern isolates "domain logic" (the application logic for the

user) from input and presentation (User Interface, UI), permitting independent development, testing and maintenance of each. A fundamental part of this is the AJAX (Asynchronous JavaScript and XML) technology used to enhance user-friendly interfaces for web mapping applications. AJAX technologies have a number of benefits; the essential one is removing the need to reload and refresh the whole page after every event. Careful application design and component selection results in a measurably smaller web server load in geodata rendering and publishing, as there is no need to link and send the whole html document, just the relevant part that needs to be changed.

GloFAS uses OpenLayers (OpenLayers, 2018) as a WMS client. OpenLayers is a JavaScript-based web mapping toolkit designed to make it easy to put a dynamic map on any web page. It doesn't depend on the server technology and can display a set of vector data, such as points, with aerial photographs as backdrop maps from different sources. Closely coupled to the map widget is a layer manager that controls which layers are displayed with facilities for adding, removing and modifying layers. The new layers associated with GloFAS-Seasonal are described in the following section.

5.3 Forecast Products

The GloFAS seasonal outlook is provided as three new forecast layers in the GloFAS forecast viewer: the basin overview, river network and reporting point layers. Each of the three layers represents a different forecast product, described in the following sections. Information on each of the layers is also provided for end users of the forecasts under the dedicated 'Seasonal Outlook' page of the GloFAS website.

5.3.1 Basin Overview Layer

The first GloFAS seasonal outlook product is designed to provide a quick global overview of areas that are likely to experience unusually high or low river flow over the coming 4 months. The "Basin Overview" layer displays a map of 306 major world river basins, colour-coded according to the maximum probability of exceeding the high (blue) or low (orange) flow thresholds (the 80th and 20th percentiles of the reference climatology, respectively) during the 4-month forecast horizon. This value is calculated for each river basin by taking the average of the maximum exceedance probabilities at each grid cell within the basin (using only river pixels with an upstream area >1500km^2). The three different shades of orange / blue indicate the probability: dark (>90%), medium (75-90%) and light (50-75%). Basins that remain white are

those where the probability of unusually high or low flow does not exceed 50% during the 4-month forecast horizon. An example is shown in Figure 4.

As mentioned in Section 5.2.2.3, the Lisflood river network is based on HydroSHEDS (Lehner et al., 2008). In order to generate the river basins used in GloFAS-Seasonal, the corresponding HydroBASINS (Lehner and Grill, 2013) data were used. HydroBASINS consists of a suite of polygon layers depicting watershed boundaries at the global scale. These watersheds were manually merged using QGIS (QGIS Development Team, 2017) to create a global polygon layer of major river basins based on the river network used in the model.

Figure 4. Example screenshot of the seasonal outlook layers in the GloFAS web interface. Shown here are both the "basin overview" layer and "river network" layer, both indicating the maximum probability of unusually high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability: darkest shading = >90% probability, medium shading = 75-90% probability, light shading = 50-75% probability. A white basin or light grey river pixel indicates that the forecast does not exceed 50% probability of high or low flow during the forecast horizon. Legends providing this information are available for each layer by clicking on the green "i" next to the layer toggle (shown at the bottom left in this example).

5.3.2 River Network Layer

The second map layer provides similar information at the sub-basin scale, by colour-coding the entire model river network according to the maximum exceedance probability during the 4-month forecast horizon. This allows the user to zoom in to their region of interest and view the forecast maximum exceedance probabilities in more detail. Again, only river pixels with an

upstream area >1500km^2 are shown. The same colour scheme is used for both the basin overview and river network layers, with blue indicating high flow (exceeding the 80th percentile) and orange low flow (falling below the 20th percentile) and darker colours indicating higher probabilities. In the river network layer, additional colours also represent areas where the forecast does not exceed 50% probability of exceeding either the high or low flow threshold (light grey), and where the river pixel lies in a climatologically arid area and the forecast probability cannot be defined (darker grey-brown). Examples of the river network layer can be seen in both Figure 4 (globally) and Figure 5 (zoomed in).
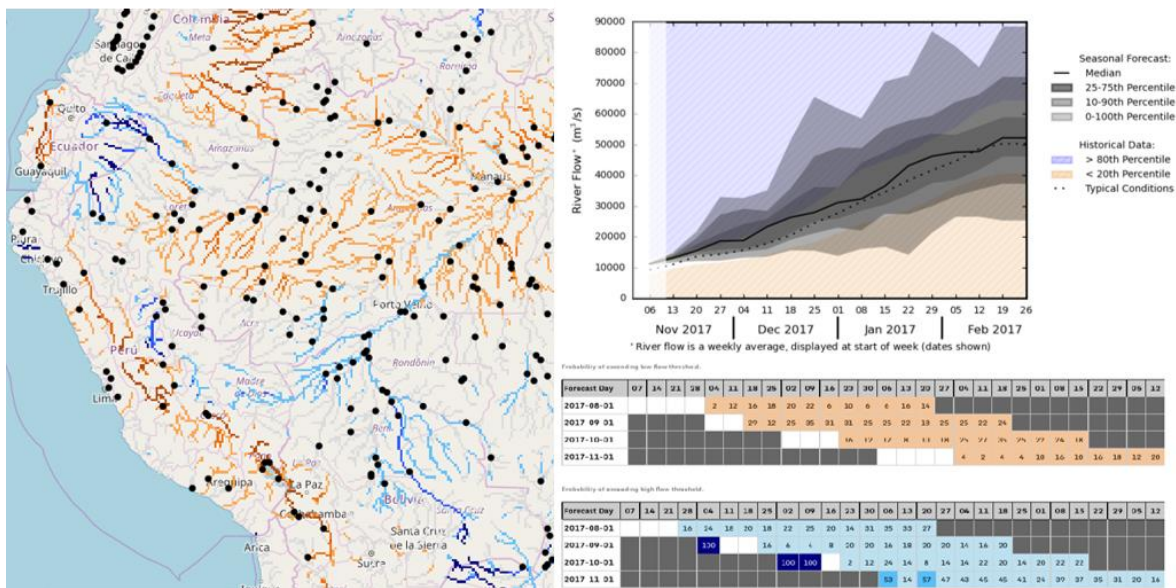
Figure 5: Example of the "reporting points" GloFAS seasonal outlook layer in the web interface. Black circles indicate the reporting points, which provide the ensemble hydrograph (top right) and persistence diagrams for both low flow (centre right) and high flow (bottom right). Also shown is an example section of the "river network" seasonal outlook layer, indicating the maximum probability of high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability.

5.3.3 Reporting Points Layer

In addition to the two summary map layers, reporting points are provided at both static and dynamic locations throughout the global river network, providing additional forecast information; an ensemble hydrograph and a persistence diagram.

Static points originally consisted of a selection of gauged river stations included in the Global Runoff Data Centre (GRDC; BfG, 2017); this set of points has since been expanded to further include points at locations of particular interest to GloFAS partners. There now exist ~2200 static reporting points in the GloFAS interface.

Dynamic points are generated to provide the additional forecast information throughout the global river network, including river reaches where there are no static points. These points are obtained for every new forecast based on a set of selection criteria, adapted from the GloFAS flood forecast dynamic point selection criteria (Alfieri et al., 2013):

- The maximum probability of high [low] river flow (exceeding [falling below] the 80th [20th] percentile of the reference climatology) during the 4-month forecast horizon must be ≥50% for at least 5 contiguous pixels of the river network.

- The upstream area of the selected point must be ≥4000km².

- Dynamic reporting points are generated starting from the most downstream river pixel complying with the previous two selection criteria. A new reporting point is then generated every 300km upstream along the river network, unless a static reporting point already exists within a short distance of the new dynamic point, or the forecasts further upstream no longer comply with the previous two criteria.

Reporting points are displayed as black circles in the "reporting points" seasonal outlook layer. An example is shown in Figure 5. Clicking on a reporting point brings up a new window, containing a hydrograph and persistence diagram alongside some basic information about the location, such as the latitude and longitude, and the upstream area of the point in the model river network. The number of dynamic reporting points can vary from one forecast to the next due to the criteria applied; for example, the March 2018 forecast included ~1600 dynamic points in addition to the static points, thus ~3800 reporting points were available globally.

The ensemble hydrographs (also shown in Figure 5) display a fan plot of the ensemble forecast of weekly-averaged river flow out to 4 months, indicating the spread of the forecast and associated probabilities. Also shown are thresholds based on the reference climatology; the median, and the 80th and 20th percentiles. These thresholds are displayed as a three-week moving average of the weekly-averaged river flow for the given threshold, for the same months of the climatology as that of the forecast (i.e. a forecast for J-F-M-A also displays thresholds based on the reference climatology for J-F-M-A). This allows comparison of the forecast to typical and extreme conditions for the time of year.

Persistence diagrams (see Figure 5) show the weekly probability of exceeding the high and low flow thresholds, for the current forecast (bottom row) and previous three forecasts, colour-coded to match the probabilities indicated in the map layers. These diagrams are provided in

order to highlight the evolution of the forecast, which can indicate whether the forecast is progressing consistently, or whether behaviour is variable from month to month.

5.4 Forecast Evaluation

In this section, the GloFAS-Seasonal reforecasts are evaluated using historical river flow observations. Benchmarking a forecasting system is important to evaluate and understand the value of the system, and in order to communicate the skill of the forecasts to end users (Pappenberger et al., 2015b). This evaluation is designed to measure the ability of the forecasts to predict the correct category of an 'event', i.e. the ability of the forecast to predict that weekly-averaged river flow will fall in the upper 80th or lower 20th percentile of climatology, using a climatology of historical observations as a benchmark. This can be referred to as the potential usefulness of the forecasts, and is of particular importance for decision-making purposes (Arnal et al., 2018). Another key aspect of probabilistic forecasts to consider is their reliability, which indicates the agreement between forecast probabilities and the observed frequency of events.

The potential usefulness is assessed using the relative operating characteristic (ROC) curve, which is based on ratios of the proportion of events (the probability of detection, POD) and non-events (the false alarm rate, FAR) for which warnings were provided (Mason and Graham, 1999), where in this case warnings are treated as forecasts of river flow exceeding the 80th or falling below the 20th percentile of the reference climatology (see Section 5.2.2.4). These ratios allow for estimation of the probability that an event will be predicted.

For each week of the forecast (out to 16 weeks, corresponding to the forecasts provided via the interface, for example the hydrograph shown in Figure 5), the POD (eq. 1) and FAR (eq. 2) are calculated for both the 80th and 20th percentile events at each observation station:

$$POD = \frac{\text{hits}}{\text{hits+misses}} \tag{1}$$

$$FAR = \frac{\text{false alarms}}{\text{false alarms+correct negatives}} \tag{2}$$

where a *hit* is defined when the forecast correctly exceeded [fell below] the 80th [20th] percentile of the reference climatology during the same week that the observed river flow exceeded [fell below] the 80th [20th] percentile of the observations at that station. It follows that a *miss* is defined when an event was observed but the forecast did not exceed the threshold, and a *false alarm* when the forecast exceeded the threshold but no event was observed. From these, the area under the ROC curve (AROC) is calculated, again for both the 80th and 20th percentile events.

The AROC ($0 \leq$ AROC ≤ 1, where 1 is perfect) indicates the skill of the forecasts compared to the long-term average climatology (which has an AROC of 0.5) and is used here to evaluate the potential usefulness of the forecasts. The maximum lead time at which forecasts are more skilful than climatology (AROC > 0.5) is identified; a forecast with an AROC < 0.5 would be less skilful than climatology, and thus not useful.

The reliability of the forecasts is assessed using attributes diagrams, which show the relationship between the forecast probability and the observed frequency of the events. While the ROC measures the ability of a forecasting system to predict the correct category of an event, the reliability assesses how closely the forecast probabilities correspond to the actual chance of observing the event. As such, these evaluation metrics are useful to consider together. As with the ROC calculations, the reliability is assessed for each week of the forecast (out to 16 weeks), and for both the 80th and 20th percentile events. The range of forecast probabilities is divided into 10 bins (0-10%, 10-20%, etc.), and the forecast probability is plotted against the frequency at which an event was observed for forecasts in each probability bin. Perfect reliability is exhibited when the forecast probability and the observed frequency are equal, for example if a forecast predicts that an event will occur with a probability of 60%, then the event should occur on 60% of the occasions that this forecast was made. Attributes diagrams can also be used to assess the sharpness and resolution of the forecasts. Forecasts that do not discriminate between events and non-events are said to have no resolution (a forecast of climatology would have no resolution), and forecasts which are capable of predicting events with probabilities that differ from the observed frequency, such as forecasts of high or 0 probability, are said to have sharpness.

The GloFAS-Seasonal reforecasts (of which there are 216, covering 18 years, as described in Section 5.2.2.4 and Figure 2) are compared to river flow observations that have been made available to GloFAS, covering 17 years of the study period up to the end of 2015, when the data were collated (see Figure 2). To ensure a large enough sample size for this analysis, alongside the best possible spatial coverage, the following criteria are applied to the data:

- The weekly river flow data record available for each station must contain no more than 53% (9 years) missing data. The high and low flow thresholds (the 80th and 20th percentile, respectively) are calculated using the observations for each station, and for each week, across the 17 years of data, so a sample size of 17 is the maximum possible. A threshold of (up to) 53% missing data allows for a minimum sample size of 8.

Selecting a smaller threshold reduced the number of stations, and the spatial coverage across the globe, significantly. The percentage of missing data is calculated at each station and for each week of the dataset independently, and as such the number of stations used can vary slightly with time.

– The upstream area of the corresponding grid point in the model river network must be at least 1500km^2.

These criteria allow for the use of 1140±14 stations globally. While the dataset contains 6122 stations, just 1664 of these contain data during the 17-year period, and none have the full 17 years of data available. Data from human-influenced rivers have not been removed, as in this study we are interested in identifying the ability of the forecasting system in its current state to predict observed events, rather than the ability of the hydrological model to represent natural flow.

5.4.1 Potential Usefulness

In order to gain an overview of the potential usefulness of the GloFAS-Seasonal forecasts across the globe, we map the maximum lead time at which the forecasts are more skilful than climatology (i.e. AROC > 0.5), at each observation station, averaged across all forecast months. These results are shown in Figure 6, and it is clear that forecasts of both high and low flow events are more skilful than climatology across much of the globe, with potentially useful forecasts at many stations out to 4 months ahead. However, there are regions where the forecasts are (on average, across all forecast months) not useful (i.e. AROC < 0.5), such as the western USA and Canada (excluding coastlines), much of Africa, and additionally across parts of Europe for low flow events. As forecasts with an AROC larger than but close to 0.5 could be deemed as only marginally more skilful than climatology, we apply a skill buffer, setting the threshold to AROC > 0.6 for a forecast to be deemed as potentially useful. These results are mapped in Figure 7, and clearly indicate the reduction in the lead time at which forecasts are potentially useful (for both high and low flow events) at many stations, implying that in some locations, forecasts beyond the first 1-2 months are only marginally more skilful than climatology. There are, however, stations in some rivers with an AROC > 0.6 out to 4 months lead time, and many locations across the globe that still indicate that forecasts are potentially useful 1-2 months ahead for both high and low flow events.

These results can be further broken down by season, indicating whether the forecasts are more potentially useful at certain times of the year. Maps showing the maximum lead time at which
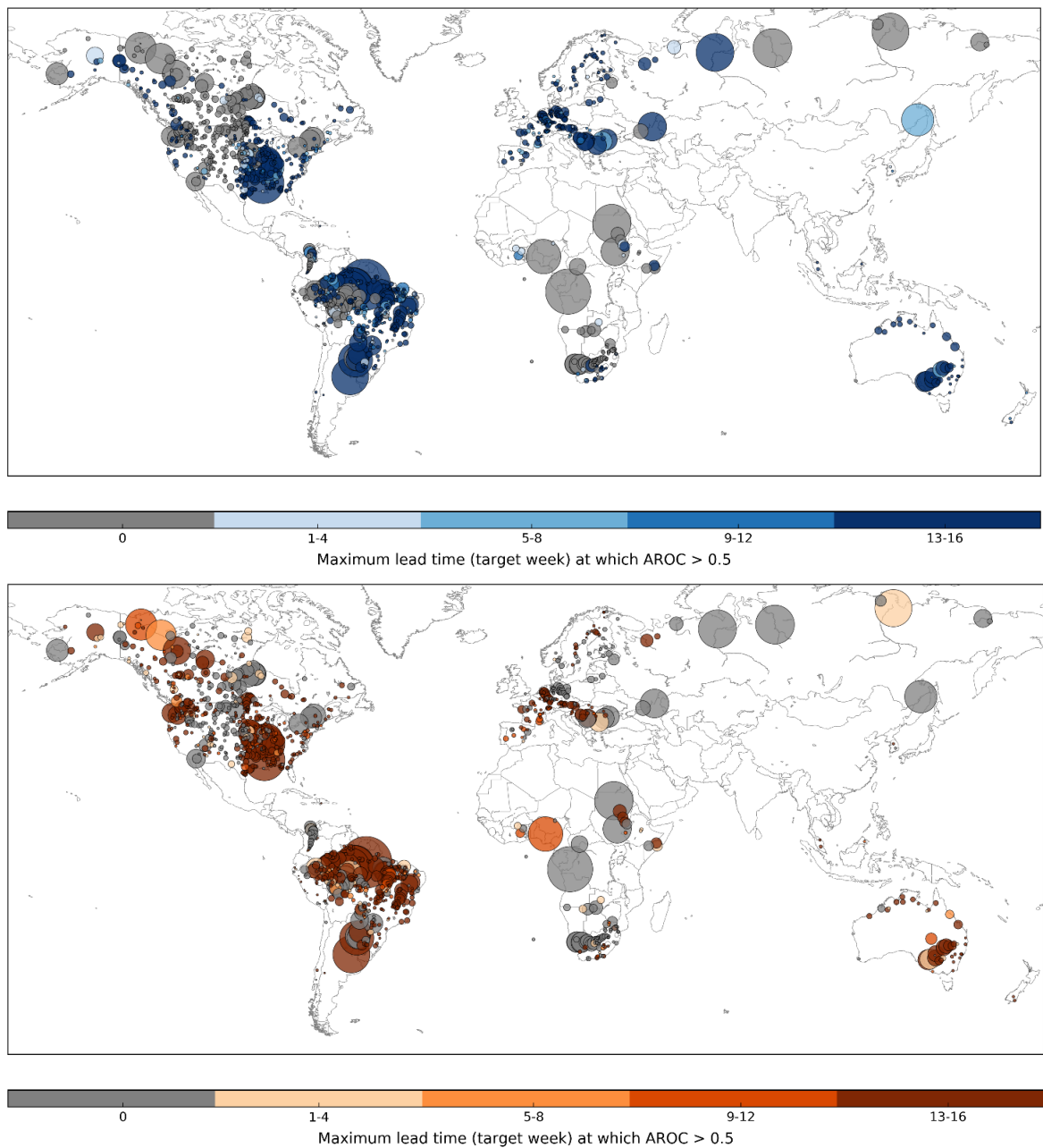
Figure 6: Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.5 for high flow events (flow exceeding the 80th percentile of climatology, top panel) and low flow events (flow below the 20th percentile of climatology, bottom panel), at each observation station. This is used to indicate the maximum lead time at which forecasts are more skilful than the long-term average. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skilful than climatology at all lead times.

Figure 7: Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.6 for high flow events (flow exceeding the 80th percentile of climatology, top panel) and low flow events (flow below the 20th percentile of climatology, bottom panel), at each observation station. This is used to indicate the maximum lead time at which forecasts are deemed skilful. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skilful than climatology at all lead times. Maps for each season are provided in the supplementary material.

AROC > 0.6 for each season (for forecasts started during the season; e.g. DJF indicates the average results for forecasts produced on 1st December, 1st January and 1st February) are

provided for high and low flow events in the supplementary material, Figure S1 and S2, respectively. The following paragraphs provide an overview of these results for each continent; for further detail please refer to the maps.

South America: For high flow events, forecasts for the Amazon basin in DJF and MAM are potentially useful out to longer lead times (up to 3-4 months) and at more stations than in JJA and SON, with similar results in MAM for low flow events. In contrast, further south, forecasts are most potentially useful in JJA and SON, up to 4 months ahead. In the more mountainous regions of western South America, forecasts in JJA and SON are generally less skilful than climatology for high and low flow events. In the northwest, however, for some stations, forecasts started in DJF and MAM are potentially useful up to 3 months ahead.

North America: In eastern North America, JJA and SON forecasts are most potentially useful, with more stations indicating an AROC > 0.6 out to 2-3 months ahead. However, during all seasons there are several stations in the east showing skill out to varying lead times. Much of the western half of the continent (excluding coastal areas) sees forecasts that are less skilful than climatology during all seasons, although some stations do indicate skill up to 4 months ahead for high flow, for forecasts started in MAM and JJA, and for low flow in MAM. At many coastal stations in the west, forecasts of high flow events started in DJF, MAM and JJA do indicate skill out to 3-4 months, and out to ~6 weeks in SON.

Europe: Forecasts for European rivers generally perform best for high flow events in SON and DJF, with the exception of some larger rivers in eastern Europe, for which the forecasts are more potentially useful in JJA and SON. In MAM and JJA, the number of stations indicating no skill is generally higher. In contrast, forecasts for low flow events are less skilful than climatology across much of Europe. Particularly in northeast Europe and Scandinavia, forecasts produced in the summer months of JJA have an AROC < 0.6 at all stations, with only a few stations indicating any skill in other seasons, whereas in central and southeast Europe forecasts of low flow events are most skilful in JJA and SON, out to 3-4 months ahead in the larger rivers. These results are similar to those of Arnal et al. (2018) for the potential usefulness of the EFAS seasonal outlook.

Asia: Although the number of available stations is very limited, the few stations available in southeast Asia indicate that the forecasts are potentially useful out to 3-4 months ahead, particularly for forecasts started in DJF and MAM, preceding the start of the wet season. For

low flow events, this skill extends into JJA, whereas forecasts made in SON, towards the end of the wet season, tend to be less skilful than climatology.

Australia & New Zealand: Forecasts are most skilful out to longer lead times in the Murray-Darling river basin in the southeast, in particular for forecasts started in JJA and SON during the southern hemisphere winter and spring. In northern Australia, forecasts started in DJF and MAM for high flow events, and MAM and JJA for low flow events, are potentially useful out to 3-4 months ahead. This corresponds with the assessment of the skill of the Bayesian joint probability modelling approach for sub-seasonal to seasonal streamflow forecasting in Australia by Zhao et al. (2016), who found that forecasts in northern Australian catchments tend to be more skilful for the dry season (May to October) than the wet season (December to March). At the 3 stations in New Zealand, forecasts are only skilful for high flow events during the first month of lead time, in DJF and MAM; however, for low flow events forecasts made in SON for the southern stations are potentially useful out to 4 months ahead.

Africa: While the spatial distribution of stations is limited, for high flow events forecasts are seen to be potentially useful at some of the stations in eastern Africa, particularly in SON and to a lesser extent in DJF. In southern Africa, there is skill in DJF and MAM, although the maximum lead time varies significantly from station to station. For low flow, there is little variation between the seasons; forecasts are generally less skilful than climatology across the continent, with some stations in DJF in southern and western Africa indicating skill in the first 1-2 months only.

5.4.2 Reliability

To provide an overall picture of the reliability of the GloFAS-Seasonal forecasts, attributes diagrams are produced for forecasts aggregated across all observation stations globally, for both the 80th and 20th percentile events. In order to assess geographical differences in forecast reliability, attributes diagrams are also produced for forecasts aggregated across the stations within each of the major river basins used in the GloFAS-Seasonal forecast products (see Section 5.3.1). Many of these river basins do not contain a large enough number of stations to produce useful attributes diagrams, and as such, results in this section are presented for one river basin per continent for this initial evaluation. The river basin chosen for each continent is that which contains the largest number of observation stations.

The globally aggregated results (Figure 8) indicate that, in general, the forecasts have more reliability than a forecast of climatology, though the reliability is less than perfect. It is important

to note that the globally aggregated results shown in Figure 8 mask any variability between river basins. Overall, the reliability appears to be slightly better for forecasts of high flow events than low flow events, and for lower probabilities, indicated by the steeper positive slope showing that as the forecast probability increases, so does the verified chance of the event. The forecasts for both high and low flow events exhibit sharpness, although more so for high flow events, meaning that they have the ability to forecast probabilities that differ from the climatological average. This is indicated by the histograms inset within the attributes diagrams in Figure 8; a forecast with sharpness will show a range of forecast probabilities differing from the climatological average (20%), and a forecast with perfect sharpness will show peaks in the forecast frequency at 0% and 100%. Forecasts with no, or low, sharpness will show a peak in the forecast frequency near to the climatological average. A forecast can have sharpness but still be unreliable. Figure 8 also suggests that in general, GloFAS-Seasonal forecasts have a tendency to over-predict the likelihood of an event occurring.

Figure 8: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right), aggregated across all observation stations globally. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness. Attributes diagrams for selected river basins are provided in the supplementary material.

The following paragraphs summarise the forecast reliability for one river basin per continent; for a map of the location of these river basins, please refer to Figure S3. The attributes diagrams for these river basins for both the 80th and 20th percentile events, and for each season, are

provided in Figure S4 – S8. Each attributes diagram displays the results for forecast weeks 4,8,12 and 16, representing the reliability out to 1,2,3 and 4 months ahead. There are no river basins in Asia containing enough stations to produce an attributes diagram.

South America, Tocantins River (Figure S4). For high flow events, forecasts for the Tocantins River indicate good reliability in all seasons, particularly up to 50% probability. Forecasts in the higher probability bins tend to over-predict, and this over-prediction worsens with lead time. In MAM and JJA, the forecasts tend to slightly under-predict in the lower probability bins. The forecasts have sharpness, but it is clear that the sample size of high probability forecasts is limited. There is a tendency to over-predict the likelihood of low flow events in all seasons, but the forecasts show good reliability for the lower probability bins, particularly in SON and DJF. In JJA, the resolution of the forecasts is low.

North America, Lower Mississippi River (Figure S5). For high flow events, the sample size of high probability forecasts is small, and as such it is difficult to evaluate the reliability of these forecasts. The forecasts at lower probabilities have good reliability, particularly out to 2 months ahead in MAM and JJA. In SON and DJF, forecasts are more reliable at longer lead times. There is a tendency to under-predict at low probabilities and over-predict at high probabilities. For low flow events, the forecasts have a tendency to over-predict in all seasons, and the resolution of the forecasts is lower than for high flow events. At higher probabilities, forecasts of low flow events are more reliable than climatology, but the resolution is particularly low for probabilities up to 50-60%. The forecasts for both high and low flow events have sharpness.

Europe, River Rhone (Figure S6). For the River Rhone, the reliability is better than climatology at all lead times for high flow events, although there is a lack of forecasts of higher probabilities, particularly in MAM and JJA, as may be expected in the summer months. In SON, the reliability of forecasts up to 60-70% is good at all lead times, and in DJF the forecasts are more reliable in the first 2 months of lead time for most probability bins. The reliability is less good for low flow events, but is generally better than climatology, particularly in summer (JJA). In winter (DJF), the resolution and reliability of the forecasts are poor. For all seasons and lead times, and for both events, the forecasts have sharpness.

Australia, River Murray (Figure S7). The attributes diagrams for both high and low flow events indicate that forecasts are often over-confident in this river basin, with probabilities of 0-10% for low flow events, and 0-30% and 90-100% for high flow events, occurring frequently. As such, the sample size of forecasts in several of the bins is low. For high flow events, forecasts

tend to over-predict at high probabilities, and under-predict at low probabilities. The reliability is very good up to ~30%, after which the sample size is too small. For low flow events, there is a tendency to under-predict, but based on the forecasts available, the reliability is better than climatology at all lead times. The reliability for low flow events is better in SON and DJF (spring and summer), than MAM and JJA (autumn and winter) and for high flow events there is less differentiation between the seasons.

Africa, Orange River (Figure S8). For the Orange River, forecasts of high flow events exhibit good reliability for lower probabilities in SON, DJF and MAM (spring through autumn), particularly at longer lead times in SON and DJF, with a tendency to over-predict at higher probabilities. Resolution and reliability are poor for high flow events in JJA (winter), with probabilities of 90-100% predicted too frequently. For low flow events, forecasts of 0-10% are very frequent, and the forecasts under-predict in all seasons, although the reliability is better than climatology at all lead times (based on a limited sample of forecasts for most probability bins). Reliability for low flow events is best in DJF (summer).

5.4.3 Discussion

The results presented provide an initial evaluation of the potential usefulness and reliability of GloFAS-Seasonal forecasts. For decision-making purposes, it is important to measure the ability of a forecasting system to predict the correct category of an event. As such, an event-based evaluation of the forecasts is used to assess whether the forecasts were able to correctly predict observed high and low river flow events over a 17-year period, and whether it is able to do so with good reliability. The initial results are promising, indicating that the forecasts are, on average, potentially useful up to 1-2 months ahead in many rivers worldwide, and up to 3-4 months ahead in some locations. The GloFAS-Seasonal forecasts have sharpness, i.e. they are able to predict forecasts with probabilities that differ from climatology, and overall have better reliability than a forecast of climatology, but with a tendency to over-predict at higher probabilities. It is also clear that there exists a frequency bias in the reliability results, as often there is a small sample of high probability forecasts. Typically, the reliability is seen to be better when there is a higher forecast frequency on which to base the results. As would be expected, the potential usefulness and reliability of the forecasts vary by region, season and forecast lead time.

Considering the evaluation results by season allows further analysis of the times of year in which the forecasts are potentially useful and/or reliable. For example, in southeast Australia, forecasts

are seen to be potentially useful up to 4 months ahead in JJA and SON, but for forecasts produced in DJF the skill only extends to 1 month ahead, and forecasts are less skilful than climatology at several of the stations in MAM. In many rivers across the globe, it is the case that forecasts are potentially useful in some seasons, but not in others, and may be more reliable in certain seasons than others. As such, the maps provided in Figure S1 and S2 are intended to highlight where and when the forecasts are likely to be useful, information that is key in terms of decision-making.

It is clear that there are regions and seasons where the forecasts are less skilful than climatology and do not have good reliability, and thus in these rivers it would be more useful to use a long-term average climatology than seasonal hydro-meteorological forecasts of river flow. This lack of skill could be due to several factors, such as certain hydrological regimes that may not be well-represented in the hydrological model or may be difficult to forecast at these lead times (for example snow dominated-catchments, or regions where convective storms produce most of the rainfall in some seasons), poor skill of the meteorological forecast input, poor initial conditions from the ERA5-R reanalysis, extensive management of rivers that cannot be represented by the current model, or the lack of model calibration. While this initial evaluation is designed to provide an overview of whether the forecasts are potentially useful and reliable in predicting high and low flow events, more extensive analysis is required to diagnose the sources of predictability in the forecasts and the potential causes of poor skill. Additionally, it is evident that observations of river flow, particularly covering the reforecast period, are both spatially and temporally limited across large areas of the globe. A more extensive analysis should make use of the globally consistent ERA5-R river flow reanalysis as a benchmark in order to fully assess the forecast skill worldwide, including in regions where no observations are available.

The verification metrics used also require that a high or low flow event is predicted with the correct timing, in the same week as that in which it occurred. This is asking a lot of a seasonal forecasting system and for many applications, such as water resources and reservoir management, a forecast of the exact week in which an event is expected at a lead time of several months ahead may not be necessary. That such a system shows real skill despite this being a tough test for the model, and is able to successfully predict observed high or low river flow in a specific week, several weeks or months ahead, provides optimism for the future of global scale seasonal hydro-meteorological forecasting. Further evaluation should aim to assess the skill of the forecasts with a more relaxed constraint on the event timing, and also make use of alternative skill measures to cover different aspects of the forecast skill, such as the spread and bias of the

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system

forecasts. It will also be important to assess whether the use of weekly-averaged river flow is the most appropriate way to display the forecasts. While this is commonly used for applications such as drought early awareness and water resources management, there may be other aspects of decision-making, such as flood forecasting, for which other measures may be more appropriate, for example daily averages or floodiness (Stephens et al., 2015).

Future development of GloFAS-Seasonal will aim to address these evaluation results and improve the skill and reliability of the current forecasts, and will also aim to overcome some of the grand challenges in operational hydrological forecasting, such as seamless forecasting and the use of data assimilation. Seamless forecasting will be key in the future development of GloFAS; the use of two different meteorological forecast inputs for the medium-range and seasonal versions of the model means that discrepancies can occur between the two timescales thus providing confusing, inconsistent forecast information to users. Additionally, the use of river flow observations could lead to significant improvements in skill, through calibration of the model using historical observations, and assimilation of real-time data to adjust the forecasts. This remains a grand challenge due to the lack of openly available river flow data, particularly in real time.

5.5 Conclusions

In this paper, the development and implementation of a global scale operational seasonal hydro-meteorological forecasting system, GloFAS-Seasonal, was presented, and an event-based forecast evaluation was carried out using two different but complementary verification metrics, to assess the capability of the forecasts to predict high and low river flow events.

GloFAS-Seasonal provides forecasts of high or low river flow out to 4 months ahead for the global river network through three new forecast product layers via the openly available GloFAS web interface at *www.globalfloods.eu*. Initial evaluation results are promising, indicating that in many rivers, forecasts are both potentially useful, i.e. more skilful than a long-term average climatology, out to several months ahead in some cases, and overall more reliable than a forecast of climatology. Forecast skill and reliability vary significantly by region and by season.

The initial evaluation however also indicates a tendency of the forecasts to over-predict, in general, and in some regions forecasts are currently less skilful than climatology; future development of the system will aim to improve the forecast skill and reliability with a view to providing potentially useful forecasts across the globe. Development of GloFAS-Seasonal will

continue based on results of the forecast evaluation, and on feedback from GloFAS partners and users worldwide, in order to provide a forecast product that remains state-of-the-art in hydro-meteorological forecasting, and caters to the needs of its users. Future versions are likely to address some of the grand challenges in hydro-meteorological forecasting in order to improve forecast skill, such as data assimilation, and will also include more features, such as flexible percentile thresholds and indication of the forecast skill via the interface. A further grand challenge that is important in terms of global scale hydro-meteorological forecasting and indeed for the development of GloFAS, is the need for more observed data (Emerton et al., 2016), which is essential not only for providing initial conditions to force the models, but also for evaluation of the forecasts and continuous improvement of forecast accuracy.

While such a forecasting system requires extensive computing resources, the potential for use in decision-making across a range of water-related sectors, and the promising results of the initial evaluation, suggest that it is a worthwhile use of time and resources to develop such global scale systems. Recent papers have highlighted that seasonal forecasts of precipitation are not necessarily a good indicator of potential floodiness, and called for investment in better forecasts of seasonal flood risk (Coughlan De Perez et al., 2017; Stephens et al., 2015). Coughlan de Perez et al. (2017) state that "ultimately, the most informative forecasts of flood hazard at the seasonal scale could be seasonal streamflow forecasts using hydrological models", and that better seasonal forecasts of flood risk could be hugely beneficial for disaster preparedness.

GloFAS-Seasonal represents a first attempt at overcoming the challenges of producing and providing openly-available seasonal hydro-meteorological forecast products, which are key for organisations working at the global scale, and for regions where no other forecasting system exists. We provide, for the first time, seasonal forecasts of hydrological variables for the global river network, by driving a hydrological model with seasonal meteorological forecasts. GloFAS-Seasonal forecasts could be used in addition to other forecast products such as seasonal rainfall forecasts and short-range forecasts from national hydro-meteorological centres across the globe, to provide useful added information for many water-related applications, from water resources management and agriculture to disaster risk reduction.

Code Availability. The ECMWF IFS source code is available subject to a license agreement, and as such access is available to the ECMWF member-state weather services and other approved partners. The IFS code is also available for educational and academic purposes as part of the OpenIFS project (ECMWF, 2011, 2018b), with full forecast capabilities and including the HTESSEL land surface scheme, but without modules for data assimilation. Similarly, the GloFAS river routing component source code is not openly available; however, the *'forecast product'* code (prior to implementation in ecFlow) that was newly developed for GloFAS-Seasonal, used for a number of tasks such as computing exceedance probabilities and producing the graphics for the interface, is provided in the supplementary material.

Data Availability. ECMWF's ERA5 reanalysis and SEAS5 reforecasts are available through the Copernicus Climate Data Store (Copernicus, 2018a). The ERA5-R river flow reanalysis and the GloFAS-Seasonal reforecasts (daily data) are currently available from the authors on request, and will be made available through ECMWF's data repository in due course. The majority of the observed river flow data was provided by the Global Runoff Data Centre (GRDC; BfG, 2017). This data is freely available from *www.bafg.de/GRDC*. Additional data was provided by the Russian State Hydrological Institute (SHI, 2018), the European Flood Awareness System (EFAS, 2017), Somalia Water and Land Information Management (SWALIM, 2018), South Africa Department for Water and Sanitation (DWA, 2018), Colombia Institute of Hydrology, Meteorology and Environmental Studies (IDEAM, 2014), Nicaragua Institute of Earth Studies (INETER, 2016), Dominican Republic National Institute of Hydraulic Resources (INDRHI, 2017), Brazil National Centre for Monitoring and Forecasting of Natural Hazards (Cemaden, 2017), Environment Canada Water Office (Environment Canada, 2014), Nepal Department of Hydrology and Meteorology (DHM, 2017), Red Cross Red Crescent Climate Centre (RCCC, 2018), Chile General Water Directorate (DGA, 2018), Historical Database on Floods (BDHI, 2018).

Emergencies and Resilience (SHEAR) research programme (project FATHUM (Forecasts for AnTicipatory HUManitarian action)).

5.6 Supplementary Figures
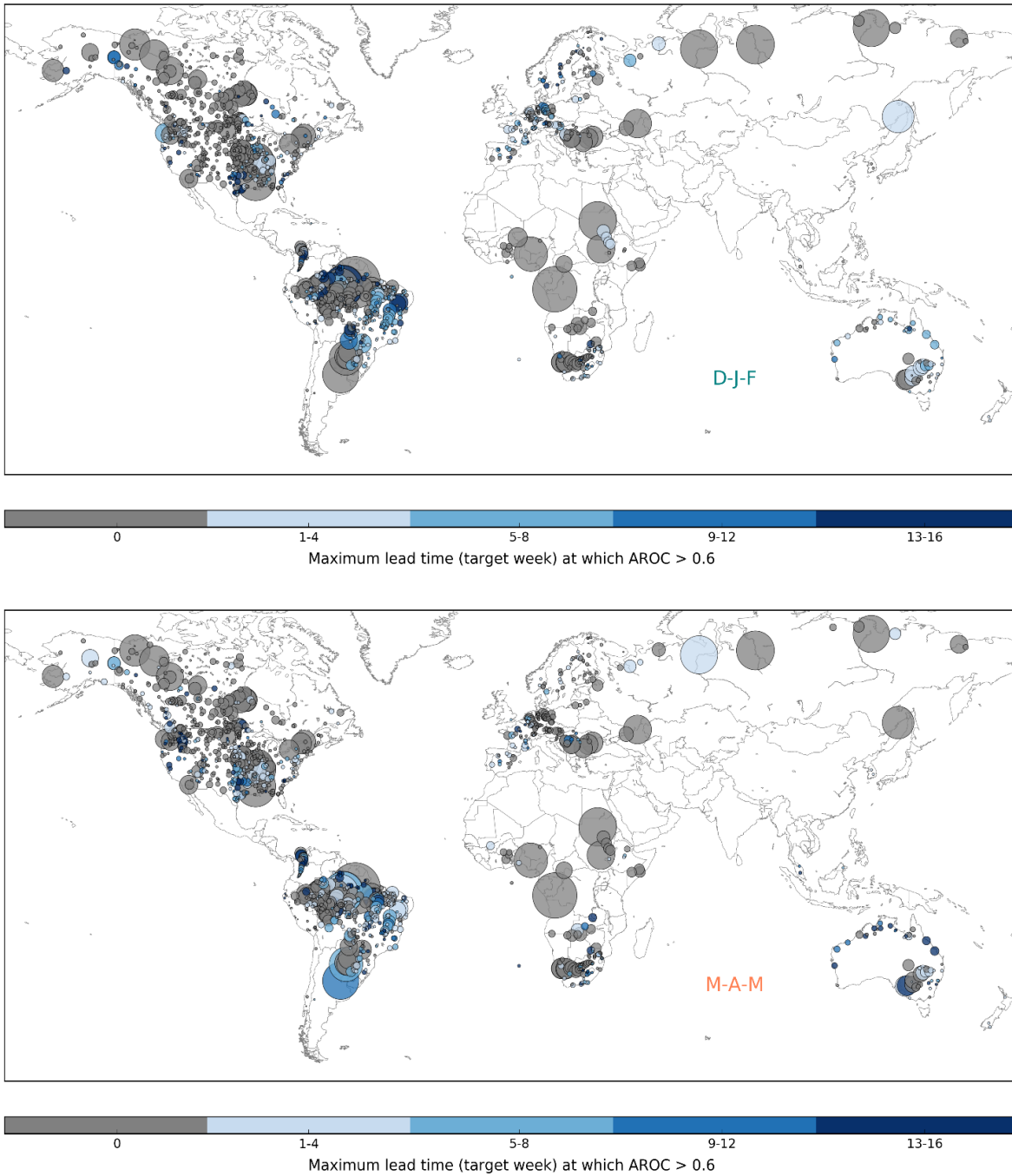
Figure S1 (continued on next page)
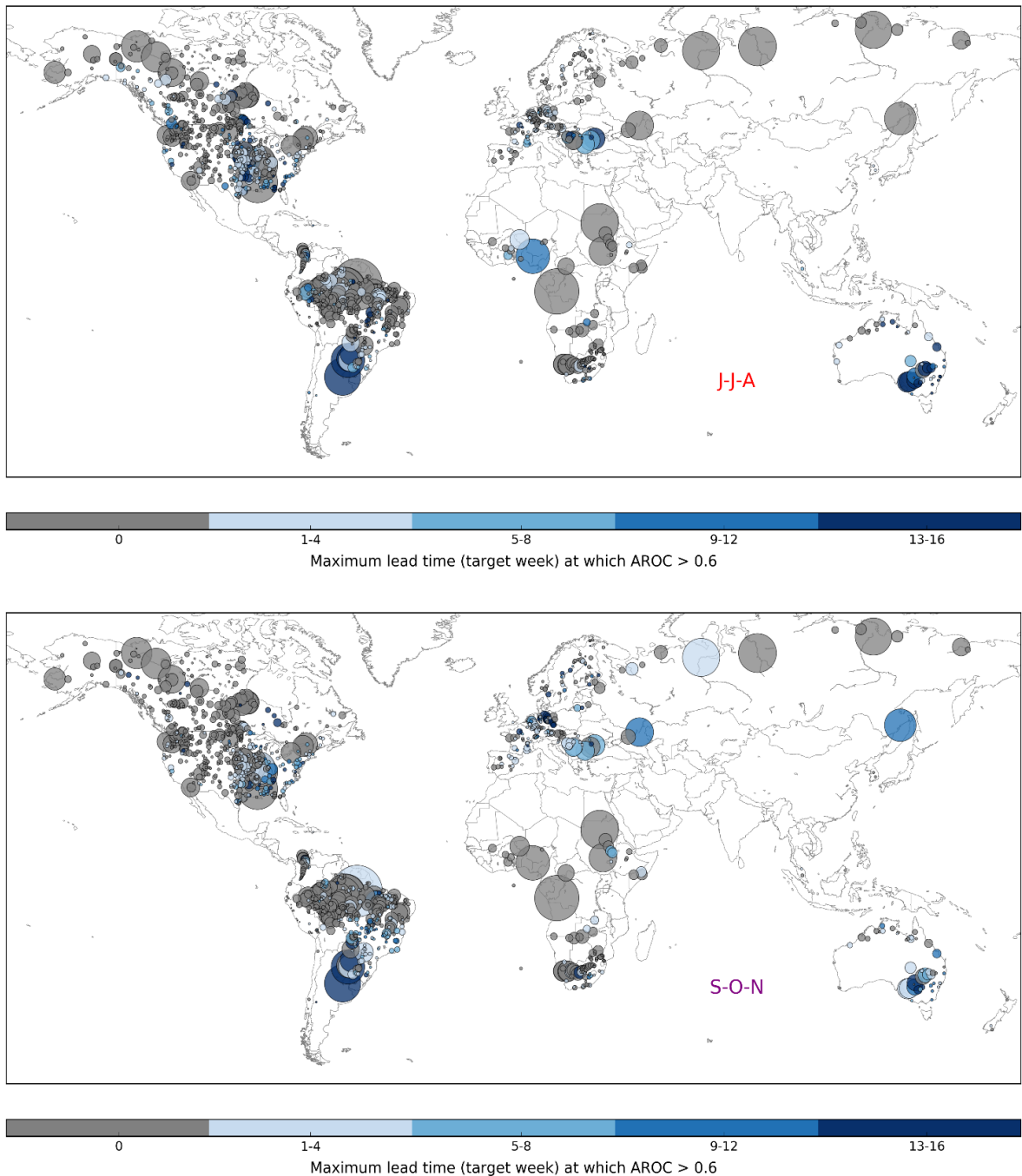
Figure S1: Maximum forecast lead time at which the area under the ROC curve (AROC) is greater than 0.6 for high flow events (flow exceeding the 80th percentile of climatology), at each observation station, for forecasts started in each season. This is used to indicate the maximum lead time at which forecasts are skilful. Grey dots indicate that forecasts started in that season have an AROC < 0.6 at all lead times.

Maximum lead time (target week) at which AROC > 0.6

Maximum lead time (target week) at which AROC > 0.6
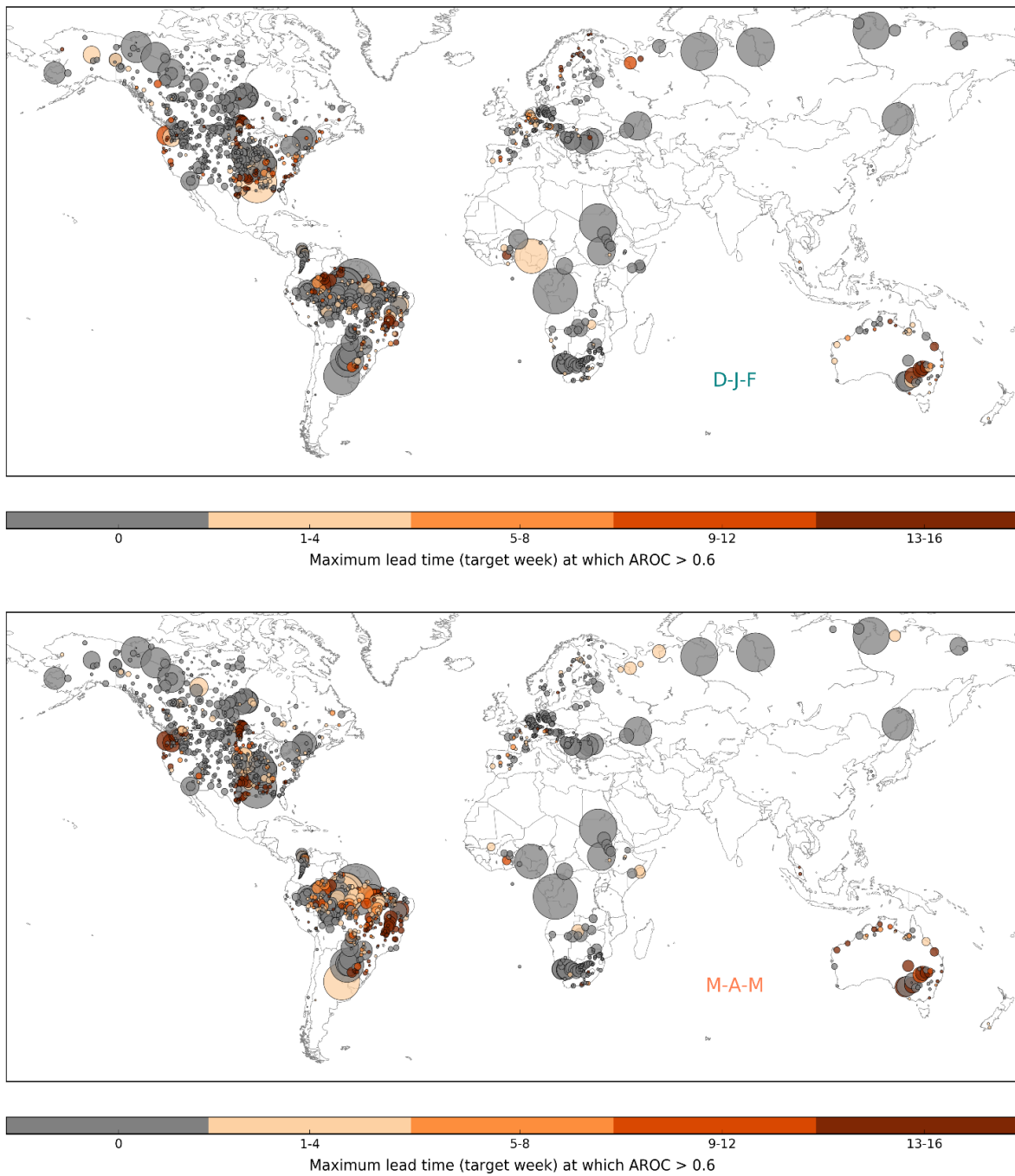
Figure S2 (continued on next page)
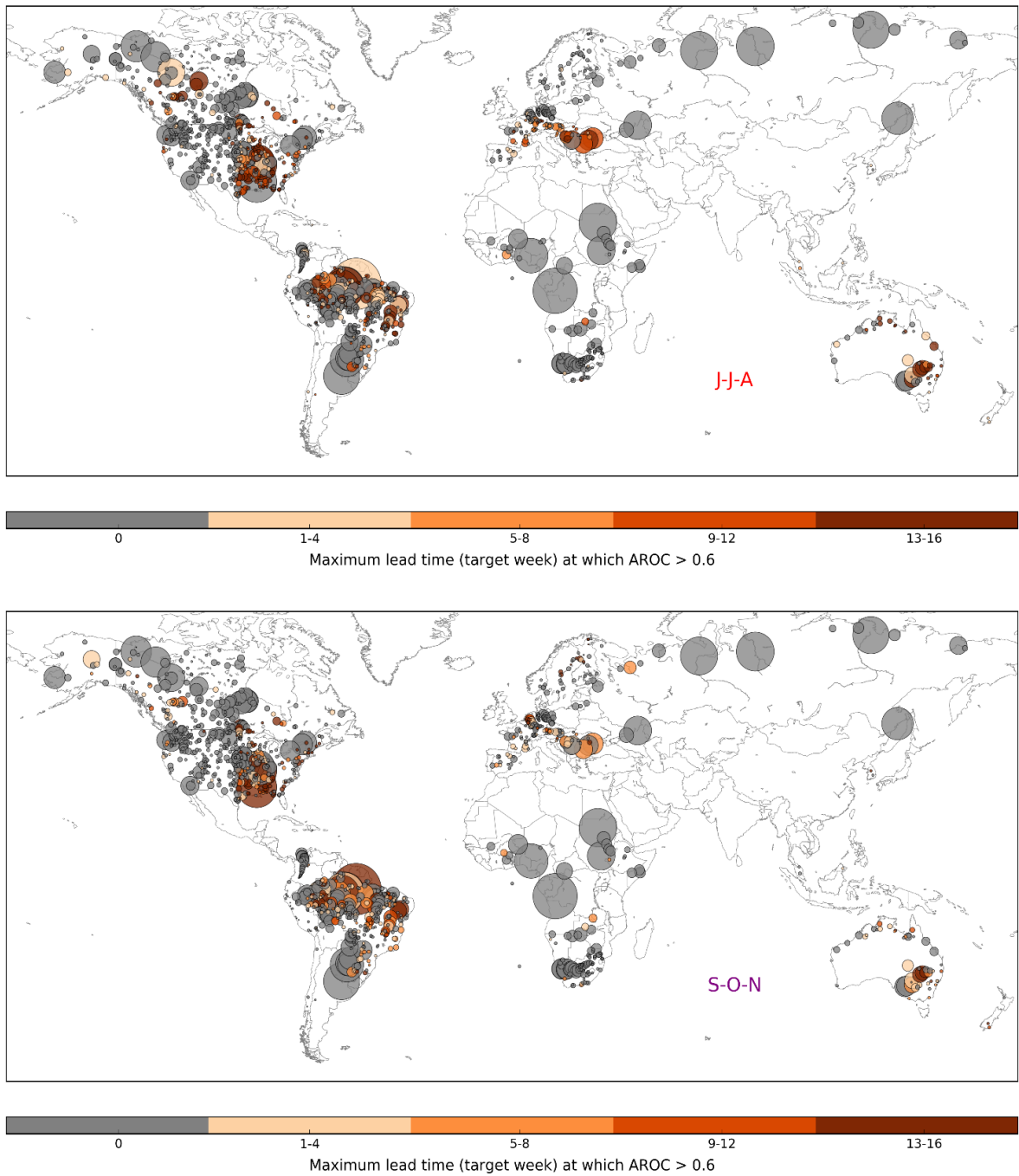
Figure S2: Maximum forecast lead time at which the area under the ROC curve (AROC) is greater than 0.6 for low flow events (flow below the 20th percentile of climatology), at each observation station, for forecasts started in each season. This is used to indicate the maximum lead time at which forecasts are skilful. Grey dots indicate that forecasts started in that season have an AROC < 0.6 at all lead times.
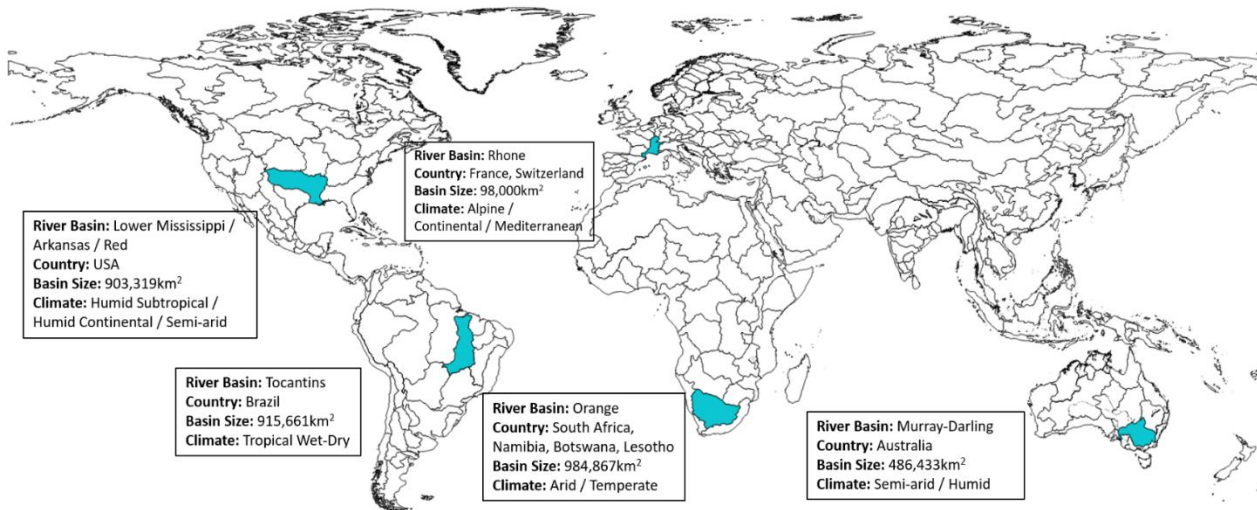
Figure S3: Map of the GloFAS-Seasonal major river basins, highlighting the river basins used for the forecast reliability evaluation.
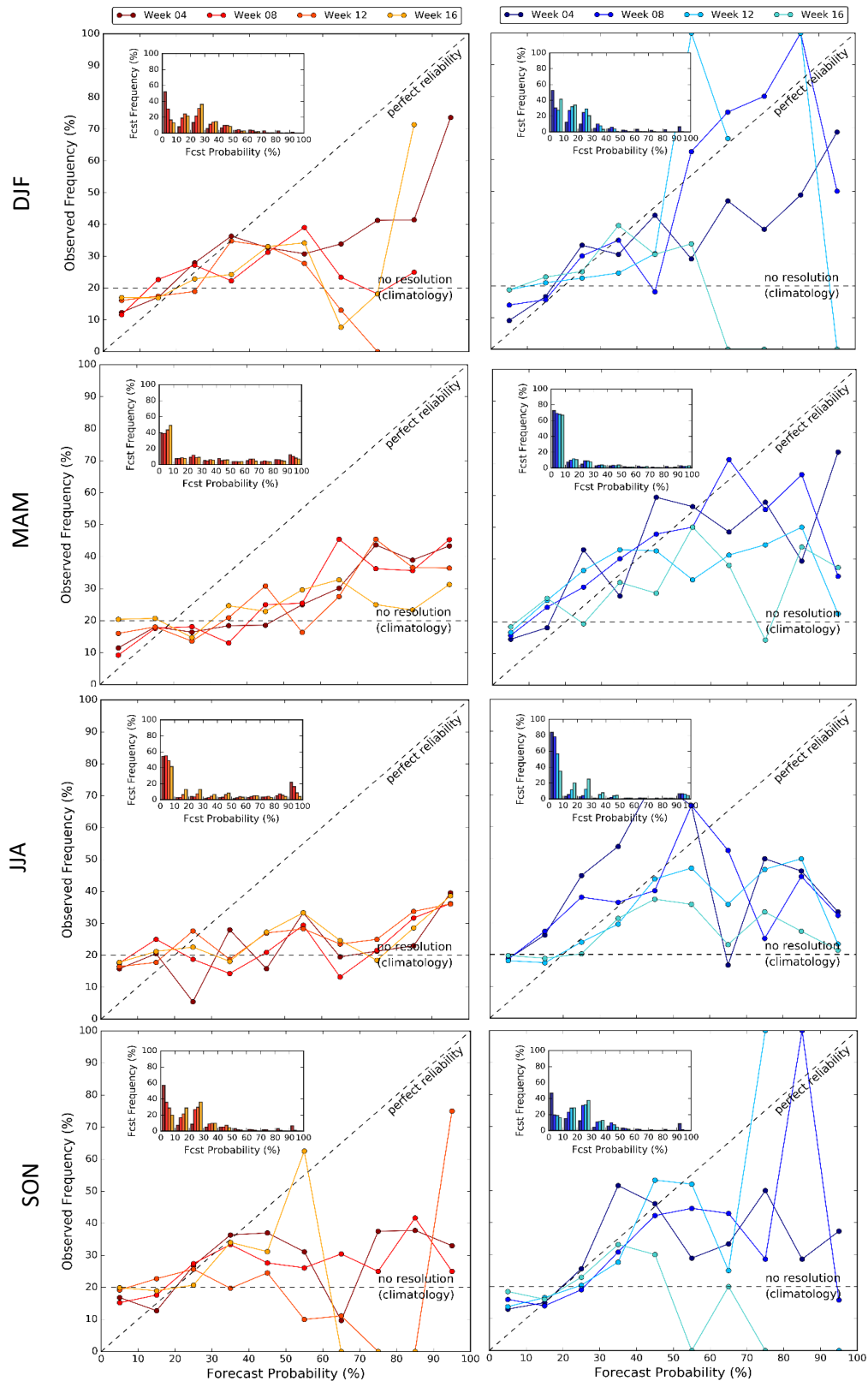
Figure S4: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Tocantins river basin (40 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system
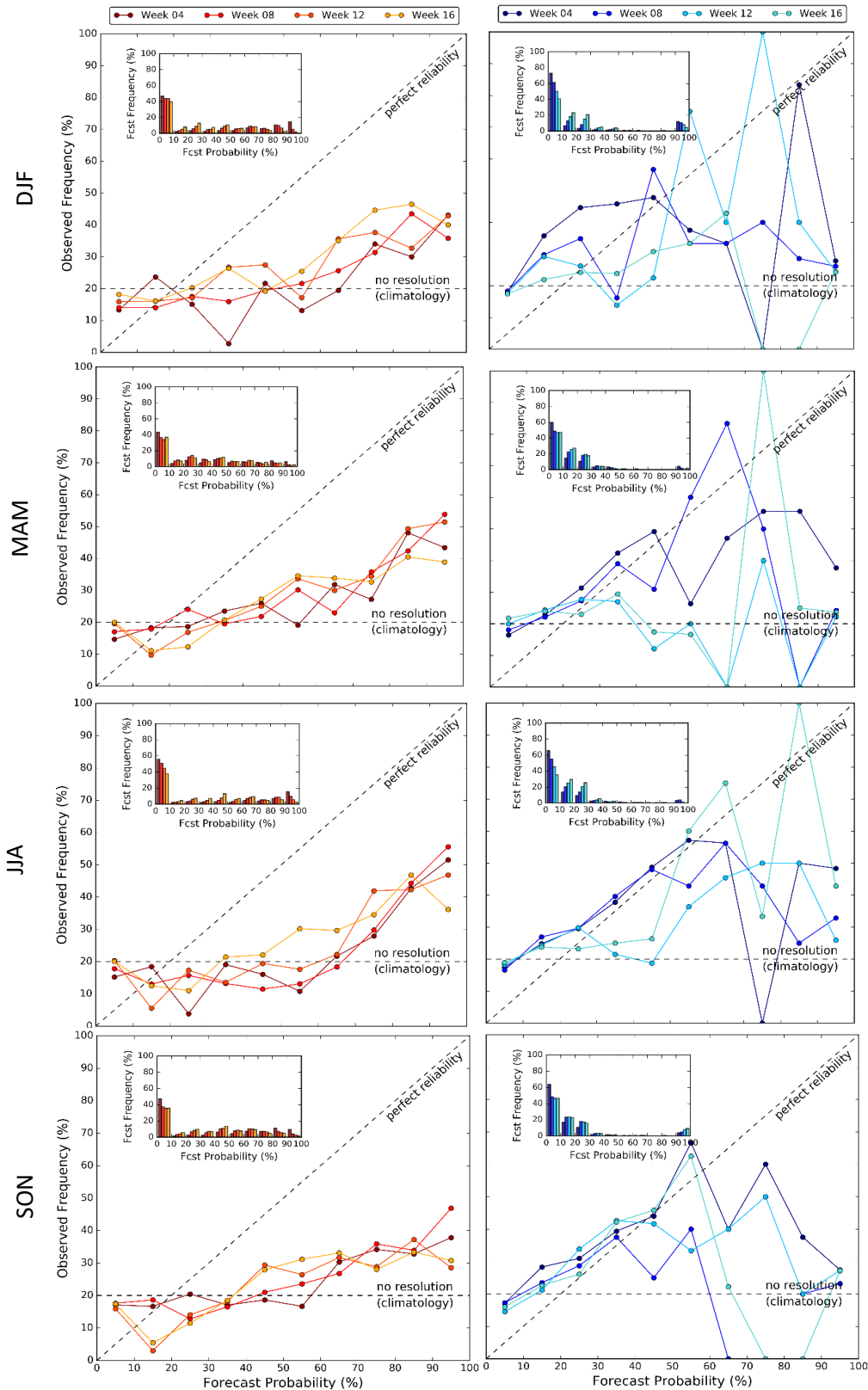
Figure S5: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Lower Mississippi river basin (35 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system
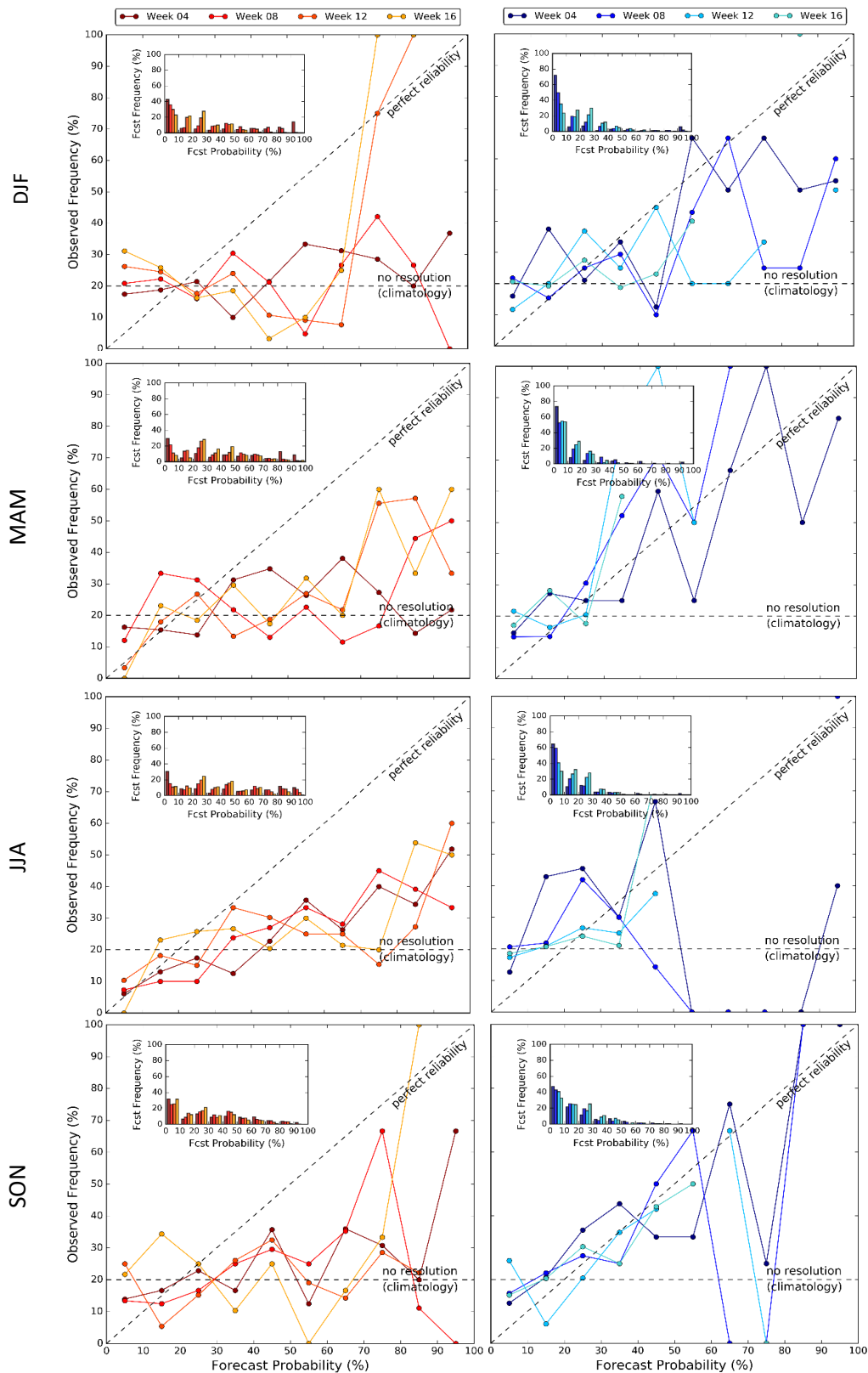
Figure S6: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Rhone river basin (8 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system
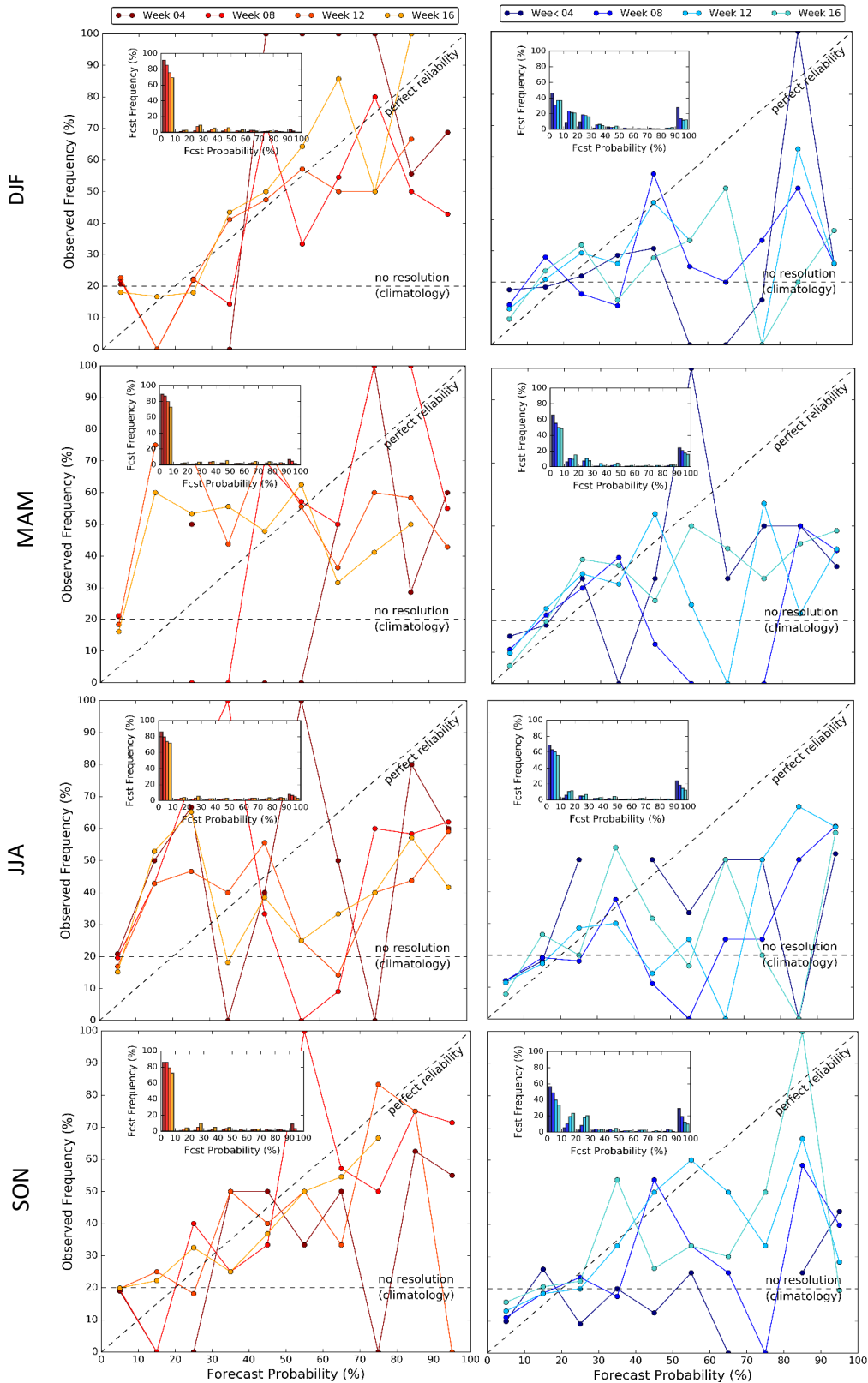
Figure S7: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Murray river basin (12 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system
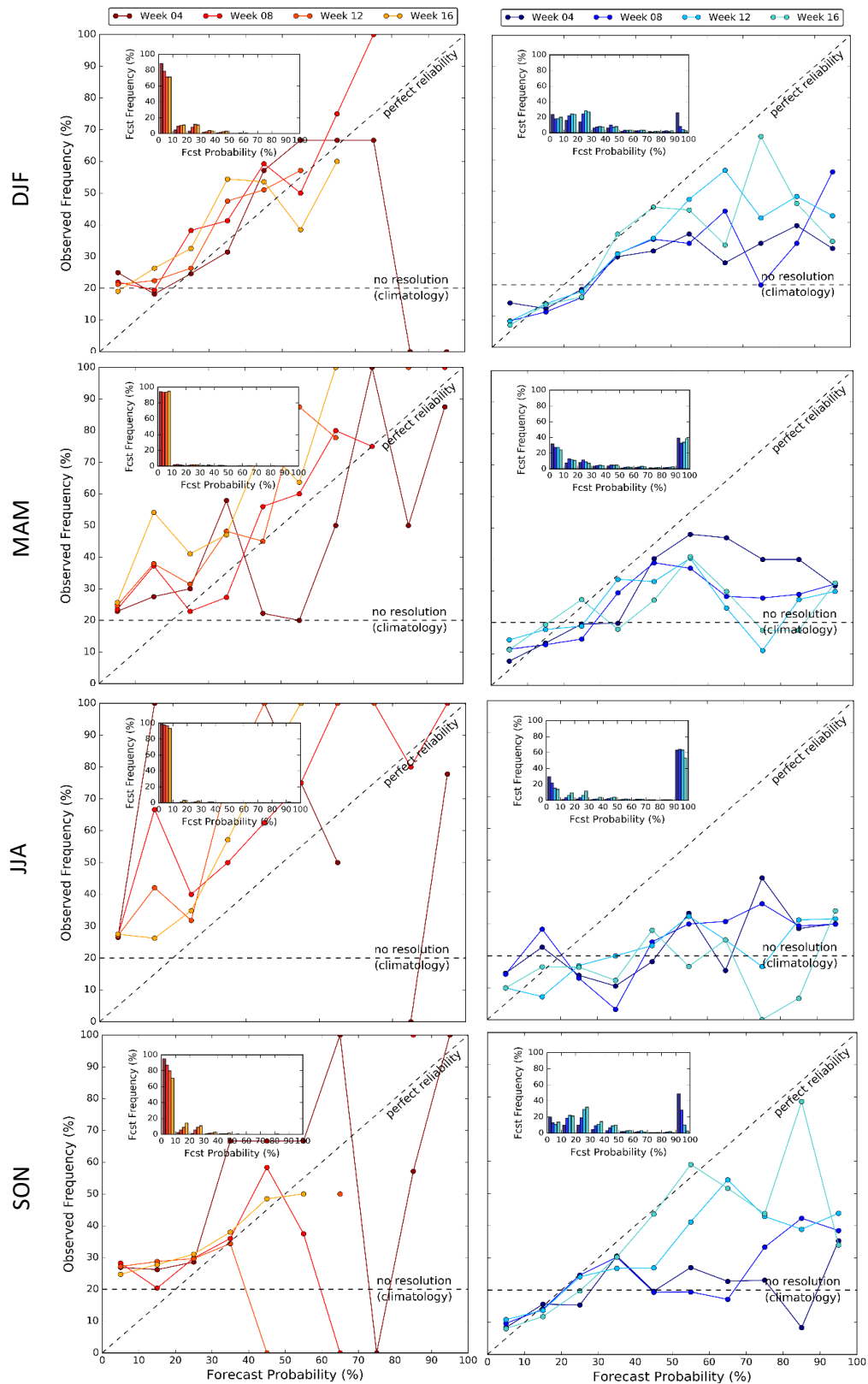
Figure S8: Attributes diagram for forecasts of high flow events (flow exceeding the 80th percentile of climatology, left) and low flow events (flow below the 20th percentile of climatology, right) aggregated across all observation stations in the Orange river basin (46 stations), for each season. Results are shown for lead time weeks 4, 8, 12 and 16, and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin, and are used to indicate forecast sharpness.

Chapter 5. Developing a global operational seasonal hydro-meteorological forecasting system

Chapters 4 and 5 have explored the two key ways in which we can use the inherent predictability of the atmosphere and land surface to extend flood hazard predictability, through statistical analysis based on large-scale climate variability and teleconnections (Chapter 4), and through seasonal forecasting using coupled ocean-atmosphere GCMs (Chapter 5). The following chapter addresses the third aim of this thesis; assessing the potential usefulness of both of these approaches to extending flood hazard predictability at the global scale, for decision-making purposes.

Chapter 6

What is the Most Useful Approach for Forecasting Hydrological Extremes During El Niño?

This chapter has been published in Environmental Research Communications with the following reference:

The contributions of the authors of this paper are as follows: R.E. posed the research question, designed the study with the assistance of E.M.S. and H.L.C., and carried out the analysis. R.E. led the interpretation of the results and writing of the paper, with input from E.M.S. and H.L.C. Overall, 90% of the research and 85% of the writing was undertaken by R.E.

Abstract. In the past, efforts to prepare for the impacts of El Niño-driven flood and drought hazards have often relied on seasonal precipitation forecasts as a proxy for hydrological extremes, due to a lack of hydrologically relevant information. However, precipitation forecasts are not the best indicator of hydrological extremes. Now, two different global scale hydro-meteorological approaches for predicting river flow extremes are available to support flood and drought preparedness. These approaches are statistical forecasts based on large-scale climate variability and teleconnections, and resource-intensive dynamical forecasts using coupled ocean-atmosphere general circulation models. Both have the potential to provide early warning information, and both are used to prepare for El Niño impacts, but which approach provides the most useful forecasts?

This study uses river flow observations to assess and compare the ability of two recently-developed forecasts to predict high and low river flow during El Niño: statistical historical probabilities of ENSO-driven hydrological extremes, and the dynamical seasonal river flow outlook of the Global Flood Awareness System (GloFAS-Seasonal). Our findings highlight regions of the globe where each forecast is (or is not) skilful compared to a forecast of

climatology, and the advantages and disadvantages of each forecasting approach. We conclude that in regions where extreme river flow is predominantly driven by El Niño, or in regions where GloFAS-Seasonal currently lacks skill, the historical probabilities generally provide a more useful forecast. In areas where other teleconnections also impact river flow, with the effect of strengthening, mitigating or even reversing the influence of El Niño, GloFAS-Seasonal forecasts are typically more useful.

6.1 Introduction

Global overviews of upcoming flood and drought events provide valuable information for organisations working at the global scale, across a range of water-related sectors from agriculture to humanitarian aid. Producing such forecasts at the global scale has only become possible in recent years due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations, and increased computer power (Alfieri et al., 2012, 2013; Bierkens, 2015; Brown et al., 2012). While several forecasting centres now produce operational forecasts of floods in the medium-range, up to ~2 weeks ahead (Emerton et al., 2016), earlier indications, many weeks or even months in advance, could be beneficial for water resources and disaster risk management.

Broadly speaking, there are two key ways to extend the predictability of river flow and provide earlier indications of flood hazard: statistical forecasts, typically based on large-scale climate variability and teleconnections, and dynamical forecasts using coupled ocean-atmosphere general circulation models (GCMs).

Operational seasonal forecasts, using both statistical and dynamical approaches, are widely available for meteorological variables, but the hydrology is often not represented, particularly for large or global scales. This means that forecasts of precipitation are often used as a proxy for flooding. However, research has shown that the link between precipitation and flood magnitude is nonlinear (Stephens et al., 2015), and as such, precipitation may not be the best indicator of potential flood hazard (Coughlan De Perez et al., 2017). Recently, there has been an effort to provide the equivalent early awareness information for hydrological variables, as exists for meteorological variables.

Global scale statistical forecasts often rely on ENSO (El Niño Southern Oscillation) teleconnections. ENSO is the largest signal of interannual climate variability (McPhaden et al., 2006); a phenomenon in which sea surface temperatures (SSTs) in the central and eastern equatorial Pacific fluctuate between warm (El Niño) and cool (La Niña) conditions. ENSO is

known to influence various aspects of weather and climate, including river flow (Chiew and McMahon, 2002) and flooding (Ward et al., 2014b, 2014a, 2016), worldwide. Historical probabilities, such as those provided by the International Research Institute for Climate and Society (IRI, 2018) for precipitation and temperature, are an example of a statistical forecast that is often used for El Niño preparedness activities.

In response to a lack of hydrologically-relevant information on ENSO impacts, Emerton et al. (2017) estimated historical probabilities of high and low river flow during El Niño and La Niña. These historical probabilities provide statistical forecasts of extreme river flow, based on the links between past ENSO events and river flow across the globe.

The recent move towards the development of coupled atmosphere-ocean-land models means that it is also now becoming possible to produce seasonal dynamical hydro-meteorological forecasts. The first operational global seasonal river flow forecasting system was implemented in 2017, as part of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013). GloFAS-Seasonal (Emerton et al., 2018) provides openly-available dynamical forecasts of high and low river flow out to 4 months ahead by forcing a hydrological river routing model with seasonal forecast output from a GCM.

Both forecast approaches have the potential to provide early warning information through provision of hydrologically-relevant global scale forecasts, and both are used to prepare for El Niño impacts, but more research is required to explore whether statistical forecasts are able to provide stronger indications of changes in hydrological extremes than seasonal dynamical forecasts.

This study uses river flow observations to compare the potential usefulness of these two global scale forecasts of river flow during El Niño events. Both forecasts are compared to a forecast of climatology and then against each other, using an event-based verification approach.

6.2 Forecasting Approaches

6.2.1 Dynamical Approach: GloFAS-Seasonal

GloFAS-Seasonal provides global scale seasonal hydro-meteorological forecasts using a GCM. Implemented in 2017, it is run by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission Joint Research Centre (JRC), as part of the Copernicus Emergency Management Services. It uses surface and subsurface runoff forecasts from ECMWF's latest seasonal meteorological forecasting system, SEAS5 (ECMWF, 2017a;
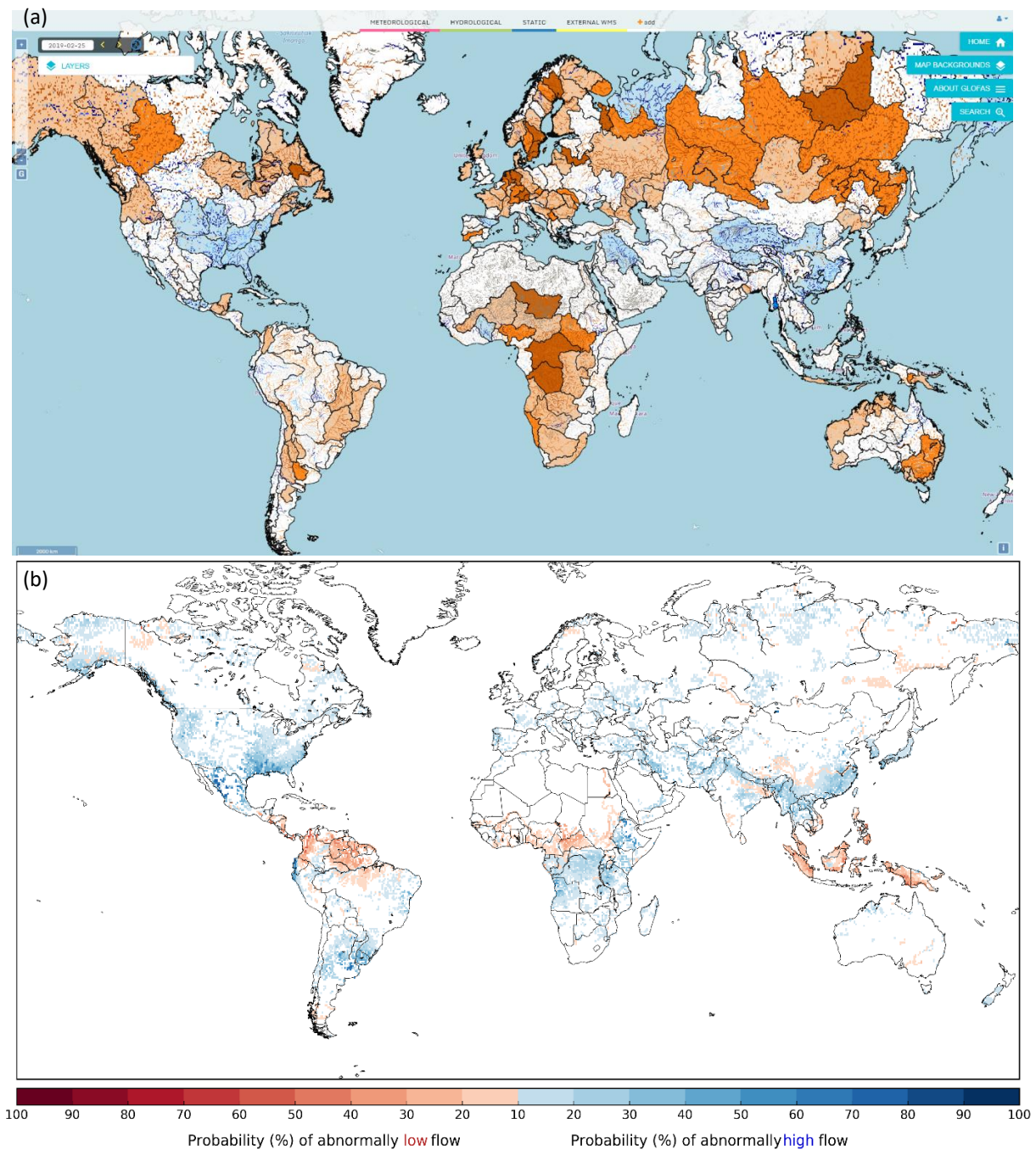
Figure 1: (a) Example of the GloFAS-Seasonal forecast website, displaying the probability of exceeding both the high (blue) and low (orange) river flow thresholds. **(b)** Example of the HistProbs forecast for one week during an El Niño. The map displays the probability of exceeding both the high (blue) and low (red) river flow thresholds. While both examples display forecasts for February during an El Niño event, (a) indicates the maximum probability over the 4-month lead time for a forecast started in February, and (b) indicates the probability for one week in February only.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

Stockdale et al., 2018), to drive a river routing model, Lisflood (Van Der Knijff et al., 2010), producing forecasts of river flow out to 4 months ahead. The GloFAS website (www.globalfloods.eu, see Figure 1a for example) provides seasonal outlooks of the likelihood of exceeding / falling below the climatological thresholds of high (80th percentile) and low (20th percentile) weekly-averaged river flow.

For this study, we make use of the GloFAS-Seasonal *reforecasts*, which were produced using the SEAS5 reforecasts (ECMWF, 2018d; Emerton et al., 2018) initialised with the ERA5-R river flow reanalysis (Emerton et al., 2018). ERA5 (Hersbach and Dee, 2016) is currently still in production, and as such, 34 years of data were available with which to produce the reforecasts: 1981-1983, and 1986-2016.

6.2.2 Statistical Approach: Historical Probabilities

Historical Probabilities (hereafter referred to as HistProbs) provide information about typical El Niño impacts based on historical evidence (Bradley et al., 1987; Mason and Goddard, 2001). The probability of an impact is predicted based on the frequency of occurrence during past El Niños.

The HistProbs of high and low river flow during ENSO events from Emerton et al. (2017) have been reproduced in this study for weekly-averaged river flow, in order to directly compare them with GloFAS-Seasonal. Following the method of Emerton et al. (2017), we used the ERA-20CM-R 10-member, 110-year (1901-2010) river flow climatology to calculate the upper and lower 20th percentile of river flow for each grid . We then calculate, for each week of an El Niño, the percentage of historical El Niños during which the high or low flow threshold was exceeded. The use of ERA-20CM-R allows for more El Niños to be included in the calculation of the HistProbs, with 30 El Niños identified over the 110-year period. An El Niño is identified when the SST anomaly in the central equatorial Pacific Ocean (Niño3.4 region; 5°S - 5°N, 170°-120°W) exceeds +0.5°C for at least five consecutive (overlapping) three-month periods.

The HistProbs (Figure 1b) were estimated for each grid point, through calculation of the percentage of the 30 historical El Niños in which the river flow exceeded the high flow threshold, or fell below the low flow threshold, during the same week. This was repeated for each of the 10 ensemble members of ERA-20CM-R. The ensemble mean probability was then interpolated from the 0.5° (~50km) resolution of ERA-20CM-R, to the 0.1° (~10km) resolution of GloFAS-Seasonal; it is this higher-resolution ensemble mean that is used throughout this study.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

6.3 Evaluation Data and Methods

This study evaluates the predictability of hydrological extremes during El Niño in both GloFAS-Seasonal and the HistProbs by assessing the ability of each system to predict high and low river flow, with the correct timing, during an El Niño. The ability of a forecast to predict events of the correct category is referred to as the "potential usefulness" and is of particular importance for decision-making purposes (Arnal et al., 2018).

The potential usefulness is calculated using the relative operating characteristic (ROC) curve, based on ratios of the probability of detection (POD) and the false alarm rate (FAR) (Mason and Graham, 1999). These ratios are calculated by assessing whether a forecast correctly predicted an observed event, or whether it missed the event or provided a false alarm, and allow for estimation of the probability that an event will be predicted. The POD (eq. 1) and FAR (eq. 2) are calculated as follows:

$$POD = \frac{\text{hits}}{\text{hits+misses}} \tag{1}$$

$$FAR = \frac{\text{false alarms}}{\text{false alarms+correct negatives}} \tag{2}$$

where a *hit* is defined when the forecast correctly predicted flow exceeding [falling below] the 80th [20th] percentile during the same week that the observed river flow exceeded [fell below] the 80th [20th] percentile of the observations at that location. It follows that a *miss* is defined when an event was observed but the forecast did not exceed the threshold, a *false alarm* when the forecast exceeded the threshold but no event was observed, and a *correct negative* when no event was observed and the forecast did not exceed the threshold.

The ROC curve is constructed from the FAR (horizontal axis) and POD (vertical axis) at different probability thresholds (in this case, in 10% bins), therefore providing information on the likelihood that an event will be predicted at a given probability threshold. The geometrical area under the ROC curve (AROC; $0 \leq AROC \leq 1$) provides a summary statistic for the performance of a probabilistic forecast, where a forecast that correctly predicts every observed event (with no recorded false alarms or missed events) would have an AROC of 1. An AROC < 0.5 indicates that the skill of the forecasts is less than a forecast of climatology, which has an AROC of 0.5.

The AROC is used to infer the potential usefulness of the forecast; a forecast that is more skilful than a forecast of climatology is said to be potentially useful, whereas a forecast that is less skilful

than a forecast of climatology is not useful. This approach has previously been used in the evaluation of seasonal river flow forecasts (Arnal et al., 2018; Emerton et al., 2018). Often, seasonal forecasts are provided in terms of the likelihood that a given variable will be above or below normal (based on terciles) in the coming months. The evaluation technique used in this study presents a significant challenge for both forecasting systems, requiring that they predict more extreme weekly-averaged river flow, in the same week as that in which it was observed, several weeks to months ahead.

6.3.1 Observed Data

The two forecasts are evaluated over the same 34-year period (1981-2015), using river flow observations obtained from the Global Runoff Data Centre (GRDC; BfG, 2017), alongside observations that have been made available to GloFAS (Emerton et al., 2018). To ensure a large enough sample size for the forecast evaluation, alongside the best possible spatial coverage, the following criteria are applied to the data:

- The weekly-averaged river flow record at each station must contain data for at least 50% (17 years) of the evaluation period, in order to calculate the observed high and low flow thresholds (80th and 20th percentiles) for each station, and for each week of the year.

- The weekly-averaged river flow record at each station must contain at least 6 El Niños over which to evaluate the forecasts.

- The upstream area of the corresponding grid point in the model river network must be at least 1500km^2.

Data from human-influenced rivers have not been removed, as we are interested in identifying the ability of both forecasting approaches to predict observed events, rather than their ability to represent natural flow. Of the 2355 stations in the database, ~1250 contain enough data to meet the above criteria and are used in this study.

6.3.2 Calculating Potential Usefulness of GloFAS-Seasonal

To evaluate the potential usefulness of GloFAS-Seasonal we calculate the AROC for each season during an El Niño using the observations as a benchmark. The AROC for a season is calculated by grouping together forecasts for every week during the season for all 11 El Niño events between 1981 and 2015.

The AROC is also calculated for lead times of 1-4 months ahead, by selecting the GloFAS-Seasonal weekly-averaged river flow forecast that would have been available 1, 2, 3 and 4 months ahead of each week of the El Niño event. For example, for the fourth week in January the forecast available one month ahead would be the fourth week of the forecast produced at the start of January, the forecast available two months ahead would be the 8th week of the forecast produced in December, and three months ahead the 12th week of the forecast produced in November. Following the same method, for the second week in December, the forecast available one month ahead for that week, would be the 6th week of the forecast produced in November. This is necessary because while GloFAS-Seasonal predicts weekly-averaged river flow, the forecasts are updated just once per month.

6.3.3 Calculating Potential Usefulness of the Historical Probabilities

To evaluate the potential usefulness of the HistProbs we calculate the AROC for each season during an El Niño event using the observations as a benchmark.

The HistProbs are a "static" forecast, that is, the forecasts do not change with lead time and there is just one probability for high or low river flow during each week of an El Niño. As such, the AROC is calculated by comparing the river flow in each week of the 11 El Niño events in the observations, with the HistProb of high or low river flow for the corresponding week of the year. The AROC for a season is calculated by grouping together forecasts for every week during the season, for all 11 El Niño events between 1981 and 2015.

6.4 Results

The results presented in this section compare the "potential usefulness" of both GloFAS-Seasonal and the HistProbs during an El Niño. The following criteria are used to define the "most useful" forecast, based on the null hypothesis that the potential usefulness of the two forecasts is not significantly different:

- If GloFAS-Seasonal has an AROC > 0.5 and the HistProbs < 0.5, or both exceed 0.5 but GloFAS-Seasonal has an AROC > 0.1 larger than the HistProbs, *GloFAS-Seasonal* is most useful

- If the HistProbs have an AROC > 0.5, and GloFAS-Seasonal < 0.5, or both exceed 0.5 but the HistProbs have an AROC > 0.1 larger than GloFAS-Seasonal, the *HistProbs* are most useful

- If both forecasts have an AROC > 0.5, and within 0.1 of each other, both are useful and *similar*

- If both forecasts have an AROC < 0.5, *neither* are useful

The statistical significance of the difference in AROC between the two forecasts was investigated using a bootstrap procedure. For each season and each observation location, all available forecasts for both GloFAS-Seasonal (132 forecasts per season across the 11 El Niño events, at each lead time of 1-4 months ahead) and the HistProbs (143 forecasts per season, providing an independent probability for each week of the season, but the same probability for a given week across all 11 El Niño events), were resampled with replacement, and the resulting AROC was calculated. This process was repeated 1000 times.
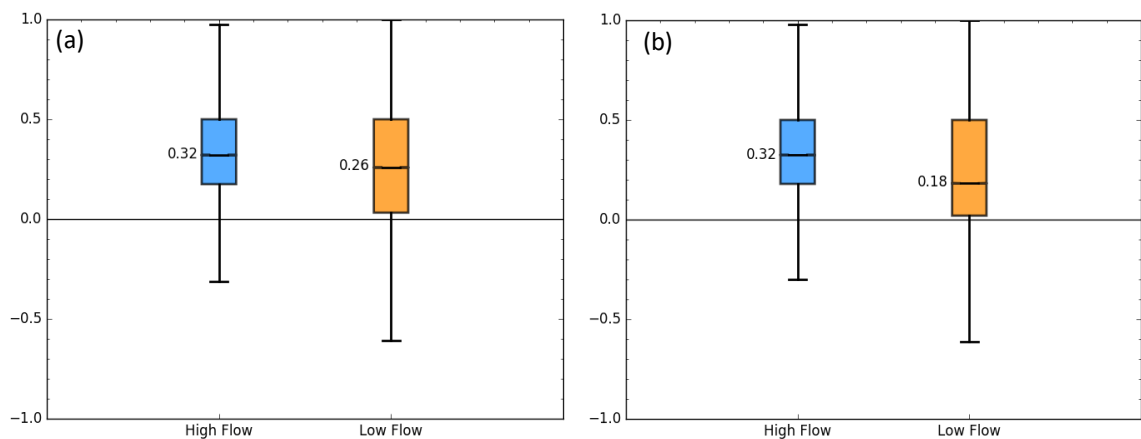
Figure 2: Box plots of the AROC differences (GloFAS - HistProbs) at lead times of **(a)** 1, and **(b)** 3 months ahead for both high (blue) and low (orange) river flow in MAM globally (for stations where at least one of the forecasts has an AROC > 0.5), calculated from a bootstrap procedure that was repeated 1000 times using resampling of the 132 [144] GloFAS-Seasonal [HistProbs] forecasts, with replacement. The bottom and top of the boxes correspond to the 25th and 75th percentiles, respectively. The notch represents the 95% confidence interval around the median from a 1000-bootstrapped sample.

Figure 2 displays box plots of the global bootstrapped AROC differences (GloFAS-Seasonal - HistProbs) at lead times of 1 and 3 months ahead for high and low river flow in MAM during an El Niño. These results indicate that, aggregated globally, there is evidence that GloFAS-Seasonal provides an improved AROC for forecasts of both high and low river flow, however, this is not statistically significant. For high [low] flow 3 months ahead, the median AROC difference is 0.32 [0.18], across all stations where at least one of the two forecasts is potentially useful (AROC > 0.5). Further assessment of the bootstrapped AROC differences for each individual station indicates that at ~95.5% of the locations where the median AROC difference

of the 1000-bootstrapped sample exceeds ±0.1, the choice of the most useful forecast is statistically significant to the 95% confidence level (at ~4.5% of stations, this is not the case, and using a threshold of ±0.1 does not provide a statistically significant result). At locations where the median AROC difference is <0.1, choosing a 'most useful' forecast would not provide a statistically significant result, and therefore it is reasonable to class the forecasts as 'similar' (or 'not useful' depending on the AROC values).

6.4.1 Probability of High Flow

Figure 3a indicates that for forecasts of high river flow 3 months ahead, for MAM during an El Niño, the most useful forecast varies by region, and there are many locations where neither forecast is more skilful than a forecast of climatology (grey dots).

Across much of North America, the HistProbs provide a more useful forecast of high river flow than GloFAS-Seasonal, except along the east coast, where GloFAS-Seasonal forecasts are more skilful. In the regions of South America that are more likely to see high flow during an El Niño, GloFAS-Seasonal is more useful at several locations, particularly in northern Peru, while the HistProbs are more useful in southern Brazil. In Europe, the HistProbs are more useful in the west, and GloFAS-Seasonal is more useful in the east.

Figure 4 shows the AROC values for each forecast at locations where they are more skilful than climatology. Generally, the AROC for the HistProbs lies in the 0.5-0.6 range, meaning they are only marginally more skilful than climatology, except in some small regions, such as north-west USA where the AROC reaches 0.7-0.8. There are also regions where GloFAS-Seasonal forecasts are only marginally more skilful than climatology, such as the east coast of North America, but the majority of locations show an AROC of 0.6-0.8.

Results for all seasons and lead times are provided in the supplementary material. In general, the results tend to be consistent with lead time, although as may be expected, the skill of GloFAS-Seasonal is reduced at longer lead times in some locations. The skill of both forecasts varies more significantly with season than with lead time. Figure S1 shows that areas where neither is useful are more widespread in JJA, when El Niño typically begins to develop, and both become more widely skilful through SON and DJF as El Niño intensifies. The timing of El Niño onset varies from one event to the next, which results in more uncertainty in the HistProbs for JJA than for other seasons. For GloFAS-Seasonal, forecasts made ahead of JJA are likely to be more uncertain due to uncertainty in forecasting the timing and magnitude of El Niño. Forecasts of El Niño produced before and during spring tend to be much less successful

(the infamous "spring predictability barrier"), although the cause of this remains controversial (Barnston et al., 2012; Duan and Wei, 2013; McPhaden, 2003; Wang-Chun Lai et al., 2018).
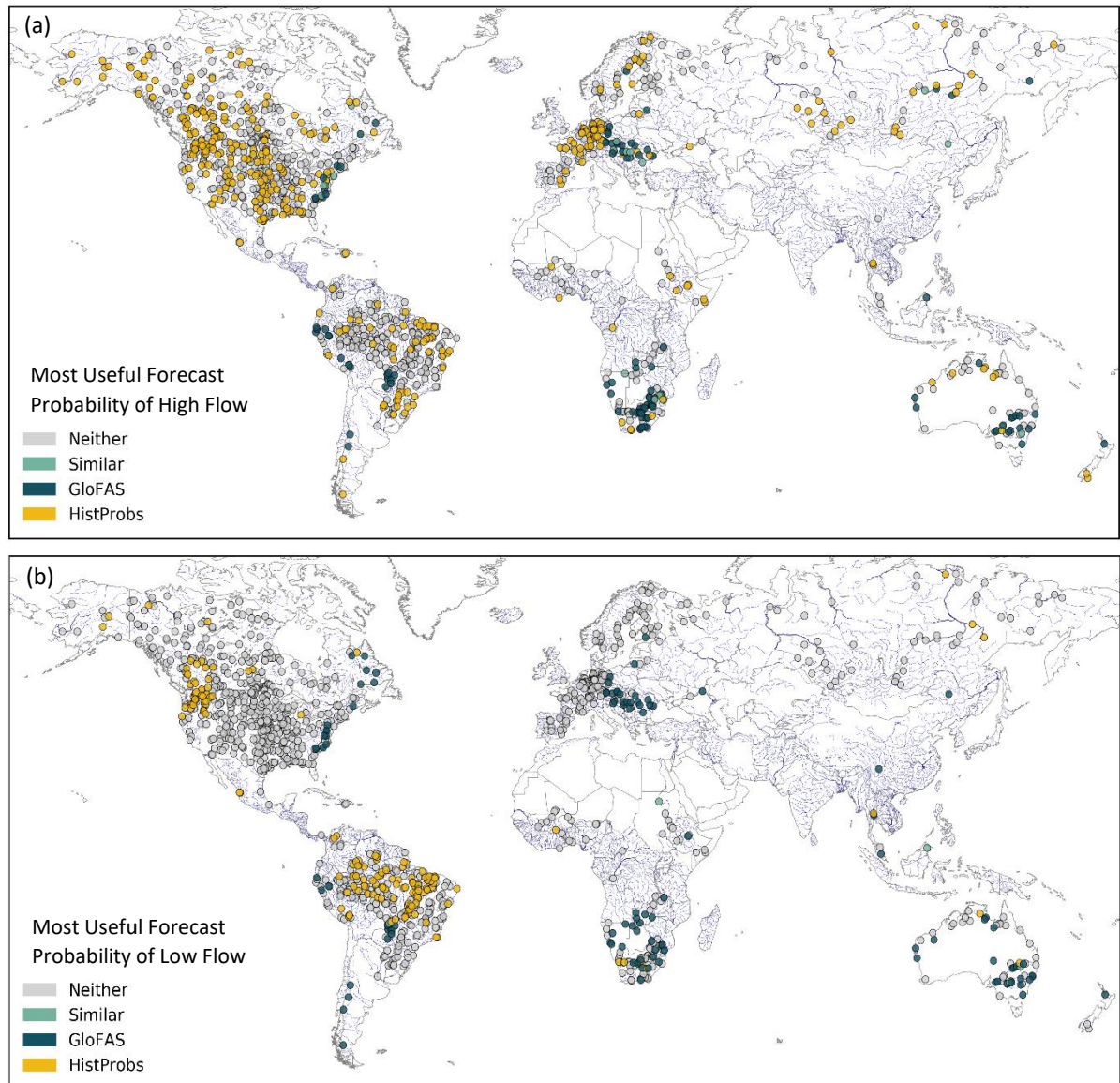
Figure 3: Maps indicating the most potentially useful forecast 3 months ahead for **(a)** high river flow (>80th percentile of climatology) and **(b)** low river flow (<20th percentile of climatology) in MAM, at each observation location.

6.4.2 Probability of Low Flow

Figure 3b provides the same results for forecasts of low river flow. Locations where neither forecast is more skilful than climatology are more widespread. However, some of these regions, such as the USA, are more likely to see high river flow during an El Niño.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

In the low flow regions in the USA, South America, Africa and Australia, there are locations at which the HistProbs are potentially useful (see Figures 2b and 3a), but the variability from one location to the next is much higher than for forecasts of high river flow. The skill of the HistProbs increases during and after the peak of El Niño, in DJF and MAM. This is likely due to the delayed response of river flow to the El Niño-driven precipitation, which is more prominent for low flow and drought, than for high flow and flooding. This is also reflected in the HistProbs themselves (not shown), which highlight the lagged response of river flow to El Niño, and that the influence on rivers can continue beyond the return to neutral ENSO conditions.

In general, GloFAS-Seasonal is the most useful forecast for low river flow in the same regions as for high flow, while the HistProbs are more useful over the Amazon basin and north-west USA, particularly in DJF and MAM. Interestingly, Figure 4 indicates that for low river flow, the AROC values for the two forecasts tend to be very similar; within ±0.2. The GloFAS-Seasonal AROC values are similar to those for high river flow, reaching 0.6-0.8 in many locations, but where the HistProbs are potentially useful, the AROC can also reach 0.6-0.7, and 0.8 at some locations. As with the forecasts for high river flow, some variations in the results are seen with lead time, but these are less significant than the variations from one season to the next. Additional results for all seasons and lead times are provided in the supplementary material.

6.4.3 Discussion

The results presented in sections 4.1 and 4.2 highlight areas of the globe where potentially useful forecasts of hydrological extremes during El Niño are available, and indicate that the skill of both forecasts varies by region and season, and to some extent with lead time.

Overall, where there is a strong El Niño influence on river flow the HistProbs are able to provide a potentially useful forecast of high flow in regions where GloFAS-Seasonal lacks skill. The HistProbs presented here are estimated based only on SSTs in the Niño3.4 region in the central Pacific, and therefore are not able to reflect ENSO diversity. For example, flooding in Peru is known to be driven by El Niños which exhibit larger SST anomalies in the eastern Pacific than the central Pacific.

In fact, the impact of ENSO diversity provides some indication as to why GloFAS-Seasonal is more useful than the HistProbs in specific regions (e.g. northern Peru, east coast of North America, southern Africa, eastern Europe and Australia). All of these regions are similarly, if not more strongly, influenced by other modes of climate variability, such as the Indian Ocean
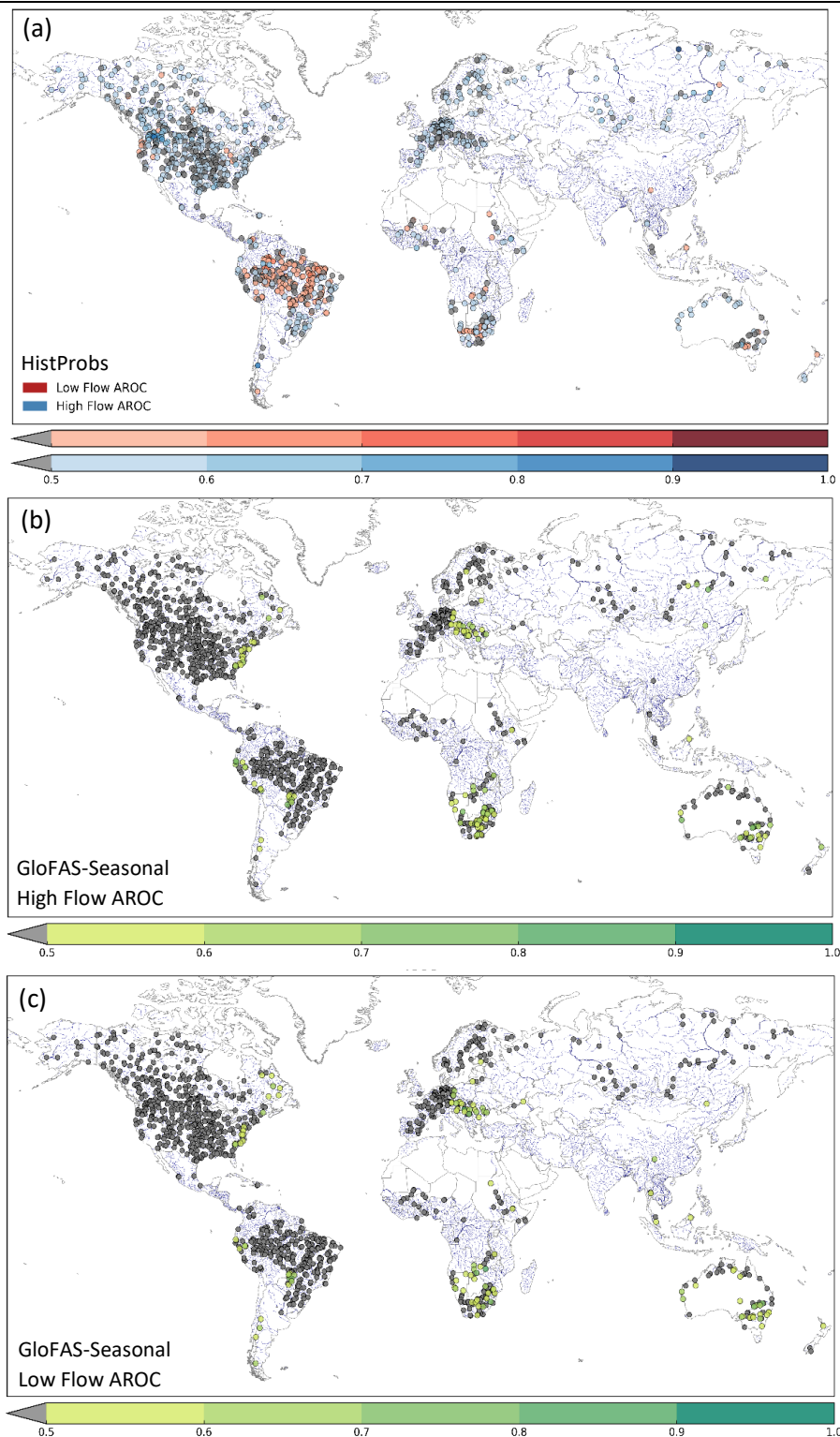
Figure 4: Maps indicating **(a)** the AROC of the HistProbs for both high river flow (>80ᵗʰ percentile of climatology, blue) and low river flow (<20ᵗʰ percentile of climatology, red) in MAM, **(b)** the AROC of GloFAS-Seasonal 3 months ahead for high river flow in MAM, and **(c)** the AROC of GloFAS-Seasonal 3 months ahead for low river flow in MAM. On all 3 maps, the darker the colour, the higher the skill (and potential usefulness) of the forecast. Grey dots indicate that the forecast is not useful at that location; i.e. the forecast has an AROC ≤ 0.5.

Dipole (IOD), Pacific Decadal Oscillation (PDO), and North Atlantic Oscillation (NAO). A NWP model, by design, should be able to represent the impact of these other modes of variability on weather patterns.

Wang et al. (2015) show that generally, an El Niño combined with a warm phase PDO gives a similar, but stronger, pattern of influence on wet-dry anomalies. However, in some regions the wet-dry anomaly during El Niño is reversed when combined with a cold phase PDO. In regions where the impact is similar regardless of the PDO phase, the HistProbs are generally more useful than GloFAS-Seasonal, particularly for high flow. Regions where the wet-dry anomaly is reversed depending on the PDO phase, tend to correspond to those where GloFAS-Seasonal is more useful. There are some exceptions, however, such as high latitude Canada and Siberia, where the HistProbs are more useful. These correspond to regions where GloFAS-Seasonal has been shown to generally be less skilful than climatology (Emerton et al., 2018). As the PDO is a decadal oscillation varying on much longer timescales than ENSO, it is likely to influence El Nino impacts over several events in turn. It is therefore a potential source of uncertainty in the HistProbs (see Emerton et al., 2017), as they are conditioned only on ENSO, and a change in the PDO may represent a change in the climate state from the period over which the HistProbs are estimated. The state of the PDO, however, is accounted for within a dynamical seasonal forecasting system.

Further regions where GloFAS-Seasonal tends to provide a more useful forecast, for both high and low river flow, include southern Africa and Australia, which are known to be influenced by the IOD (Behera et al., 2005; Hoell et al., 2017; Marchant et al., 2007; Washington and Preston, 2006). Saji and Yamagata (2003) show that the IOD impacts African rain variability regardless of the ENSO phase, but ENSO only has an impact when combined with an IOD event. As mentioned previously, the skill can vary significantly by season, and recent research (MacLeod, 2018) has also shown that SEAS5, the meteorological forecast input of GloFAS-Seasonal, is more skilful at predicting short rains (OND) than long rains (MAM) in east Africa, as the short rains have much stronger teleconnections with ENSO and the IOD than the long rains. In Australia and south-east Asia, the IOD increases [decreases] the chance of rainfall during its negative [positive] phase (Ashok et al., 2003). Additionally, the NAO has been shown to influence flood occurrence in Europe, with extreme rainfall more likely in parts of eastern Europe during the positive phase of the NAO (Guimarães Nobre et al., 2017).

While the HistProbs are able to, in general, provide a more skilful forecast than climatology in the majority of regions influenced by El Niño, there are locations where GloFAS-Seasonal is less skilful than climatology in all seasons and at all lead times. In these locations, GloFAS-Seasonal is unable to correctly predict the magnitude, and/or the timing, of the observed events. A study by Hirpa et al. (2018) identifies regions of bias in GloFAS river flow simulations. Regions of negative bias generally correspond to those where GloFAS-Seasonal is not skilful in this study. Future work should determine whether calibration of GloFAS, such as that presented by Hirpa et al. (2018) for the medium-range GloFAS forecasts, could improve the skill of the seasonal forecasts. As GloFAS-Seasonal is further developed, it will also be important to consider a wider range of skill metrics for verification, taking into account both the skill and the value of the forecasting system (Cloke et al., 2017). The evaluation technique used in this study presents a significant challenge for both forecasting systems, requiring that they predict high or low weekly-averaged river flow, in the same week as that in which it was observed, several weeks to months ahead.

Prediction of El Niño events is also key for both types of forecast. As a dynamical model, GloFAS-Seasonal incorporates forecasts of SSTs and therefore ENSO. Decision-makers often rely on forecasts of El Niño before consulting forecasts such as the HistProbs, when an El Niño event is forecast or developing. ECMWF's seasonal forecasts of ENSO events are world-leading (Barnston et al., 2012; ECMWF, 2018d), and SEAS5 represents an improvement in the skill of these forecasts over the previous version of the forecasting system, S4. However, there is a decrease in the skill of the IOD in SEAS5, with forecasts producing cold events that are too large and too frequent, alongside a slight deterioration in the skill of upper level winds (ECMWF, 2018d), which are important for representing teleconnections across the globe. While dynamical models are better able to represent the complex interactions between the various modes of climate variability and their associated teleconnections by design, it is still possible that the evolution of El Nino may be uncertain or incorrectly predicted, or that even a perfect forecast of El Nino evolution may poorly simulate the teleconnections due to the nonlinearity of the teleconnections and their impacts. This can have important implications for seasonal predictability of ENSO teleconnections using GCMs (Turner et al., 2005).

A further point of consideration is that while this study makes use of >1200 river flow observation stations around the globe, there are large areas of the world, including some that are significantly impacted by El Niño, where there is very sparse to no data coverage. At many of the stations used, management of water resources will be evident in the river flow records,

particularly during periods of low flow conditions, and this is likely to affect the evaluation results.

Statistical forecasts such as the HistProbs are limited in that they can only forecast the response to events which we have previously observed. With recent research suggesting that the frequency of extreme El Niño events, such as those in 1982-83, 1997-98 and 2015-16, is likely to increase with future climate change (Cai et al., 2014, 2015a), this limitation could become more and more relevant. The HistProbs were also estimated using the longer ERA-20CM-R dataset. This dataset provides more El Niños over which to calculate the probabilities, and has been shown to represent ENSO teleconnections, but is unable to reproduce synoptic situations as no atmospheric observations were assimilated (Hersbach et al., 2015). Future work should explore whether the skill of statistical forecasts such as the HistProbs could be improved using different reanalysis products, such as ERA5.

While currently there are areas of the globe where GloFAS-Seasonal is less skilful than climatology, this is the just the first version of the first global scale operational seasonal river flow forecasting system. Future improvements to the input datasets (e.g. topography, river flow observations, lakes and reservoirs), seasonal precipitation forecasts and hydrological models could result in a dynamical forecasting system that consistently provides a more useful forecast of hydrological extremes, with the benefit that such dynamical forecasts are not constrained to periods of time when there is an El Niño. A third approach, not considered in this study, could be to combine statistical and dynamical forecasts to produce a hybrid system; recent studies suggest this approach could enhance prediction skill at seasonal timescales (Schepen et al., 2012; Slater and Villarini, 2018). Research shows that seasonal hydrological forecasts are able to inform local decisions and actions, and that while uncertainty is not necessarily a barrier to the use of such forecasts, a range of information, including forecast skill, different forecast types and local knowledge are important, alongside a need for higher resolutions to aid local decision-making (Neumann et al., 2018).

6.5 Conclusions

This paper has evaluated the ability of two different seasonal forecasting approaches, statistical historical probabilities and the dynamical GloFAS-Seasonal, to predict both high and low river flow during El Niño, with the correct timing. Previous research has highlighted the importance of considering the hydrology in addition to meteorological variables, with precipitation often used by decision-makers as a proxy for river flow. These recently-developed forecasts, both of

which are used for El Niño preparedness activities, aim to provide hydrologically relevant predictions of hydrological extremes.

While the results presented indicate that the skill of both forecasts varies by location, season and lead time, and it is important to remember that both approaches have uncertainties associated with them and regions where they lack skill, we are able to draw the following conclusions, to answer the question: **what is the most useful approach for forecasting hydrological extremes during El Niño?**

1. In regions that are strongly influenced by central Pacific El Niños, and in those where GloFAS-Seasonal forecasts currently lack skill, Historical Probabilities generally provide a more useful forecast.

2. In regions where river flow is also influenced by other teleconnections, GloFAS-Seasonal forecasts are typically more useful, as they are better able to account for the characteristics of each El Niño, including the location, timing and magnitude of the SST anomalies, and simulate the response to other modes of climate variability coinciding with El Niño. For example, the phase of the PDO, IOD, NAO, can act to strengthen, mitigate or even reverse the river flow response to El Niño at a regional scale.

3. At lead times of a season ahead, dynamical seasonal forecasts, such as the GloFAS-Seasonal river flow forecasts and seasonal precipitation forecasts, are better able to account for the interaction between various modes of climate variability. Historical Probabilities are, however, available at even earlier lead times, when an El Niño is first forecast or begins to develop.

We further emphasise that while there is often significant interest in the impacts of El Niño due to its global teleconnections, in some regions, it is important to consider that other modes of climate variability can play a key role in addition to ENSO, or may be able to provide added predictability over the use of ENSO as a predictor of hydrological extremes. As more global scale seasonal hydro-meteorological forecasting systems are developed and forecasts are improved, it will be important to revisit the question of which approach is more useful for forecasting hydrological extremes. To forecast high and low river flow on seasonal timescales, and with the correct timing, is a challenging endeavour. That either or both of these forecasts has some ability to predict these events, several weeks to months in advance, provides optimism for the future of seasonal hydro-meteorological forecasting and its use in decision-making across many water-related sectors.

6.6 Supplementary Figures

Figure S1: Maps indicating the most potentially useful forecast for high river flow (>80th percentile of climatology, left) and low river flow (<20th percentile of climatology, right) in JJA (El Niño onset), at each observation location, for lead times (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5+ months ahead. The 5+ months ahead map is used to indicate whether the HistProbs are potentially useful ahead of the lead time at which GloFAS-Seasonal forecasts are available.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

High River Flow

Low River Flow

(a)

(b)

(c)

(d)

(e)

Most Useful
- Neither
- Similar
- GloFAS
- HistProbs
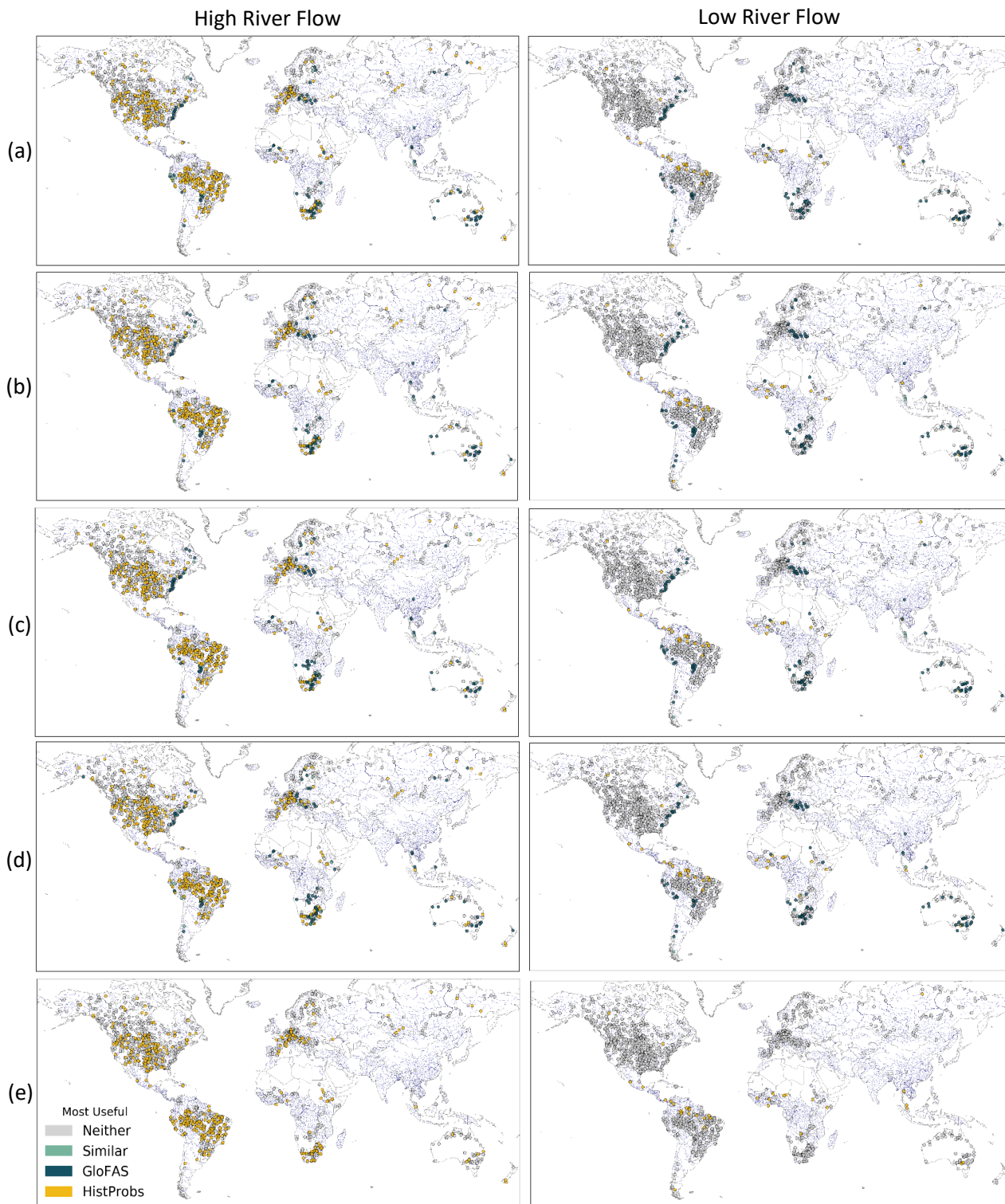
Figure S2: Maps indicating the most potentially useful forecast for high river flow (>80th percentile of climatology, left) and low river flow (<20th percentile of climatology, right) in SON (El Niño onset), at each observation location, for lead times (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5+ months ahead. The 5+ months ahead map is used to indicate whether the HistProbs are potentially useful ahead of the lead time at which GloFAS-Seasonal forecasts are available.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?
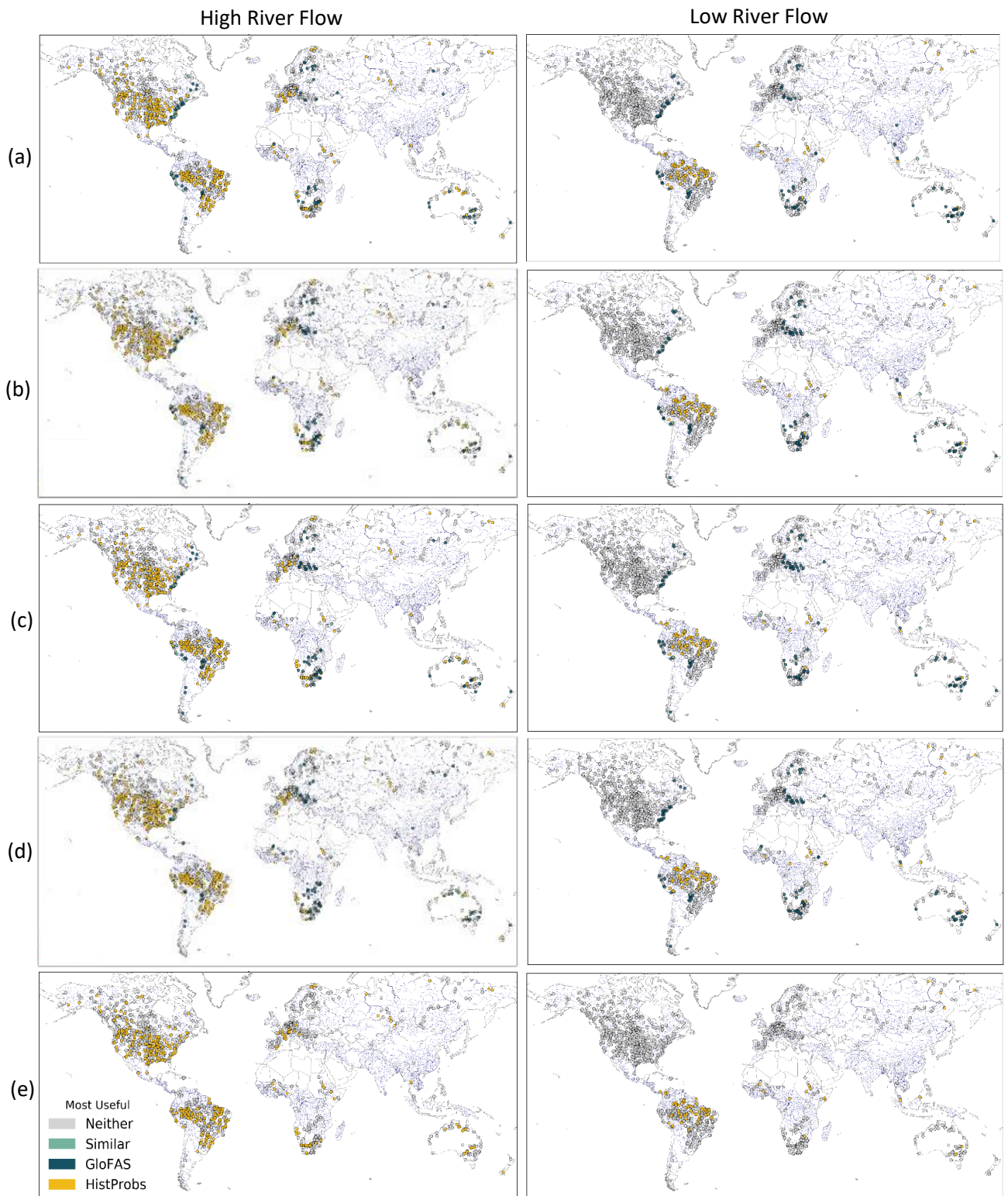
Figure S3: Maps indicating the most potentially useful forecast for high river flow (>80th percentile of climatology, left) and low river flow (<20th percentile of climatology, right) in DJF (El Niño peak), at each observation location, for lead times (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5+ months ahead. The 5+ months ahead map is used to indicate whether the HistProbs are potentially useful ahead of the lead time at which GloFAS-Seasonal forecasts are available.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

High River Flow
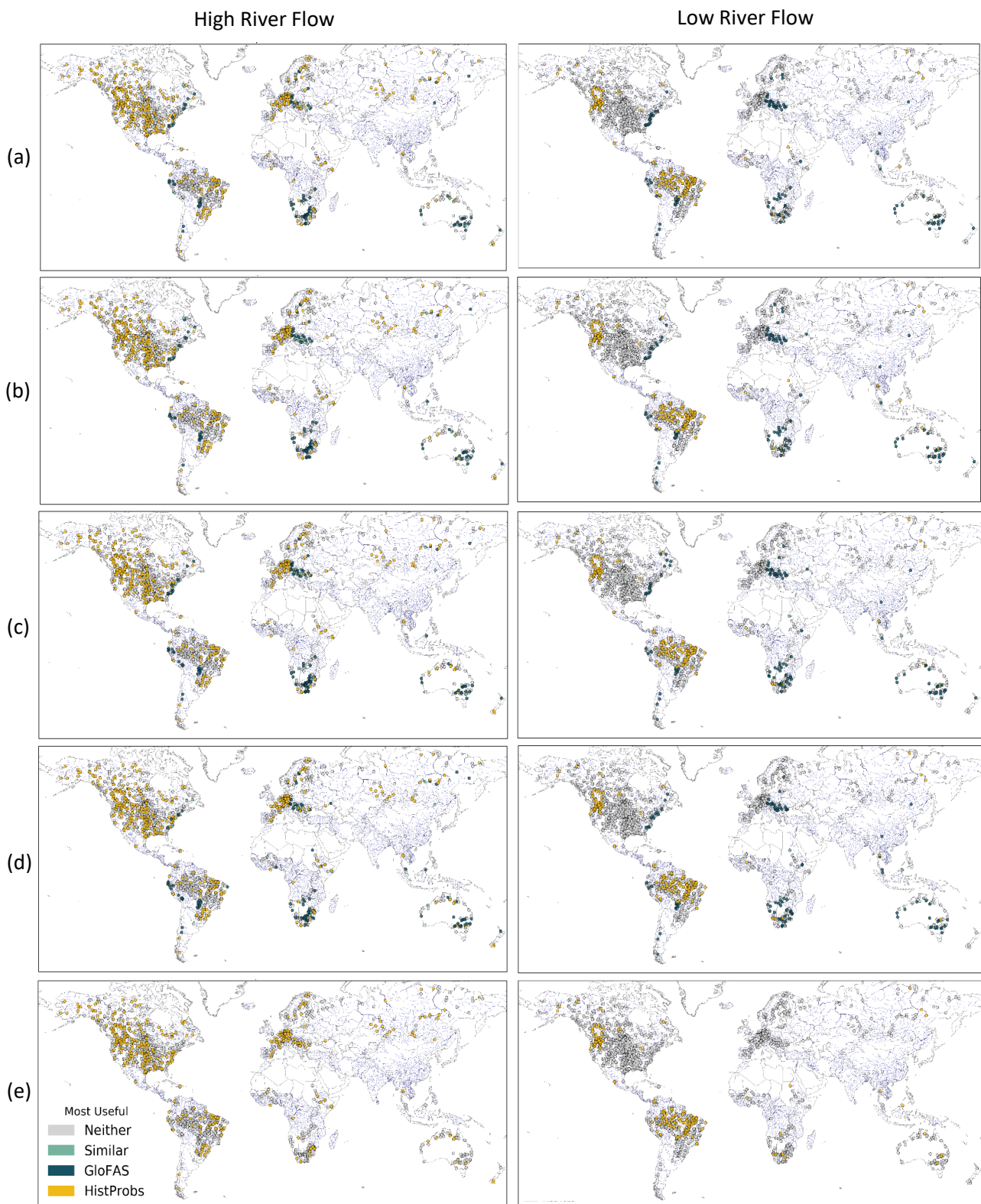
Low River Flow

Figure S4: Maps indicating the most potentially useful forecast for high river flow (>80th percentile of climatology, left) and low river flow (<20th percentile of climatology, right) in MAM (El Niño decay), at each observation location, for lead times (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5+ months ahead. The 5+ months ahead map is used to indicate whether the HistProbs are potentially useful ahead of the lead time at which GloFAS-Seasonal forecasts are available.

Chapter 6. What is the most useful approach for forecasting hydrological extremes during El Niño?

Chapter 7

Conclusions

The aim of this thesis has been to explore ways in which to extend the predictability of flood hazard at the global scale, and provide earlier indications of potential flood events. While several forecasting centres produce operational forecasts of floods in the medium-range, earlier indications, many weeks or even months in advance, could provide crucial information for flood preparedness and disaster risk reduction. For example, preparedness actions such as distributing humanitarian funds, providing training to humanitarian actors and prepositioning aid items, require longer lead times than can be provided by medium-range forecasts. There are two key ways to use the inherent predictability of the atmosphere and land surface to provide early warning information:

- Statistical analysis based on large-scale climate variability and teleconnections

- Seasonal forecasting using coupled ocean-atmosphere general circulation models

Both of these have been widely studied and/or implemented for meteorological variables, but the hydrology has often not been considered or included, particularly for large or global scales. This thesis has made progress towards providing the equivalent early awareness information for river flow, as exists for meteorological variables, through three main objectives:

1. Analyse the link between El Niño Southern Oscillation (ENSO), the most dominant mode of interannual large-scale climate variability, and river flow across the globe, using historical events to answer the question "what is the likelihood of flooding during El Niño?".

2. Develop and test seasonal forecasts of flood hazard for the global river network, by driving the hydrological component of the Global Flood Awareness System (GloFAS) with seasonal meteorological forecasts from the European Centre for Medium-Range Weather Forecasts' (ECMWF) coupled ocean-atmosphere general circulation model.

3. Assess the potential usefulness of both the statistical (1) and dynamical (2) approaches to extending flood predictability and providing early indications of flood hazard at the global scale, for decision-making purposes.

This thesis has been structured around four papers, presented as a detailed review of the current state of large-scale flood forecasting (Chapter 2; Emerton et al., 2016) followed by an additional chapter providing background information on ENSO (Chapter 3), and three papers (Chapters 4-6; Emerton et al., 2017, 2018, 2019) addressing each of the above objectives. The following sections summarise the key conclusions from each results chapter, highlight the scientific advances presented in this thesis, and discuss reflections on this research and the potential for future work.

7.1 Key Conclusions

7.1.1 Objective 1: Historical Probabilities of ENSO-Driven Flood Hazard

The first objective of this thesis was addressed in paper 2 (presented in Chapter 4). ENSO is known to influence river flow and flooding at the global scale, with the literature suggesting the possibility of using this link to provide probabilistic predictions of flooding during ENSO events (El Niño and La Niña). The objective of this paper was to assess the likelihood of flooding during El Niño using hydrological data; this was achieved through estimation of the historical probabilities of high (and low) river flow using a new 110-year (1901-2010) river flow reconstruction, and was further extended to include results for La Niña. Historical probabilities are designed to provide useful information about typical ENSO impacts based on historical evidence. **This paper provides, for the first time, the equivalent historical probabilities for river flow during ENSO events**, to those which existed for meteorological variables such as precipitation and temperature.

In addition to providing global maps of historical probabilities for high and low river flow, this paper further addressed several other key aspects of using ENSO as a predictor of flood hazard. **The importance of addressing the hydrology in addition to precipitation was highlighted** by the differences between the probabilities of high river flow and precipitation, and in the ability to further evaluate areas likely to see a lagged influence of El Niño and La Niña on river flow. This has implications for decision-making, particularly in light of the fact that El Niño preparedness activities have often relied on forecasts of precipitation as a proxy for flood hazard. A key conclusion from this paper is that **the reality of using historical probabilities to evaluate regions of the globe that are more likely to be at risk of flooding during El Niño or La Niña is much more complex than is often perceived or reported**. In the run-up to an El Niño or La Niña, potential impacts are often communicated by circling

large areas of the globe under one banner of wetter or drier. However, probabilities can vary significantly from one month to the next, and the uncertainty surrounding such historical probabilities is evident in the results. The implication here is that **while some regions see high probabilities of increased flood hazard, across much of the globe, the likelihood is much lower and more uncertain than might be useful for decision-making purposes.**

7.1.2 Objective 2: Seasonal Hydro-Meteorological Forecasts using GloFAS

In paper 1 (Chapter 2, section 2.6.2), extended-range hydrological forecasting was introduced as a likely future advance in global scale operational flood forecasting. While seasonal forecasts are already used across a wide range of weather-sensitive sectors, seasonal hydrological forecasting has only begun to emerge during the past decade. There exist challenges such as how to effectively combine global scale meteorological and hydrological models for seasonal applications, the computing resources and costs involved with producing global scale probabilistic seasonal forecasts, and how to effectively communicate seasonal forecasts and transfer the forecasts into warnings and actions.

The second objective of this thesis was to develop and test seasonal forecasts of river flow at the global scale using coupled ocean-atmosphere general circulation models (GCMs), by driving the GloFAS river routing model (Lisflood; Van Der Knijff et al., 2010) with seasonal forecasts from ECMWF (SEAS5; Stockdale et al., 2018). During the course of this PhD, a four-month placement working in the Environmental Forecasts (EFAS/GloFAS) team at ECMWF led to the **operational implementation of this research as the first global scale seasonal hydro-meteorological forecasting system**, as part of GloFAS. Paper 3 (Chapter 5) introduced GloFAS-Seasonal, providing an overview of the new forecast products provided, the hydro-meteorological components of the forecasting system, the computational framework used to run the models and produce the forecasts, and an initial evaluation of the forecast skill and reliability for predicting high and low river flow events.

GloFAS-Seasonal provides forecasts of high or low river flow out to 4 months ahead for the global river network that are openly available through the GloFAS website (*www.globalfloods.eu*). The initial evaluation results are promising, indicating that **in many rivers, forecasts are both potentially useful (i.e. more skilful than a long-term average climatology) out to several months ahead in some cases, and are overall more reliable than a forecast of climatology**. The forecast skill and reliability vary by region and season. The evaluation also indicated a tendency of the forecasts to over-predict, and there are regions

of the globe where the forecasts are less skilful than climatology (i.e. not useful) at all lead times. The potential for use in decision-making across a range of water-related sectors, and the promising results of the initial evaluation, suggest that it is a worthwhile use of time and resources to develop such global scale systems.

7.1.3 Objective 3: Which Approach is More Useful?

Through objectives 1 and 2, two new seasonal forecasts of high and low river flow were produced, one using a simple statistical approach, and the other using a resource-intensive dynamical forecasting approach. Both have the potential to provide early warning information at the global scale, and to be used to prepare for El Niño impacts. The third objective of this thesis was to assess the potential usefulness of these two different approaches to extending the predictability of flood hazard. This objective was addressed in paper 4 (Chapter 6), which assessed and compared the two newly-developed forecasts of hydrological extremes during El Niño, and further extended the research objective to include forecasts of low flow in addition to high flow or flood hazard. The evaluation was designed to assess the ability of the forecasts to predict high and low weekly-averaged river flow, with the correct timing. Seasonal forecasts are typically provided in terms of the likelihood of a given variable being above or below normal in the coming weeks or months. The evaluation of the ability of the forecasts to predict more extreme river flow, with the correct timing, presents a new approach and a more significant challenge, particularly on seasonal timescales.

In this paper, **information is provided on where each forecast is (or is not) skilful, and the advantages and disadvantages of each forecasting approach**. Such information is key in terms of both decision-making and evaluating recent advances in the science of seasonal hydrological forecasting. The results indicate that the skill of both forecasts varies by location, season and lead time. The key conclusions, answering the question "what is the most useful approach for forecasting hydrological extremes during El Niño?", are that **historical probabilities are generally more useful in regions that are strongly influenced by central Pacific El Niños, whereas GloFAS-Seasonal is typically more useful in regions where river flow is also influenced by other teleconnections. While dynamical forecasts such as GloFAS-Seasonal are able to account for the interaction between various modes of climate variability, up to a season ahead, historical probabilities are available at even earlier lead times**, when an El Niño event is first forecast or begins to develop.

7.2 Scientific Advances

Recent studies (e.g. Coughlan De Perez et al., 2017) have called for more investment in hydrologically relevant forecasts of flood hazard, and highlighted that more research is required to explore whether forecasts based on climate variability provide better predictions than dynamical forecasts. This thesis has not only addressed the research questions and objectives outlined in Chapter 1, but has further provided two new, openly-available, hydrologically-relevant forecasts of hydrological extremes. The key contributions of this work are summarised below:

1. Previous work (e.g. Ward et al., 2014a, 2014b, 2016) has evaluated the link between ENSO and flooding. This thesis builds on these studies in order to further contribute to the understanding of these links and to provide, for the first time, probabilistic hydrologically-relevant forecast information based on ENSO teleconnections.

2. The issue of uncertainty in these oft-used historical probabilities of ENSO impacts is explored and addressed in the context of the use of such information for decision-making purposes. The likelihood of increased or decreased flood hazard due to El Niño and La Niña was found to be much more complex and uncertain than is typically perceived and reported, both in the scientific literature and more widely.

3. The importance of considering hydrological variables when forecasting flood hazard, rather than relying on precipitation as a proxy, is highlighted throughout this thesis. In particular, the differences between the historical probabilities of ENSO-driven precipitation and extreme river flow were assessed; the forecasts of river flow are shown to provide additional information on the lagged influence of ENSO on flooding compared to increased precipitation.

4. As part of this research, the first global seasonal hydro-meteorological forecasting system was developed and implemented operationally. GloFAS-Seasonal represents a first attempt at overcoming the challenges of producing and providing openly-available seasonal hydro-meteorological forecast products, which are key for organisations working at the global scale, and for regions where no other forecasting system exists.

5. Evaluation of the GloFAS-Seasonal forecasts has highlighted regions of the globe where there is skill in predicting hydrological extremes out to several months ahead, providing

both optimism for the future of seasonal hydrological forecasting, and scope for further improving the seasonal forecasts implemented as part of this research.

6. The potential usefulness of the two newly-developed forecasts of hydrological extremes was assessed and compared to provide information on their ability to predict observed high and low river flow, with the correct timing. This addresses, for the first time, the question of whether statistical forecasts based on climate variability are able to provide more reliable predictions of hydrological extremes than dynamical forecasts.

7. Information is provided on where each of the two newly-developed forecasts is (or is not) skilful at predicting hydrological extremes during El Niño, and the advantages and disadvantages of both the statistical historical probabilities and dynamical GloFAS-Seasonal forecasts. Such information is key in terms of both decision-making, and evaluating recent advances in the science of seasonal hydrological forecasting.

8. Several new, extended-length, global river flow reanalysis products have been produced as part of this research. The 110-year ERA-20CM-R (producing using ERA-20CM; Emerton et al., 2017; Hersbach et al., 2015) is the longest globally consistent river flow dataset available, and ERA5-R (produced using ERA5; Emerton et al., 2018; Hersbach and Dee, 2016) represents the state-of-the-art in terms of reanalysis products. The ERA-Interim/Land (Balsamo et al., 2015) reanalysis has also been used to produce several river flow reanalysis products, using various hydrological model set-ups. Additionally, a dataset containing 34 years of GloFAS-Seasonal reforecasts has been produced (Emerton et al., 2018, 2019). These data are all openly available for use by the scientific community.

The findings of this thesis have implications for an international, inter-disciplinary community of scientists and decision-makers in three major, current research areas: ENSO, hydro-meteorological extremes, and seasonal forecasting. The research has potential applications in decision-making across a wide range of water-related sectors, from agriculture and water resources management to flood preparedness and disaster risk reduction. Throughout this thesis, there has been a focus on furthering research into the predictability of flood hazard at the global scale, and also on providing the equivalent hydrologically relevant forecasting information that existed for meteorological variables. Additionally, there has been an emphasis on providing the information and analysis required to effectively communicate the forecasts produced as part of this research.

In Chapter 4, significant emphasis was put on understanding the uncertainty associated with producing and using historical probabilities, from the uncertainty inherent in the probabilities themselves due to ENSO diversity, and from the underlying uncertainties in the datasets used to produce the probabilities. Both of the forecasts produced have the potential to be used for El Niño preparedness activities, and as such, Chapter 6 provided information on the potential usefulness of the forecasts at various lead times and in different seasons during an El Niño, based on comparisons with observed river flow. Such information is key in terms of decision-making, particularly in the run-up to predicted El Niño events.

In section 7.1.2, the design and implementation of GloFAS-Seasonal was mentioned. Development of the forecasts involved careful consideration of the new forecast layers that would be provided through the GloFAS website. The design was based on both the existing European Flood Awareness System (EFAS) seasonal outlooks and the GloFAS medium-range flood forecasts, and further incorporated feedback from users of the EFAS seasonal outlook. Beyond the research scope of exploring the possibility of producing such forecasts, it was also necessary to consider how the forecasting system would be implemented operationally at ECMWF; this was achieved through collaboration with Ervin Zsoter and the GloFAS team.

7.3 Reflections and Next Steps

Each of the papers presented in this thesis has contributed scientific advances towards extending the predictability of flood hazard at the global scale. While significant progress has been made, the research has also raised further questions and provided motivation for further research. While each chapter presents some aspects for future work, this section considers some reflections on the completed research and outlines several key examples where this work could be extended and built upon.

Chapter 4 provided historical probabilities of ENSO-driven flood hazard, and one of the key uncertainties highlighted in this work is that of ENSO diversity. The research presented in Chapters 4 and 6 was based on statistical analysis of the links between the Niño3.4 SST index (in the central Pacific) and river flow across the globe; future research should aim to account for ENSO diversity by considering eastern Pacific El Niño events, which can result in considerably different impacts to central Pacific events, and events of different magnitude. The importance of ENSO diversity for global teleconnections was discussed in more detail in Chapter 3, section 3.3. It would also be of interest to extend the analysis presented in Chapter 6 to assess whether similar results are obtained during La Niña events.

Building on this, in Chapter 6, the role of various other teleconnections was discussed in the context of the skill of both the historical probabilities and GloFAS-Seasonal forecasts. There is potential for enhanced predictability of flood hazard based on the interaction of multiple teleconnections, such as ENSO and the Indian Ocean Dipole (IOD), or ENSO and the Pacific Decadal Oscillation (PDO). Thus far, it has not been possible to conduct such research or produce historical probabilities for some combinations of teleconnections, due the insufficient length of data records when considering, for example, only the years in which there was a positive IOD and neutral ENSO conditions. With improved datasets, it may be possible in future to produce statistical analyses conditioned on combinations of teleconnections that result in enhanced river flow predictability.

The issue of limited data is one that has been raised several times throughout this thesis. In order to produce consistent, global scale forecasts, it is necessary to make use of reanalysis products, using models to "fill in the gaps" where observations are not available. Multiple reanalysis products have been produced and used for this research, including the 110-year ERA-20CM-R and the higher-resolution ERA5-R. Both have associated benefits and disadvantages; while ERA-20CM-R allows for a larger sample size of El Niño and La Niña events, it has a much lower resolution and does not assimilate atmospheric observations, therefore cannot reproduce synoptic situations. ERA5-R provides a higher resolution and assimilates atmospheric observations, and is produced in near-real-time, but the shorter length of the dataset results in a smaller sample size of ENSO events. As new reanalysis products are produced and their skill improved, it will be interesting to re-estimate the historical probabilities and provide a comparison of the likelihood of ENSO-driven flood hazard when deriving the probabilities using different datasets.

In addition, since these reanalysis products are used to produce the forecasts, they cannot also be used to evaluate the skill of the forecasts, unless the aim of the evaluation is to assess the forecast's ability to predict events within the model world only. As such, observed data was used for the forecast evaluations undertaken in this thesis. The use of observations allowed for assessment of the ability of the forecasts to predict observed events, but there are large areas of the globe where no observations were available. Indeed, many of the regions with sparse data coverage are located where the impacts of ENSO are most significant. In Chapter 2, both data availability and evaluation of forecasts of extreme events were discussed as two of the grand challenges for the future of global scale flood forecasting. It would be beneficial to collate data

from a larger number of national and international organisations, for use in evaluating the forecasts across as much of the globe as possible, as information on forecast skill is crucial for decision-making. Further to this, there exist a wide range of forecast verification metrics, designed to assess various aspects of the forecasts and their value (Cloke et al., 2017). For this research, the ROC score was chosen due to its relevance for decision-making. Going forward, it will also be important to assess other aspects of the forecasts, such as the timing of events, the ensemble spread and bias in the river flow, and work in collaboration with end users to provide useful and detailed information on forecast skill. Work has already begun in partnership with ECMWF and the University of Reading to conduct a more thorough investigation of the skill of both GloFAS and GloFAS-Seasonal.

GloFAS-Seasonal, as an operational forecasting system, will continue to be developed and improved based on the latest scientific advances and on feedback from GloFAS users and partners. Since the publication of Chapter 6, GloFAS-Seasonal has been upgraded to v2.0, which includes a calibrated version of the river routing model. The calibration is described in a recent paper by Hirpa et al. (2018), and it will be interesting to assess the impact of this calibration on the skill of the GloFAS-Seasonal forecasts. While making improvements to the river routing is a key aspect of improving hydro-meteorological forecasting systems such as GloFAS-Seasonal, research has suggested that the largest contribution to errors in flood forecasting comes from the precipitation forecasts used to drive the hydrological models (Sperna Weiland et al., 2015). Future improvements to GloFAS-Seasonal are likely to include post-processing of the precipitation forecasts, improvements to the land surface scheme (HTESSEL, which is currently primarily used for the meteorological forecasts of ECMWF and has limitations for river flow forecasting) and the way that anthropogenic influences (such as reservoirs, dams and water extraction) are accounted for, incorporation of data assimilation techniques, and added information such as forecast skill displayed through the forecast interface.

Finally, throughout this thesis, flood hazard and hydrological extremes were considered in terms of river flow exceeding, or falling below, a given percentile threshold. A recent study by Coughlan de Perez et al. (2017) considered various metrics of extreme rainfall and their correlation with floodiness (Stephens et al., 2015), in order to assess whether seasonal rainfall forecasts should be used for flood preparedness. Their findings show that the best indicator of floodiness varies widely across the study region of Africa. It would be worth extending the research presented in this thesis to consider different metrics of flood hazard, such as return

period thresholds or indeed floodiness, as different metrics may be more appropriate for various applications.

7.4 Closing Remarks

This thesis presents research that has provided some of the equivalent forecast information for hydrological variables as exists for meteorological variables, and extended the predictability of flood hazard at the global scale. Whilst significant improvements have been made in recent years in the field of seasonal forecasting, both for meteorological and hydrological variables, there are many grand challenges still to face in the future of global scale flood forecasting and predictability. As the forecasting community moves towards fully integrated Earth system models, we are likely to see more, and better, hydrological forecasts at the global scale. Such forecasts have the potential to provide early warning information for a range of applications worldwide, from agriculture and water resources management to flood preparedness and disaster risk reduction. The importance of effectively communicating forecasts is evident in terms of their potential use and associated uncertainties, and working with both the users of such forecasts and the centres producing the forecasts is paramount for improving not only the skill but the usability and value of seasonal forecasts and early warning information.

References

Alfieri, L., P. Salamon, F. Pappenberger, F. Wetterhall and J. Thielen, 2012: Operational early warning systems for water-related hazards in Europe, *Environ. Sci. Policy*, **21**, 35–49, doi:10.1016/J.ENVSCI.2012.01.008

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen and F. Pappenberger, 2013: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci*, **17**, 1161–1175, doi:10.5194/hess-17-1161-2013

Alfieri, L., F. Pappenberger, F. Wetterhall, T. Haiden, D. Richardson and P. Salamon, 2014a: Evaluation of ensemble streamflow predictions in Europe, *J. Hydrol.*, **517**, 913–922, doi:10.1016/J.JHYDROL.2014.06.035

Alfieri, L., F. Pappenberger, and F. Wetterhall, 2014b: The extreme runoff index for flood early warning in Europe, *Nat. Hazards Earth Syst. Sci.*, **14**(6), 1505–1515, doi:10.5194/nhess-14-1505-2014

Anderson, E., 2006: Snow Accumulation and Ablation Model – SNOW-17, Tech. Memo. NWS HYDRO-17. Accessed 18 May 2018, http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf

Arduino, G., P. Reggiani and E. Todini, 2005: Recent advances in flood forecasting and flood risk assessment, *Hydrol. Earth Syst. Sci.*, **9**(4), 280–284, doi:10.5194/hess-9-280-2005

Arnal, L., H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski and F. Pappenberger, 2018: Skilful seasonal forecasts of streamflow over Europe?, *Hydrol. Earth Syst. Sci*, **22**, 2057–2072, doi:10.5194/hess-22-2057-2018

Ashok, K., Z. Guan and T. Yamagata, 2003: Influence of the Indian Ocean Dipole on the Australian winter rainfall, *Geophys. Res. Lett.*, **30**(15), doi:10.1029/2003GL017926

Bahra, A., 2011: Managing work flows with ecFlow, *ECMWF Newsl.*, **129**, 30–32. Accessed 18 April 2018, https://www.ecmwf.int/sites/default/files/elibrary/2011/14594-newsletter-no129-autumn-2011.pdf

Balsamo, G., P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A. K. Betts and K. Scipal, 2008: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, *J. Hyrometeorol.*, **10**, 623-643, doi:10.1175/2008JHM1068.1

Balsamo, G., F. Pappenberger, E. Dutra, P. Viterbo and B. van den Hurk, 2011: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, *Hydrol. Process.*, **25**(7), 1046–1054, doi:10.1002/hyp.7808

Balsamo, G., C. Albergel, A. Beljaars, S. Boussetta, E. Brun, H. Cloke, D. Dee, E. Dutra, J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay, T. Stockdale and F. Vitart, 2015: ERA-Interim/Land: a global land surface reanalysis data set, *Hydrol. Earth Syst. Sci.*, **19**(1), 389–407, doi:10.5194/hess-19-389-2015

Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. Li and D. G. DeWitt, 2012: Skill of Real-Time Seasonal ENSO Model Predictions during 2002–11: Is Our Capability Increasing?, *Bull. Am. Meteorol. Soc.*, **93**(5), 631–651, doi:10.1175/BAMS-D-11-00111.1

Bartholmes, J. and E. Todini, 2005: Coupling meteorological and hydrological models for flood forecasting, *Hydrol. Earth Syst. Sci.*, **9**(4), 333–346, doi:10.5194/hess-9-333-2005

Battisti, D. S. and A. C. Hirst, 1989: Interannual Variability in a Tropical Atmosphere–Ocean Model: Influence of the Basic State, Ocean Geometry and Nonlinearity, *J. Atmos. Sci.*, **46**(12), 1687–1712, doi:10.1175/1520-0469(1989)046<1687:IVIATA>2.0.CO;2

BDHI, 2018: Base de Donnees Historiques sur les Inondations. Accessed 23 April 2018, http://bdhi.fr/appli/web/welcome

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, *Earth Syst. Sci. Data*, **5**(1), 71–99, doi:10.5194/essd-5-71-2013

Behera, S. K., J. Luo, S. Masson, P. Delecluse, S. Gualdi, A. Navarra and T. Yamagata, 2005: Paramount Impact of the Indian Ocean Dipole on the East African Short Rains: A CGCM Study, *J. Clim.*, **18**(21), 4514–4530, doi:10.1175/JCLI3541.1

Bell, V. A., H. N. Davies, A. L. Kay, A. Brookshaw and A. A. Scaife, 2017: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, *Hydrol. Earth Syst. Sci.*, **21**(9), 4681–4691, doi:10.5194/hess-21-4681-2017

Bell, R., K. Hodges, P. L. Vidale, J. Strachan and M. Roberts, 2014: Simulation of the Global ENSO–Tropical Cyclone Teleconnection by a High-Resolution Coupled General Circulation Model, *J. Clim.*, **27**(17), 6404–6422, doi:10.1175/JCLI-D-13-00559.1

Bennett, J. C., Q. J. Wang, M. Li, D. E. Robertson and A. Schepen, 2016: Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, *Water Resour. Res.*, **52**(10), 8238–8259, doi:10.1002/2016WR019193

Bennett, J. C., Q. J. Wang, D. E. Robertson, A. Schepen, M. Li and K. Michael, 2017: Assessment of an ensemble seasonal streamflow forecasting system for Australia, *Hydrol. Earth Syst. Sci.*, **21**, 6007–6030, doi:10.5194/hess-21-6007-2017

BfG, 2017: The GRDC. Accessed 23 April 2018, http://www.bafg.de/GRDC/EN/Home/homepage_node.html

Bierkens, M. F. P., 2015: Global hydrology 2015: State, trends, and directions, *Water Resour. Res.*, **51**(7), 4923–4947, doi:10.1002/2015WR017173

Bierkens, M. F. P., V. A. Bell, P. Burek, N. Chaney, L. E. Condon, C. H. David, A. de Roo, P. Döll, N. Drost, J. S. Famiglietti, M. Flörke, D. J. Gochis, P. Houser, R. Hut, J. Keune, S. Kollet, R. M. Maxwell, J. T. Reager, L. Samaniego, E. Sudicky, E. H. Sutanudjaja, N. van de Giesen, H. Winsemius and E. F. Wood, 2015: Hyper-resolution global hydrological modelling: what is next?, *Hydrol. Process.*, **29**(2), 310–320, doi:10.1002/hyp.10391

Bjerknes, J., 1966: A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature, *Tellus*, **18**(4), 820–829, doi:10.1111/j.2153-3490.1966.tb00303.x

BoM, 2018: Seasonal Streamflow Forecasts: Water Information: Bureau of Meteorology. Accessed 24 April 2018, http://www.bom.gov.au/water/ssf/about.shtml

Bradley, R. S., H. F. Diaz, G. N. Kiladis and J. K. Eischeid, 1987: ENSO signal in continental temperature and precipitation records, *Nature*, **327**(6122), 497–501, doi:10.1038/327497a0

Braman, L. M., M. K. van Aalst, S. J. Mason, P. Suarez, Y. Ait-Chellouche and A. Tall, 2013: Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008, *Disasters*, **37**(1), 144–164, doi:10.1111/j.1467-7717.2012.01297.x

Brown, A., S. Milton, M. Cullen, B. Golding, J. Mitchell and A. Shelly, 2012: Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey, *Bull. Am. Meteorol. Soc.*, **93**(12), 1865–1877, doi:10.1175/BAMS-D-12-00018.1

Brown, J. D., L. Wu, M. He, S. Regonda, H. Lee and D. J. Seo, 2014a: Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 1. Experimental design and forcing verification, *J. Hydrol.*, **519**, 2869–2889, doi:10.1016/J.JHYDROL.2014.05.028

Brown, J. D., M. He, S. Regonda, L. Wu, H. Lee and D. J. Seo, 2014b: Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 2. Streamflow verification, *J. Hydrol.*, **519**, 2847–2868, doi:10.1016/J.JHYDROL.2014.05.030

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, *Mon. Weather Rev.*, **133**(5), 1076–1097, doi:10.1175/MWR2905.1

Buizza, R., M. Milleer and T. N. Palmer, 2007: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Q. J. R. Meteorol. Soc.*, **125**(560), 2887–2908, doi:10.1002/qj.49712556006

Burnash, R., R. Ferral and R. McGuire, 1973: *A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers*, U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources, Sacramento, 204pp.

Cai, W., S. Borlace, M. Lengaigne, P. van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. J. McPhaden, L. Wu, M. H. England, G. Wang, E. Guilyardi and F. F. Jin, 2014: Increasing frequency of extreme El Niño events due to greenhouse warming, *Nat. Clim. Chang.*, **4**(2), 111–116, doi:10.1038/nclimate2100

Cai, W., A. Santoso, G. Wang, S. W. Yeh, S. I. An, K. M. Cobb, M. Collins, E. Guilyardi, F. F. Jin, J. S. Kug, M. Lengaigne, M. J. McPhaden, K. Takahashi, A. Timmermann, G. Vecchi, M. Watanabe and L. Wu, 2015a: ENSO and greenhouse warming, *Nat. Clim. Chang.*, **5**(9), 849–859, doi:10.1038/nclimate2743

Cai, W., G. Wang, A. Santoso, M. J. McPhaden, L. Wu, F. F. Jin, A. Timmermann, M. Collins, G. Vecchi, M. Lengaigne, M H. England, D. Dommenget, K. Takahashi and E. Guilyardi, 2015b: Increased frequency of extreme La Niña events under greenhouse warming, *Nat. Clim. Chang.*, **5**(2), 132–137, doi:10.1038/nclimate2492

Camargo, S. J., K. A. Emanuel and A. H. Sobel, 2007: Use of a Genesis Potential Index to Diagnose ENSO Effects on Tropical Cyclone Genesis, *J. Clim.*, **20**(19), 4819–4834, doi:10.1175/JCLI4282.1

Candogan Yossef, N., R. van Beek, A. Weerts, H. Winsemius and M. F. Bierkens, 2017: Skill of a global forecasting system in seasonal ensemble streamflow prediction, *Hydrol. Earth Syst. Sci*, **21**, 4103–4114, doi:10.5194/hess-21-4103-2017

Capotondi, A., A. T. Wittenberg, M. Newman, E. Di Lorenzo, J. Y. Yu, P. Braconnot, J. Cole, B. Dewitte, B. Giese, E. Guilyardi, F. F. Jin, K. Karnauskas, B. Kirtman, T. Lee, N. Schneider, Y. Xue and S. W. Yeh, 2015: Understanding ENSO Diversity, *Bull. Am. Meteorol. Soc.*, **96**(6), 921–938, doi:10.1175/BAMS-D-13-00117.1

Cemaden, 2017: Cemaden – Centro Nacional de Monitoramento e Alertas de Desastres Naturais. Accessed 23 April 2018, http://www.cemaden.gov.br/

Chan, J. C. L., 1985: Tropical Cyclone Activity in the Northwest Pacific in Relation to the El Niño/Southern Oscillation Phenomenon, *Mon. Weather Rev.*, **113**(4), 599–606, doi:10.1175/1520-0493(1985)113<0599:TCAITN>2.0.CO;2

Chand, S. S., L. McBride, K. J. Tory, M. C. Wheeler and K. J. E. Walsh, 2013: Impact of Different ENSO Regimes on Southwest Pacific Tropical Cyclones, *J. Clim.*, **26**(2), 600–608, doi:10.1175/JCLI-D-12-00114.1

Chen, D., T. Lian, C. Fu, M. A. Cane, Y. Tang, R. Murtugudde, X. Song, Q. Wu and L. Zhou, 2015: Strong influence of westerly wind bursts on El Niño diversity, *Nat. Geosci.*, **8**(5), 339–345, doi:10.1038/ngeo2399

Chiew, F. H. S. and T. A. McMahon, 2002: Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, *Hydrol. Sci. J.*, **47**(3), 505–522, doi:10.1080/02626660209492950

Chow, V. T., D. R. Maidment and L. W. Mays, 1988: *Applied hydrology*, Tata McGraw-Hill Education, 572pp.

Clark, M. P. and L. E. Hay, 2004: Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow, *J. Hydrometeorol.*, **5**(1), 15–32, doi:10.1175/1525-7541(2004)005<0015:UOMNWP>2.0.CO;2

Cloke, H. L. and F. Pappenberger, 2009: Ensemble flood forecasting: A review, *J. Hydrol.*, **375**(3–4), 613–626, doi:10.1016/J.JHYDROL.2009.06.005

Cloke, H., F. Pappenberger, J. Thielen and V. Thiemig, 2013a: Operational European Flood Forecasting. *Environmental Modelling*, J. Wainwright. and M. Mulligan, Eds., John Wiley & Sons Ltd, Chichester, UK, 415–434

Cloke, H. L., F. Pappenberger, S. J. van Andel, J. Schaake, J. Thielen and M.-H. Ramos, 2013b: Hydrological ensemble prediction systems, *Hydrol. Process.*, **27**(1), 1–4, doi:10.1002/hyp.9679

Cloke, H. L., F. Pappenberger, P. J. Smith and F. Wetterhall, 2017: How do I know if I've improved my continental scale flood early warning system?, *Environ. Res. Lett.*, **12**(4), 044006, doi:10.1088/1748-9326/aa625a

Cohen, J., 2016: Weather forecasting: El Niño dons winter disguise as La Niña., *Nature*, **533**(7602), 179, doi:10.1038/533179b

Copernicus, 2018a: Copernicus Climate Data Store. Accessed 23 April 2018, https://climate.copernicus.eu/climate-data-store

Copernicus, 2018b: SWICCA | Service for Water Indicators in Climate Change Adaptation. Accessed 12 January 2018, http://swicca.climate.copernicus.eu/

Coughlan De Perez, E., E. Stephens, K. Bischiniotis, M. Van Aalst, B. Van Den Hurk, S. Mason, H. Nissan and F. Pappenberger, 2017: Should seasonal rainfall forecasts be used for flood preparedness?, *Hydrol. Earth Syst. Sci*, **21**, 4517–4524, doi:10.5194/hess-21-4517-2017

Crochemore, L., M.-H. Ramos and F. Pappenberger, 2016: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, *Hydrol. Earth Syst. Sci*, **20**, 3601–3618, doi:10.5194/hess-20-3601-2016

Cuo, L., T. C. Pagano and Q. J. Wang, 2011: A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting, *J. Hydrometeorol.*, **12**(5), 713–728, doi:10.1175/2011JHM1347.1

Dai, A. and T. M. L. Wigley, 2000: Global patterns of ENSO-induced precipitation, *Geophys. Res. Lett.*, **27**(9), 1283–1286, doi:10.1029/1999GL011140

Davey, M. K., A. Brookshaw and S. Ineson, 2014: The probability of the impact of ENSO on precipitation and near-surface temperature, *Clim. Risk Manag.*, **1**, 5–24, doi:10.1016/J.CRM.2013.12.002

Deltares, 2015: OpenDA. Accessed 19 May 2018, https://www.deltares.nl/en/software/openda/

Deltares, 2018: Global Storm Surge Information System (GLOSSIS). Accessed 19 May 2018, https://www.deltares.nl/en/projects/global-storm-surge-information-system-glossis/

Demargne, J., L. Wu, S. K. Regonda, J. D. Brown, H. Lee, M. He, D.-J. Seo, R. Hartman, H. D. Herr, M. Fresch, J. Schaake and Y. Zhu, 2014: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, *Bull. Am. Meteorol. Soc.*, **95**(1), 79–98, doi:10.1175/BAMS-D-12-00081.1

Demeritt, D., H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes and M.-H. Ramos, 2007: Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, *Environ. Hazards*, **7**(2), 115–127, doi:10.1016/J.ENVHAZ.2007.05.001

DGA, 2018: Ministerio de Obras Públicas - Dirección de General de Aguas. Accessed 23 April 2018, http://www.dga.cl/Paginas/default.aspx

DHM, 2017: Department of Hydrology and Meteorology. Accessed 23 April 2018, http://www.dhm.gov.np/

Di Liberto, T., 2014: The Walker Circulation: ENSO's atmospheric buddy | NOAA Climate.gov ENSO Blog. Accessed 12 November 2018, https://www.climate.gov/news-features/blogs/enso/walker-circulation-ensos-atmospheric-buddy

Donnelly, C., J. C. M. Andersson and B. Arheimer, 2016: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, *Hydrol. Sci. J.*, **61**(2), 255–273, doi:10.1080/02626667.2015.1027710

Duan, W. and C. Wei, 2013: The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model, *Int. J. Climatol.*, **33**(5), 1280–1292, doi:10.1002/joc.3513

DWA, 2018: Department: WAter and Sanitation. Accessed 23 April 2018, http://www.dwa.gov.za/default.aspx

Ebert, E. and J. McBride, 2000 : Verification of precipitation in weather systems: determination of systematic errors, *J. Hydrol.*, **239**(1–4), 179–202, doi:10.1016/S0022-1694(00)00343-7

ECMWF, 2006: TIGGE. Accessed 19 May 2018, https://www.ecmwf.int/en/research/projects/tigge?referer=/

ECMWF, 2011: OpenIFS. Accessed 19 November 2018, https://www.ecmwf.int/en/research/projects/openifs

ECMWF, 2012: ecFlow Documentation. Accessed 18 April 2018, https://software.ecmwf.int/wiki/display/ECFLOW/Documentation

ECMWF, 2015: A global approach to predicting flash floods. Accessed 19 May 2018, https://www.ecmwf.int/en/about/media-centre/news/2015/global-approach-predicting-flash-floods

ECMWF, 2017a: SEAS5 user guide. Accessed 18 April 2018, https://www.ecmwf.int/sites/ default/files/medialibrary/2017-10/System5_guide.pdf

ECMWF, 2017b: What are the changes from ERA-Interim to ERA5? - Copernicus Knowledge Base. Accessed 24 April 2018, https://software.ecmwf.int/wiki/pages/viewpage. action?pageId=74764925

ECMWF, 2018a: Changes in ECMWF model. Accessed 18 May 2018, https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model

ECMWF, 2018b: About OpenIFS. Accessed 26 April 2018, https://software.ecmwf.int/wiki /display/OIFS/About+OpenIFS

ECMWF, 2018c: ECMWF IFS Documentation CY43R1. Accessed 18 April 2018, https://www.ecmwf.int/search/elibrary/IFS?secondary_title=%22IFSDocumentationCY4 3R1%22

ECMWF, 2018d: SEAS5 and the future evolution of the long-range forecast system, ECMWF Sci. Advis. Comm. 47th Sess., ECMWF SAC (Agenda item 5.1), 81

EFAS, 2017: European Flood Awareness System (EFAS). Accessed 23 April 2018, https://www.efas.eu/

Emerton, R. E., E. M. Stephens, F. Pappenberger, T. C. Pagano, A. H. Weerts, A. W. Wood, P. Salamon, J. D. Brown, N. Hjerdt, C. Donnelly, C. A. Baugh and H. L. Cloke, 2016: Continental and global scale flood forecasting systems, *Wiley Interdiscip. Rev. Water*, **3**(3), 391–418, doi:10.1002/wat2.1137

Emerton, R., H. L. Cloke, E. M. Stephens, E. Zsoter, S. J. Woolnough and F. Pappenberger, 2017: Complex picture for likelihood of ENSO-driven flood hazard, *Nat. Commun.*, **8**, 14796, doi:10.1038/ncomms14796

Emerton, R., E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E. M. Stephens, P. Salamon and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, *Geosci. Model Dev.*, **11**(8), 3327–3346, doi:10.5194/gmd-11-3327-2018

Emerton, R., E. M. Stephens and H. L. Cloke, 2019: What is the most useful approach for forecasting hydrological extremes during El Niño?, *Environmental Research Communications*, doi:10.1088/2515-7620/ab114e

Environment Canada, 2014: Water Level and Flow - Environment Canada. Accessed 23 April 2018, https://wateroffice.ec.gc.ca/

Environmental Systems Research Institute, 2018: ArcMap | ArcGIS Desktop. Accessed 26 April 2018, http://desktop.arcgis.com/en/arcmap/

Essou, G. R. C., F. Sabarly, P. Lucas-Picher, F. Brissette and A. Poulin, 2016: Can Precipitation and Temperature from Meteorological Reanalyses Be Used for Hydrological Modeling?, *J. Hydrometeorol.*, **17**(7), 1929–1950, doi:10.1175/JHM-D-15-0138.1

Fan, F. M., D. Schwanenberg, W. Collischonn and A. Weerts, 2015: Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil, *J. Hydrol. Reg. Stud.*, **4**, 196–227, doi:10.1016/J.EJRH.2015.05.012

Fekete, B. M., C. J. Vörösmarty and R. B. Lammers, 2001: Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, *Water Resour. Res.*, **37**(7), 1955–1967, doi:10.1029/2001WR900024

Fennessy, M. J. and J. Shukla, 1999: Impact of Initial Soil Wetness on Seasonal Atmospheric Prediction, *J. Clim.*, **12**(11), 3167–3180, doi:10.1175/1520-0442(1999)012<3167: IOISWO>2.0.CO;2

FloodList, 2018: FloodList. Accessed 18 May 2018, http://floodlist.com/

García-Pintado, J., D. C. Mason, S. L. Dance, H. L. Cloke, J. C. Neal, J. Freer and P. D. Bates, 2015: Satellite-supported flood forecasting in river networks: A real case study, *J. Hydrol.*, **523**, 706–724, doi:10.1016/J.JHYDROL.2015.01.084

Georgakakos, K. P., R. Graham, R. Jubach, T. M. Modrick, E. Shamir, C. Spencer and J. A. Sperfslage, 2013: Global Flash Flood Guidance System, Phase I, San Diego, USA. Accessed 19 May 2018, http://www.hrc-lab.org/projects/projectpdfs/HRC Technical Report No 9.pdf

GloFAS, 2018a: GloFAS Web Map Service Time (WMS-T) User Manual. Accessed 26 April 2018, http://www.globalfloods.eu/static/downloads/GloFAS-WMS-T_usermanual.pdf

GloFAS, 2018b: GloFAS WMS-T. Accessed 26 April 2018, http://globalfloods-ows.ecmwf.int/glofas-ows/?service=WMS&request=GetCapabilities

Goddard, L. and M. Dilley, 2005: El Niño: Catastrophe or Opportunity, *Journal of Climate*, **18**, 651-665, doi:10.1175/JCLI-3277.1

Gourley, J. J., J. M. Erlingis, Y. Hong, E. B. Wells, J. J. Gourley, J. M. Erlingis, Y. Hong and E. B. Wells, 2012: Evaluation of Tools Used for Monitoring and Forecasting Flash Floods in the United States, *Weather Forecast.*, **27**(1), 158–173, doi:10.1175/WAF-D-10-05043.1

Guha-Sapir, G., R. Below and P. Hovois, 2018: The CRED / EM-DAT International Disaster Database, Univ. Cathol. Louvain - Brussels - Belgium. Accessed 18 May 2018, http://www.emdat.be/database

Guimaraes Nobre, G., B. Jongman, J. Aerts and P. J. Ward, 2017: The role of climate variability in extreme floods in Europe, *Environ. Res. Lett.*, **12**(8), 084012, doi:10.1088/1748-9326/aa7c22

Haiden, T., M. Janousek, J. Bidlot, L. Ferranti, T. Hewson, F. Prates, D. Richardson and F. Vitart, 2014: Evaluation of ECMWF forecasts, including 2013-2014 upgrades, *ECMWF Tech. Memo.*, **742**. Accessed 18 May 2018, https://www.ecmwf.int/en/elibrary/12525-evaluation-ecmwf-forecasts-including-2013-2014-upgrades

Halpert, M. S. and C. F. Ropelewski, 1992: Surface Temperature Patterns Associated with the Southern Oscillation, *J. Clim.*, **5**(6), 577–593, doi:10.1175/1520-0442(1992)005<0577: STPAWT>2.0.CO;2

Hannah, D. M., S. Demuth, H. A. van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl and L. M. Tallaksen, 2011: Large-scale river flow archives: importance, current status and future needs, *Hydrol. Process.*, **25**(7), 1191–1200, doi:10.1002/hyp.7794

Hapuarachchi, H. A. P. and Q. J. Wang, 2008: A review of methods and systems available for flash flood forecasting. *Water for a Healthy Country National Research Flagship.* Accessed 19 May 2018, www.csiro.au/partnerships/WIRADA.html

Hapuarachchi, H. A. P., Q. J. Wang and T. C. Pagano, 2011: A review of advances in flash flood forecasting, *Hydrol. Process.*, **25**(18), 2771–2784, doi:10.1002/hyp.8040

Hersbach, H. and D. Dee, 2016: ERA5 reanalysis is in production, *ECMWF Newsl.*, **147**

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford and P. Poli, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble, *Q. J. R. Meteorol. Soc.*, **141**(691), 2350–2375

Hirpa, F. A., P. Salamon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter and S. J. Dadson, 2018: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, *J. Hydrol.*, **231**, in press

Hoell, A., A. E. Gaughan, S. Shukla and T. Magadzire, 2017: The Hydrologic Effects of Synchronous El Niño–Southern Oscillation and Subtropical Indian Ocean Dipole Events over Southern Africa, *J. Hydrometeorol.*, **18**(9), 2407–2424, doi:10.1175/JHM-D-16-0294.1

Hossain, F. and N. Katiyar, 2006: Improving flood forecasting in international river basins, *Eos, Trans. Am. Geophys. Union*, **87**(5), 49, doi:10.1029/2006EO050001

IDEAM, 2014: IDEAM. Accessed 23 April 2018, http://www.ideam.gov.co/

INDRHI, 2017: INDRHI - National Institute of Hydraulic Resources. Accessed 23 April 2018, http://indrhi.gob.do/

INETER, 2016: Ineter | Instituto Nicaragüense de Estudios Territoriales. Accessed 23 April 2018, http://www.ineter.gob.ni/

IRI, 2018: ENSO Impacts. Accessed 18 May 2018, http://iridl.ldeo.columbia.edu/maproom/ENSO/Impacts.html

Jien, J. Y., W. A. Gough and K. Butler, 2015: The Influence of El Niño–Southern Oscillation on Tropical Cyclone Activity in the Eastern North Pacific Basin, *J. Clim.*, **28**(6), 2459–2474, doi:10.1175/JCLI-D-14-00248.1

Jin, F.-F., 1997: An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model, *J. Atmos. Sci.*, **54**(7), 811–829, doi:10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2

Johnson, N. C., 2013: How Many ENSO Flavors Can We Distinguish?, *J. Clim.*, **26**(13), 4816–4827, doi:10.1175/JCLI-D-12-00649.1

Kao, H.-Y. and J.-Y. Yu, 2009: Contrasting Eastern-Pacific and Central-Pacific Types of ENSO, *J. Clim.*, **22**(3), 615–632, doi:10.1175/2008JCLI2309.1

Karnauskas, K. B., 2013: Can we distinguish canonical El Niño from Modoki?, *Geophys. Res. Lett.*, **40**(19), 5246–5251, doi:10.1002/grl.51007

Kauffeldt, A., 2014: Disinformative and Uncertain Data in Global Hydrology Challenges for Modelling and Regionalisation. PhD thesis, Uppsala Universitet, 80pp. https://www.diva-portal.org/smash/get/diva2:766354/FULLTEXT01.pdf

Khan, S. I., H. Yang, H. J. Vergara, J. J. Gourley, G. R. Brakenridge, T. De Groeve, Z. Flamig, F. Policelli, and Bin Yong, 2012: Microwave Satellite Data for Hydrologic Modeling in Ungauged Basins, *IEEE Geosci. Remote Sens. Lett.*, **9**(4), 663–667, doi:10.1109/LGRS.2011.2177807

Kiladis, G. N. and H. F. Diaz, 1989: Global Climatic Anomalies Associated with Extremes in the Southern Oscillation, *J. Clim.*, **2**(9), 1069–1090, doi:10.1175/1520-0442(1989)002<1069:GCAAWE>2.0.CO;2

Kirtman, B. P. and P. S. Schopf, 1998: Decadal Variability in ENSO Predictability and Prediction, *J. Clim.*, **11**(11), 2804–2822, doi:10.1175/1520-0442(1998)011<2804:DVIEPA>2.0.CO;2

Van Der Knijff, J. M., J. Younis and A. P. J. De Roo, 2010: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *Int. J. Geogr. Inf. Sci.*, **24**(2), 189–212, doi:10.1080/13658810802549154

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil and S. Surendran, 1999: Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble., *Science*, **285**(5433), 1548–1550, doi:10.1126/SCIENCE.285.5433.1548

Lau, K.-M. and P. J. Sheu, 1988: Annual cycle, quasi-biennial oscillation, and southern oscillation in global precipitation, *J. Geophys. Res.*, **93**(D9), 10975, doi:10.1029/JD093iD09p10975

Lau, K.-M. and S. Yang, 2015: Walker Circulation. *Encyclopedia of Atmospheric Sciences*, Elsevier Science, 177-181.

Lee, D., P. J. Ward and P. Block, 2018: Identification of symmetric and asymmetric responses in seasonal streamflow globally to ENSO phase, *Environ. Res. Lett.*, **13**(4), 044031, doi:10.1088/1748-9326/aab4ca

Lehner, B. and G. Grill, 2013: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, *Hydrol. Process.*, **27**(15), 2171–2186, doi:10.1002/hyp.9740

Lehner, B., K. Verdin and A. Jarvis, 2008: New Global Hydrography Derived From Spaceborne Elevation Data, *Eos, Trans. Am. Geophys. Union*, **89**(10), 93, doi:10.1029/2008EO100001

Leutbecher, M. and T. N. Palmer, 2008: Ensemble forecasting, *J. Comput. Phys.*, **227**(7), 3515–3539, doi:10.1016/j.jcp.2007.02.014

Li, H., L. Luo, E. F. Wood and J. SchaakeC, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res*, **114**, doi:10.1029/2008JD010969

Lindström, G., C. Pers, J. Rosberg, J. Strömqvist and B. Arheimer, 2010: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, *Hydrol. Res.*, **41**(3–4), 295, doi:10.2166/nh.2010.007

Liu, Y., A. H. Weerts, M. Clark, F. Hendricks, H.-J. Franssen, S. Kumar, H. Moradkhani, D.-J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec and P. Restrepo, 2012: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, **16**(10), 3863–3887, doi:10.5194/hess-16-3863-2012

Liu, Y., Q. Duan, L. Zhao, A. Ye, Y. Tao, C. Miao, X. Mu and J. C. Schaake, 2013: Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin, *Hydrol. Process.*, **27**(1), 57–74, doi:10.1002/hyp.9496

Lloyd, J., E. Guilyardi, H. Weller and J. Slingo, 2009: The role of atmosphere feedbacks during ENSO in the CMIP3 models, *Atmospheric Science Letters,* **10**(3), 170-176, doi:10.1002/asl.227

Lorenz, E. N., 1963: Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, **20**(2), 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion, *Tellus*, **21**(3), 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x

Lorenz, E. N., 1993: *The essence of chaos*, University of Washington Press, 240pp.

MacLeod, D., 2018: Seasonal predictability of onset and cessation of the east African rains, *Weather Clim. Extrem.*, **21**, 27–35, doi:10.1016/j.wace.2018.05.003

Marchant, R., C. Mumbi, S. Behera and T. Yamagata, 2007: The Indian Ocean dipole ? the unsung driver of climatic variability in East Africa, *Afr. J. Ecol.*, **45**(1), 4–16, doi:10.1111/j.1365-2028.2006.00707.x

Mason, S. J. and N. E. Graham, 1999: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, *Weather Forecast.*, **14**(5), 713–725, doi:10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2

Mason, S. J. and L. Goddard, 2001: Probabilistic Precipitation Anomalies Associated with ENSO, *Bull. Am. Meteorol. Soc.*, **82**, 619–638

McPhaden, M. J., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers, *Geophys. Res. Lett.*, **30**(9), 1480, doi:10.1029/2003GL016872

McPhaden, M. J. and J. Picaut, 1990: El Niño-Southern Oscillation Displacements of the Western Equatorial Pacific Warm Pool, *Science,* **250**(4986), 1385-1388, doi: 10.1126/science.250.4986.1385

McPhaden, M. J., S. E. Zebiak and M. H. Glantz, 2006: ENSO as an integrating concept in Earth science., *Science*, **314**(5806), 1740–5, doi:10.1126/science.1132588

Meißner, D., B. Klein and M. Ionita, 2017: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, *Hydrol. Earth Syst. Sci.*, **21**(12), 6401–6423, doi:10.5194/hess-21-6401-2017

Mittermaier, M., N. Roberts and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score, *Meteorol. Appl.*, **20**(2), 176–186, doi:10.1002/met.296

Mo, K. C., D. P. Lettenmaier, K. C. Mo and D. P. Lettenmaier, 2014: Hydrologic Prediction over the Conterminous United States Using the National Multi-Model Ensemble, *J. Hydrometeorol.*, **15**(4), 1457–1472, doi:10.1175/JHM-D-13-0197.1

Muchangi, J., 2016: El Niño rains not yet over - weatherman | The Star, Kenya. Accessed 18 May 2018, https://www.the-star.co.ke/news/2016/01/06/el-nino-rains-not-yet-over-weatherman_c1270441

NASA, 2015: Flood and Landslide Monitoring. Accessed 20 November 2018, https://pmm.nasa.gov/trmm/flood-and-landslide-monitoring

National Research Council, 2006: *Completing the Forecast*, National Academies Press Expert Consensus Report, Washington, D.C., 4pp, https://www.nap.edu/resource/11699/uncertainty_brief_final-lowres.pdf

Neumann, J. L., L. Arnal., R. Emerton, H. Griffith, S. Hyslop, S. Theofanidi and H. L. Cloke, 2018: Can seasonal hydrological forecasts inform local decisions and actions? A decision-making activity, *Geoscience Communication,* **1**, 35-57, doi.org/10.5194/gc-1-35-2018

Nicholls, N., 1979: A Possible Method for Predicting Seasonal Tropical Cyclone Activity in the Australian Region, *Mon. Weather Rev.*, **107**(9), 1221–1224, doi:10.1175/1520-0493(1979)107<1221:APMFPS>2.0.CO;2

NOAA, 2012: Overview of the Hydrologic Ensemble Forecast Service (HEFS). Accessed 18 May 2018, http://www.nws.noaa.gov/os/water/RFC_support/HEFS_doc/HEFS_Overview_0.1.2.pdf

NOAA, 2015: NOAA's National Weather Service Directives. Accessed 18 May 2018, http://www.nws.noaa.gov/directives/010/010.php

NOAA, 2016a: Past ENSO Events. Accessed 18 May 2018, https://www.esrl.noaa.gov/psd/enso/past_events.html

NOAA, 2016b: Understanding El Niño. Accessed 18 May 2018, http://www.noaa.gov/understanding-el-nino

NOAA, 2016c: Global impacts of El Niño and La Niña. Accessed 19 October 2018, https://www.climate.gov/news-features/featured-images/global-impacts-el-niño-and-la-niña

NOAA, 2018: Schematic Diagrams | El Nino, NOAA Pacific Mar. Environ. Lab. Accessed 18 October 2018, https://www.pmel.noaa.gov/elnino/schematic-diagrams

Novak, D. R., C. Bailey, K. F. Brill, P. Burke, W. A. Hogsett, R. Rausch and M. Schichtel, 2014: Precipitation and Temperature Forecast Performance at the Weather Prediction Center, *Weather Forecast.*, **29**(3), 489–504, doi:10.1175/WAF-D-13-00066.1

OGC, 2015: OGC Web Map Service v1.3.0, doi:10.3173/air.21.76

Olson, D. A., N. W. Junker and B. Korty, 1995: Evaluation of 33 Years of Quantitative Precipitation Forecasting at the NMC, *Weather Forecast.*, **10**(3), 498–511, doi:10.1175/1520-0434(1995)010<0498:EOYOQP>2.0.CO;2

Open Source Geospatial Foundation, 2016: MapServer 7.0.1 documentation. Accessed 26 April 2018, http://mapserver.org/uk/index.html

OpenLayers, 2018: OpenLayers. Accessed 18 April 2018, http://openlayers.org/

Pagano, T. C., D. L. Shrestha, Q. J. Wang, D. Robertson and P. Hapuarachchi, 2013: Ensemble dressing for hydrological applications, *Hydrol. Process.*, **27**(1), 106–116, doi:10.1002/hyp.9313

Pagano, T., A. Wood, K. Werner and R. Tama-Sweet, 2014a: Western U.S. Water Supply Forecasting: A Tradition Evolves, *Eos, Trans. Am. Geophys. Union*, **95**(3), 28–29, doi:10.1002/2014EO030007

Pagano, T. C., A. W. Wood, M.-H. Ramos, H. L. Cloke, F. Pappenberger, M. P. Clark, M. Cranston, D. Kavetski, T. Mathevet, S. Sorooshian and J. S. Verkade, 2014b: Challenges of Operational River Forecasting, *J. Hydrometeorol.*, **15**(4), 1692–1707, doi:10.1175/JHM-D-13-0188.1

Pappenberger, F., H. L. Cloke, G. Balsamo, T. Ngo-Duc and T. Oki, 2010: Global runoff routing with the hydrological component of the ECMWF NWP system, *Int. J. Climatol.*, **30**(14), 2155–2174, doi:10.1002/joc.2028

Pappenberger, F., E. Stephens, J. Thielen, P. Salamon, D. Demeritt, S. J. van Andel, F. Wetterhall and L. Alfieri, 2013: Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication, *Hydrol. Process.*, **27**(1), 132–146, doi:10.1002/hyp.9253

Pappenberger, F., H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson and J. Thielen, 2015a: The monetary benefit of early flood warnings in Europe, *Environ. Sci. Policy*, **51**, 278–291, doi:10.1016/J.ENVSCI.2015.04.016

Pappenberger, F., M.-H. Ramos, H. L. Cloke, F. Wetterhall, L. Alfieri, K. Bogner, A. Mueller and P. Salamon, 2015b: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *J. Hydrol.*, **522**, 697–713, doi:10.1016/J.JHYDROL.2015.01.024

Park, T.-W., Y. Deng and M. Cai, 2012: Feedback attribution of the El Niño-Southern Oscilation-related atmospheric and surface temperature anomalies, *J. Geophysical Res.,* **117**, D23101, doi:10.1029/2012JD018468

Picaut, J., F. Masia and Y. Penhoat, 1997: An Advective-Reflective Conceptual Model for the Oscillatory Nature of the ENSO, *Science*, **277**(5326), 663–666, doi:10.1126/science.277.5326.663

Prudhomme, C., J. Hannaford, S. Harrigan, D. Boorman, J. Knight, V. Bell, C. Jackson, C. Svensson, S. Parry, N. Bachiller-Jareno, H. Davies, R. Davis, J. Mackay, A. McKenzie, A. Rudd, K. Smith, J. Bloomfield, R. Ward and A. Jenkins, 2017: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, *Hydrol. Sci. J.*, **62**(16), 2753–2768, doi:10.1080/02626667.2017.1395032

QGIS Development Team, 2017: Quantum GIS Geographical Information System. Accessed 4 December 2017, https://www.qgis.org/

Rakovec, O., A. H. Weerts, P. Hazenberg, P. J. J. F. Torfs and R. Uijlenhoet, 2012: State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci.*, **16**(9), 3435–3449, doi:10.5194/hess-16-3435-2012

Ramos, M.-H., T. Mathevet, J. Thielen and F. Pappenberger 2010: Communicating uncertainty in hydro-meteorological forecasts: mission impossible?, *Meteorol. Appl.*, **17**(2), 223–235, doi:10.1002/met.202

Raynaud, D., J. Thielen, P. Salamon, P. Burek, S. Anquetin and L. Alfieri, 2015: A dynamic runoff co-efficient to improve flash flood early warning in Europe: evaluation on the 2013 central European floods in Germany, *Meteorol. Appl.*, **22**(3), 410–418, doi:10.1002/met.1469

RCCC, 2018: Home - Red Cross Red Crescent Climate Centre. Accessed 23 April 2018, http://www.climatecentre.org/

Red Cross Red Crescent Climate Centre, 2015: Forecast based financing. Accessed 18 May 2018, http://www.climatecentre.org/programmes-engagement/forecast-based-financing

Richardson, D., J. Bidlot, L. Ferranti, A. Ghelli, T. Haiden, T. Hewson, M. Janousek, F. Prates and F. Vitart, 2012: Verification statistics and evaluations of ECMWF forecasts in 2011-2012., *ECMWF Tech. Memo.*, **688**, https://www.ecmwf.int/en/elibrary/11917-verification-statistics-and-evaluations-ecmwf-

Ropelewski, C. F. and M. S. Halpert, 1987: Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation, *Mon. Weather Rev.*, **115**(8), 1606–1626, doi:10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2

Ropelewski, C. F. and M. S. Halpert, 1989: Precipitation Patterns Associated with the High Index Phase of the Southern Oscillation, *J. Clim.*, **2**(3), 268–284, doi:10.1175/1520-0442(1989)002<0268:PPAWTH>2.0.CO;2

Ropelewski, C. F. and M. S. Halpert, 1996: Quantifying Southern Oscillation-Precipitation Relationships, *J. Clim.*, **9**(5), 1043–1059, doi:10.1175/1520-0442(1996)009<1043:QSOPR>2.0.CO;2

Saji, N. H. and T. Yamagata, 2003: Possible impacts of Indian Ocean Dipole mode events on global climate, *Clim. Res.*, **25**, 151–169

Schepen, A., Q. J. Wang and D. E. Robertson, 2012: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall, *J. Geophys. Res. Atmos.*, **117**(D20), doi:10.1029/2012JD018011

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese.,2016. GPCC Full Data Reanalysis Version 7.0: Monthly Land-Surface Precipitation from Rain Gauges built on GTS based and Historic Data. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6000072 (Accessed December 2016)

Shaw, E. M., K. Beven, N. A. Chappell R. and R. Lamb, 2011: *Hydrology in practice*, CRC Press, 546pp.

SHI, 2018: State Hydrological Institute (SHI) | Russian Federal State Budgetary Organization. Accessed 23 April 2018, http://www.hydrology.ru/en

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction, *Q. J. R. Meteorol. Soc.*, **128**(580), 647–677, doi:10.1256/003590002321042135

Slater, L. J. and G. Villarini, 2018: Enhancing the Predictability of Seasonal Streamflow With a Statistical-Dynamical Approach, *Geophys. Res. Lett.*, **45**(13), 6504–6513, doi:10.1029/2018GL077945

SMHI, 2015: E-HYPE. Accessed 18 May 2018, http://hypeweb.smhi.se/

Sperna Weiland, F. C., J. A. Vrugt, R. (L.) P. H. van Beek, A. H. Weerts and M. F. P. Bierkens, 2015: Significant uncertainty in global scale hydrological modeling from precipitation data errors, *J. Hydrol.*, **529**, 1095–1115, doi:10.1016/J.JHYDROL.2015.08.061

Stephens, E. and H. Cloke, 2014: Improving flood forecasts for better flood preparedness in the UK (and beyond), *Geogr. J.*, **180**(4), 310–316, doi:10.1111/geoj.12103

Stephens, E., J. J. Day, F. Pappenberger and H. Cloke, 2015: Precipitation and floodiness, *Geophys. Res. Lett.*, **42**(23), 10,316-10,323, doi:10.1002/2015GL066779

Stoeckenius, T., 1981: Interannual Variations of Tropical Precipitation Patterns, *Mon. Weather Rev.*, **109**(6), 1233–1247, doi:10.1175/1520-0493(1981)109<1233:IVOTPP>2.0.CO;2

Stockdale, T., S. Johnson, L. Ferranti, M. Balmaseda and S. Briceag, 2018: ECMWF's new long-range forecasting system SEAS5, *ECMWF Newsl.*, **154**, 15–20, www.ecmwf.int/en/about/news-centre/media-resources

Suarez, M. J. and P. S. Schopf, 1988: A Delayed Action Oscillator for ENSO, *J. Atmos. Sci.*, **45**(21), 3283–3287, doi:10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2

SWALIM, 2018: FAO SWALIM: Somalia Water and Land Information Management. Accessed 23 April 2018, http://www.faoswalim.org/

Takahashi, K., 2015: One forecaster's view on extreme El Niño in the eastern Pacific, NOAA ENSO Blog. Accessed 18 October 2018, https://www.climate.gov/news-features/blogs/enso/one-forecaster's-view-extreme-el-niño-eastern-pacific

Takahashi, K., A. Montecinos, K. Goubanova and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and Modoki El Niño, *Geophys. Res. Lett.*, **38**, L10704, doi:10.1029/2011GL047364

Tang, Y., H. W. Lean and J. Bornemann, 2013: The benefits of the Met Office variable resolution NWP model for forecasting convection, *Meteorol. Appl.*, **20**(4), 417–426, doi:10.1002/met.1300

Thielen, J., J. Bartholmes, M.-H. Ramos and A. de Roo, 2009.: The European Flood Alert System – Part 1: Concept and development, *Hydrol. Earth Syst. Sci.*, **13**(2), 125–140, doi:10.5194/hess-13-125-2009

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, *J. Geophys. Res. Ocean.*, **103**(C7), 14291–14324, doi:10.1029/97JC01444

Tsonevsky, I., 2015: New EFI parameters for forecasting severe convection, *ECMWF Newsl.*, **144**, 27–32, doi:10.21957/2t3a904u

Turner, A., P. M. Inness and J. M. Slingo, 2005: The role of the basic state in the ENSO-monsoon relationship and implications for predictability, *Q. J. R. Meteorol. Soc.,* **131**, 781-804, doi:10.1256/qj.04.70

UNISDR, 2015: Sendai Framework for Disaster Risk Reduction 2015-2030. Accessed 18 May 2018, https://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf

Wanders, N., D. Karssenberg, A. de Roo, S. M. de Jong and M. F. P. Bierkens, 2014: The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrol. Earth Syst. Sci.*, **18**(6), 2343–2357, doi:10.5194/hess-18-2343-2014

Wang-Chun Lai, A., M. Herzog and H.-F. Graf, 2018: ENSO Forecasts near the Spring Predictability Barrier and Possible Reasons for the Recently Reduced Predictability, *J. Clim.*, **31**(2), 815–838, doi:10.1175/JCLI-D-17-0180.1

Wang, C., 2001: A Unified Oscillator Model for the El Niño–Southern Oscillation, *J. Clim.*, **14**(1), 98–115, doi:10.1175/1520-0442(2001)014<0098:AUOMFT>2.0.CO;2

Wang, C., 2002: Atmospheric Circulation Cells Associated with the El Niño–Southern Oscillation, *J. Clim.*, **15**(4), 399–419, doi:10.1175/1520-0442(2002)015<0399:ACCAWT>2.0.CO;2

Wang, C., 2004: ENSO, Atlantic Climate Variability, and the Walker and Hadley Circulations. *The Hadley Circulation: Present, Past and Future,* Springer, 173-202

Wang, C., R. H. Weisberg and J. I. Virmani, 1999: Western Pacific interannual variability associated with the El Niño-Southern Oscillation, *J. Geophys. Res. Ocean.*, **104**(C3), 5131–5149, doi:10.1029/1998JC900090

Wang, C., C. Deser, J.-Y. Yu, P. DiNezio and A. Clement, 2016: El Niño and Southern Oscillation (ENSO): A Review, *Coral Reefs of the World*, Springer, 85-106

Wang, S., J. Huang, Y. He and Y. Guan, 2015: Combined effects of the Pacific Decadal Oscillation and El Niño-Southern Oscillation on Global Land Dry–Wet Changes, *Sci. Rep.*, **4**(1), 6651, doi:10.1038/srep06651

Ward, P. J., W. Beets, L. M. Bouwer, J. C. J. H. Aerts and H. Renssen, 2010: Sensitivity of river discharge to ENSO, *Geophys. Res. Lett.*, **37**, L12402, doi:10.1029/2010GL043215

Ward, P. J., S. Eisner, M. Flörke, M. D. Dettinger and M. Kummu, 2014a: Annual flood sensitivities to El Niño–Southern Oscillation at the global scale, *Hydrol. Earth Syst. Sci.*, **18**(1), 47–66, doi:10.5194/hess-18-47-2014

Ward, P. J., B. Jongman, M. Kummu, M. D. Dettinger, F. C. Sperna Weiland and H.C. Winsemius, 2014b: Strong influence of El Niño Southern Oscillation on flood risk around the world., *Proc. Natl. Acad. Sci.*, **111**(44), 15659–64, doi:10.1073/pnas.1409822111

Ward, P. J., M. Kummu and U. Lall, 2016: Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis, *J. Hydrol.*, **539**, 358–378, doi:10.1016/J.JHYDROL.2016.05.045

Washington, R. and A. Preston, 2006: Extreme wet years over southern Africa: Role of Indian Ocean sea surface temperatures, *J. Geophys. Res.*, **111**, D15104, doi:10.1029/2005JD006724

Weisberg, R. H. and C. Wang, 1997: A Western Pacific Oscillator Paradigm for the El Niño-Southern Oscillation, *Geophys. Res. Lett.*, **24**(7), 779–782, doi:10.1029/97GL00689

Weisheimer, A. and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts., *J. R. Soc. Interface*, **11**(96), 20131162, doi:10.1098/rsif.2013.1162

Werner, M., J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker and K. Heynert, 2013: The Delft-FEWS flow forecasting system, *Environ. Model. Softw.*, **40**, 65–77, doi:10.1016/J.ENVSOFT.2012.07.010

Williams, I. N. and C. M. Patricola, 2018: Diversity of ENSO Events Unified by Convective Threshold Sea Surface Temperature: A Nonlinear ENSO Index, *Geophys. Res. Lett.*, **45**(17), 9236–9244, doi:10.1029/2018GL079203

WMO, 2005: WMO Executive Council 57th Session Abridged Final Report with Resolutions, Geneva, Switzerland. Accessed 21 May 2018, https://library.wmo.int/pmb_ged/wmo_988_en.pdf

WMO, 2006: Capacity Assessment of National Meteorological and Hydrological Services in Support of Disaster Risk Reduction, Geneva, Switzerland. Accessed 19 May 2018, https://www.wmo.int/pages/prog/drr/documents/CR/_TOC.pdf

WMO, 2015: WWRP S2S Project. Accessed 21 September 2015, http://www.wmo.int/pages/prog/arep/wwrp/new/S2S_project_main_page.html

WMO, 2017: Multi-Hazard Early Warning Systems (MHEWS). Accessed 18 May 2018, http://www.wmo.int/pages/prog/drr/projects/Thematic/MHEWS/MHEWS_en.html

Wood, A. W., E. P. Maurer, A. Kumar and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res.*, **107**(D20), 4429, doi:10.1029/2001JD000659

Wood, A. W., A. Kumar and D. P. Lettenmaier, 2005: A retrospective assessment of National Centers for Environmental Prediction climate model–based ensemble hydrologic forecasting in the western United States, *J. Geophys. Res.*, **110**(D4), D04105, doi:10.1029/2004JD004508

Wood, A. W. and D. P. Lettenmaier, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, **35**(14), L14401, doi:10.1029/2008GL034648

Wood, E. F., J. K. Roundy, T. J. Troy, L. P. H. van Beek, M. F. P. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek, J. Famiglietti, D. Gochis, N. van de Giesen, P. Houser, P. R. Jaffé, S. Kollet, B. Lehner, D. P. Lettenmaier, C. Peters-Lidard, M. Sivapalan, J. Sheffield, A. Wade and P. Whitehead, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, **47**(5), doi:10.1029/2010WR010090

Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li and J. Wang,S 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, *Water Resour. Res.*, **50**(3), 2693–2717, doi:10.1002/2013WR014710

Yamazaki, D., S. Kanae, H. Kim and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model, *Water Resour. Res.*, **47**(4), doi:10.1029/2010WR009726

Yamazaki, D., F. O'Loughlin, M. A. Trigg, Z. F. Miller, T. M. Pavelsky and P. D. Bates, 2014: Development of the Global Width Database for Large Rivers, *Water Resour. Res.*, **50**(4), 3467–3480, doi:10.1002/2013WR014664

Yilmaz, K. K., R. F. Adler, Y. Tian, Y. Hong and H. F. Pierce, 2010: Evaluation of a satellite-based global flood monitoring system, *Int. J. Remote Sens.*, **31**(14), 3763–3782, doi:10.1080/01431161.2010.483489

Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang and K. Guan, 2013: Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, *J. Hydrometeorol.*, **14**(6), 1706–1720, doi:10.1175/JHM-D-13-054.1

Yuan, X., E. F. Wood and Z. Ma, 2015a: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdiscip. Rev. Water*, **2**(5), 523–536, doi:10.1002/wat2.1088

Yuan, X., J. K. Roundy, E. F. Wood and J. Sheffield, 2015b: Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins, *Bull. Am. Meteorol. Soc.*, **96**(11), 1895–1912, doi:10.1175/BAMS-D-14-00003.1

Zhang, C., 2005: Madden-Julian Oscillation, *Reviews of Geophysics,* **43**(2), RG2003, doi:10.1029/2004RG000158

Zhao, T., A. Schepen and Q. J. Wang, 2016: Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, *J. Hydrol.*, **541**, 839–849, doi:10.1016/J.JHYDROL.2016.07.040

References

Appendix

This appendix contains the typeset versions of each of the published chapters presented in this thesis, alongside further publications co-authored during this PhD. Author contribution statements are provided for A1, A2 and A3 in Chapters 2, 4 and 5 respectively, and are provided ahead of each publication in the Appendix for A4, A5 and A6. All author contribution statements have been approved by Professor Hannah Cloke, supervisor.

Hannah L. Cloke

A1: Continental and global scale flood forecasting systems

This paper presents the published version of chapter 2 of this thesis, with the following reference:

Emerton, R. E., E. M. Stephens, F. Pappenberger, T. C. Pagano, A. H. Weerts, A. W. Wood, P. Salamon, J. D. Brown, N. Hjerdt, C. Donnelly, C. A. Baugh and H. L. Cloke, 2016: Continental and Global Scale Flood Forecasting Systems, *WIREs Water*, **3** (3), 391-418, doi:10.1002/wat2.1137[*]

Continental and global scale flood forecasting systems

Rebecca E. Emerton,[1,2,3]* Elisabeth M. Stephens,[1] Florian Pappenberger,[3,4]
Thomas C. Pagano,[5] Albrecht H. Weerts,[6,7] Andy W. Wood,[8]
Peter Salamon,[9] James D. Brown,[10] Niclas Hjerdt,[11] Chantal Donnelly,[11]
Calum A. Baugh[3] and Hannah L. Cloke[1,2]

Floods are the most frequent of natural disasters, affecting millions of people across the globe every year. The anticipation and forecasting of floods at the global scale is crucial to preparing for severe events and providing early awareness where local flood models and warning services may not exist. As numerical weather prediction models continue to improve, operational centers are increasingly using their meteorological output to drive hydrological models, creating hydrometeorological systems capable of forecasting river flow and flood events at much longer lead times than has previously been possible. Furthermore, developments in, for example, modelling capabilities, data, and resources in recent years have made it possible to produce global scale flood forecasting systems. In this paper, the current state of operational large-scale flood forecasting is discussed, including probabilistic forecasting of floods using ensemble prediction systems. Six state-of-the-art operational large-scale flood forecasting systems are reviewed, describing similarities and differences in their approaches to forecasting floods at the global and continental scale. Operational systems currently have the capability to produce coarse-scale discharge forecasts in the medium-range and disseminate forecasts and, in some cases, early warning products in real time across the globe, in support of national forecasting capabilities. With improvements in seasonal weather forecasting, future advances may include more seamless hydrological forecasting at the global scale alongside a move towards multi-model forecasts and grand ensemble techniques, responding to the requirement of developing multi-hazard early warning systems for disaster risk reduction. © 2016 The Authors. *WIREs Water* published by Wiley Periodicals, Inc.

---

*Correspondence to: r.e.emerton@pgr.reading.ac.uk

[1]Department of Geography and Environmental Science, University of Reading, Reading, UK

[2]Department of Meteorology, University of Reading, Reading, UK

[3]European Centre for Medium-Range Weather Forecasts, Reading, UK

[4]School of Geographical Sciences, University of Bristol, Bristol, UK

[5]Bureau of Meteorology, Perth, Australia

[6]Deltares, Delft, The Netherlands

[7]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands

[8]National Center for Atmospheric Research, Boulder, CO, USA

[9]Joint Research Centre, Ispra, Italy

[10]Hydrologic Solutions Ltd., Southampton, UK

[11]Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

---

INTRODUCTION

Flooding has the highest frequency of occurrence of all types of natural disasters across the globe, accounting for 39% of all natural disasters since 2000, with >94 million people affected by floods each year worldwide[1] through displacement from homes, unsafe drinking water, destruction of infrastructure, injury, and loss of life. With an increasing population living in flood-prone areas, the forecasting of floods is key to managing and preparing for imminent disaster.

Investment in building resilience is prioritized in the Sendai Framework for Disaster Risk Reduction (DRR) 2015–2030,[2] with one component of this being the development and use of multi-hazard warning systems.[3] The World Meteorological Organization (WMO) states that economic losses due to severe hydrometeorological events have increased by nearly 50 times over the past 50 years. However, the global loss of life has decreased by a factor of 10[3]. This significant decrease in loss of life is attributed to improved monitoring and forecasting of hydrometeorological events alongside more effective preparation and planning. Four components are suggested by the WMO[3] for effective early warning systems: detection, monitoring, and forecasting hazards; analyses of risks involved; dissemination of timely warnings; and activation of emergency plans to prepare and respond.

The development of forecasting systems producing forecasts and warnings of severe hazards such as floods, droughts, storms, fires, and tropical cyclones on a global scale are critical for disaster risk reduction and further decreases in loss of life. The Sendai Framework for DRR 2015–2030[2] states that at global and regional levels, it is important to 'promote co-operation between academic, scientific and research entities and networks and the private sector to develop new products and services to help reduce disaster risk, in particular those that would assist developing countries and their specific challenges',[2] and forecasting systems such as those discussed here are essential in achieving this, particularly in providing forecasts for countries and regions where no other forecasts and early warnings are available.

The need for large-scale flood forecasting systems can be broken down into three key factors:

(i) to provide information on floodiness[4] across areas larger than a catchment, for example, to indicate where flooding during the rainy season will be worse than normal; information that is of high importance to humanitarian organizations[5];

(ii) to provide forecasts in basins across the globe where there are currently no forecasts available, which is not a massive scale-up of resources; large-scale forecasting is therefore cost-effective compared to focusing on developing and providing hydrometeorological forecasts for single catchments and greatly aids disaster risk reduction and flood early warning efforts globally;

(iii) to support existing capabilities, for example, by using ensemble forecasting techniques to enable probabilistic flood forecasts, or at longer lead times for earlier warnings; probabilistic and extended-range forecasting is computationally expensive, and in addition, many countries do not currently pay for access to these distributed meteorological forecast products and therefore are unable to produce any form of hydrometeorological forecast.

This review outlines the developments that have led to forecasting floods on the global scale, the current state-of-the-art technology in operational large-scale (continental and global) flood forecasting, and future developments in global-scale flood forecasting and early warning.

ADVANCES IN THE SCIENCE AND TECHNIQUES OF GLOBAL FORECASTING

Producing forecasts at the global scale has only become possible in recent years due to the integration of meteorological and hydrological modeling capabilities, improvements in data, satellite observations and land-surface hydrology modeling, and increased resources and computer power.[6–10] While several meteorological and hydrological forecasting centers now run operational flood forecasting models, many of these are for specific locations, river basins, or countries.[8]

Global hydrological modeling is complex due to the geographical variation of rainfall-runoff processes and river regimes,[11] but large-scale flood forecasting systems are now emerging with recent scientific and technological advances and increasing integration of hydrological and meteorological communities, allowing for uncertainty to be cascaded from the meteorological input to the river flow forecasts.[12]

In this section, we analyze the key advances that have enabled the forecasting of floods at the global scale.

The Increasing Skill of Precipitation Forecasts

The skill of precipitation forecasts in global numerical weather prediction (NWP) models has increased significantly in recent years[13–15] (e.g., gaining ~2 days precipitation skill since 2000[16]). With skilful medium-range quantitative precipitation forecasts (QPFs) being produced by NWP models across the globe, it has become possible to produce skilful forecasts of river flow and flooding at large scales for the purpose of early warning.[17] Table 1 outlines the resolutions and forecast ranges of some of the main QPF products used in operational large-scale flood forecasting systems.[8]

Precipitation is challenging to forecast due to the chaotic nature of the atmosphere,[18] where a small change in the initial conditions of the system can result in an unpredictable outcome. The underlying physical processes of precipitation generation are complex to model, and modeling deficiencies can lead to forecast inaccuracies, particularly at longer lead times.[19] In general, due to the lack of observations, precipitation predictions are less skilful in the southern hemisphere, although the difference in the skill of forecasts between the hemispheres has reduced significantly since the introduction of satellite observations and data assimilation.[19,20] Limited data are also an issue in much of the tropics alongside difficulties associated with the simulation of convective precipitation.[21] While QPF skill depends heavily on the region, season, intensity, and storm type,[19] precipitation skill is generally good for rainfall generated by synoptic-scale frontal weather systems.[22] The intensity of precipitation tends to be one of the major problems in QPFs, with convective[21] and orographic enhancement[23] processes tending to result in an under-prediction of intensity alongside the tendency of most global models to over-predict the intensity of light precipitation.[24] Many NWP models struggle with displacement;[19,25] while the areal extent, timing, and intensity of precipitation may be correct, precipitation displacement can be extremely detrimental to forecasts of river flow and flooding.

With ongoing improvements to NWP models[13,14,16,26] (resolution increases, new methods of simulating the physical processes, and increasing computer power), precipitation forecasts have become more useful to hydrological applications.

Ensemble Flood Forecasting— Representing Uncertainty

Over the past 2 decades, NWP has moved from single-solution forecasts of the future state of the atmosphere to probabilistic forecasts using ensemble prediction systems (EPS).[27] Probabilistic forecasts allow the inherent uncertainties in NWP to be represented.[15,28] In hydrological modeling, the four main sources of uncertainty are input data, evaluation data, model structure, and model parameters.[29–32] The relative importance of these uncertainties tends to vary according to catchment characteristics, event magnitude, and lead time of the forecast,[12,27] but it is generally accepted that the greatest uncertainty in flood forecasting beyond 2–3 days lead time stems from the meteorological input.[27,29]

The standard approach in NWP is to produce a single (deterministic) forecast from the initial state, whereas EPS recognise and represent the uncertainty in the initial conditions by perturbing them to produce several initial states.[33,34] The forecast model is run from each of the perturbed initial states, producing many varying, but valid and equally probable, forecast scenarios. In addition to sampling the error in the initial state, many centers also incorporate stochastic physics, which involves applying random perturbations of the parameterized physical processes.[35]

TABLE 1 | Technical details of quantitative precipitation forecasts used in large-scale flood forecasting[8]

Product Type	Spatial Extent	Spatial Resolution	Temporal Resolution	Forecast Range	Uncertainty
Radar nowcasting	~10,000–50,000km²	1–4 km	5–60 min	1–6 h	Low
Ensemble radar nowcasting	~10,000–50,000 km²	1–4 km	5–60 min	1–6 h	
Radar-NWP blending	Regional	~2 km	15–60 min	~6 h	
Limited-area NWP	Regional–Continental	2–25 km	1–6 h	1–3 days	
Ensemble limited-area NWP	Regional–Continental	2–25 km	3–6 h	~5–30 days	
Global NWP	Global	~15–100 km	~3–6 h	~5–30 days	
Seasonal forecasts	Global	~15–100 km	~6–24 h	Months	High

Predictions of river discharge are usually produced by providing the EPS as input to a hydrological model.[27,32,36,37] Prior to this, some pre-processing may be required[32,37]; scale corrections (downscaling or disaggregating) are made as the scale (temporal and spatial) does not usually correspond between the EPS and the hydrological model due to the irregular shape of catchments.[15] Bias or spread corrections may also need to be made.[27]

The use of EPS in flood forecasting allows probabilistic forecasts of flood events at much longer lead times than has previously been possible and is useful in producing forecasts in catchments where no other input data is available.[27] Cloke and Pappenberger[27] give a detailed review of the benefits of ensemble over deterministic flood forecasts, particularly looking at advantages for issuing flood alerts and warnings. Probabilistic forecasts of upcoming events have been shown to provide greater skill than deterministic forecasts[38] and provide key information about the possibility of occurrence of an extreme event.

Operational Large-Scale Flood Forecasting

There exist various large-scale hydrological models run by communities around the globe; Bierkens et al.[39] give a detailed overview of the properties of 14 global scale and 4 continental scale models. Not all of these models are used operationally for the purpose of flood forecasting, and as such, a list of operational continental and global scale flood forecasting models, alongside key system information, is provided in Table 2.

Figure 1 shows a simplified conceptual model for a large-scale flood forecasting system, the components required and the output generated within each component. The operational systems outlined in Table 2 are the focus of this review, and each takes a different approach to the components of the conceptual model. In the following sections, we benchmark the state of current science and technology in undertaking operational continental- and global-scale flood forecasting and early warning.

CONTINENTAL-SCALE FLOOD FORECASTING SYSTEMS

There are currently four operational continental-scale flood forecasting systems, two for Europe: the European Flood Awareness System (EFAS) of the European Commission (EC) and the European HYdrological Predictions for the Environment

(E-HYPE) model of the Swedish Meteorological and Hydrological Institute (SMHI). The Bureau of Meteorology (BoM) runs the Flood Forecasting and Warning Service (FFWS) for Australia, and the U.S. National Weather Service (NWS) run a model covering the continental USA, the Hydrologic Ensemble Forecasting Service (HEFS). This section outlines the components of, and the forecast products produced by, each system.

The European Flood Awareness System

EFAS is an EC initiative developed by the Joint Research Centre (JRC) to increase preparedness for riverine floods across Europe. It was in development from 2002, tested from 2005 to 2010, and has been operational since 2012. After devastating, widespread flooding on the Elbe and Danube rivers in 2002, the EC began development of EFAS, with the aim of providing transnational, harmonized early warnings of flood events and hydrological information to national agencies, complementing local services.[42] Various consortia execute different aspects (e.g., computation and dissemination) of the EFAS operational suite.[43]

Model Components

Rather than using just one meteorological NWP forecast as input, EFAS uses four different forecasts, two ensemble forecasts and two deterministic. Figure 2 details the various components of the EFAS suite, including key information regarding the NWP models. The precipitation, temperature, and evaporation from each of the four forecasts are used as input to the Lisflood hydrological model, which is used as both the rainfall-runoff and the routing components shown in Figure 1 and simulates canopy, surface, and sub-surface processes such as snowmelt (including accounting for accelerated snowmelt during rainfall) and preferential (macropore) flow, soil, and groundwater processes.[42]

Simulated ensemble hydrographs are produced by Lisflood; however, these alone do not constitute a flood forecast. A decision-making element needs to be incorporated.[42] Due to the often limited number of discharge observations in many areas of the globe, these critical thresholds cannot be derived directly from observations. Meteorological data are run through Lisflood to calculate 22-year time series of discharge, to provide a reference threshold for minor or major flooding at each grid cell.

TABLE 2 | Operational large-scale flood forecasting systems

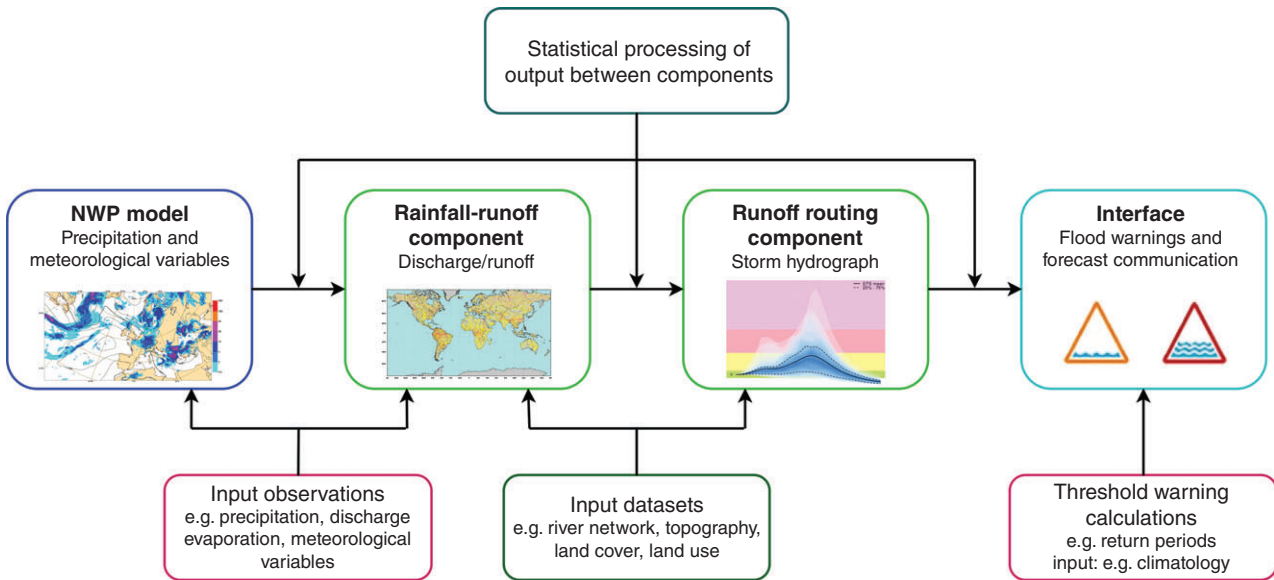| Forecasting System	EFAS (European Flood Awareness System)	E-HYPE (European Hydrological Predictions for the Environment)	FFWS (Flood Forecasting & Warning Service)	HEFS (Hydrologic Ensemble Forecast Service)	GloFAS (Global Flood Awareness System)	GLOFFIS (Global Flood Forecasting Information System)
Domain	Continental (Europe)	Continental (Europe)	Continental (Australia)	Continental (USA)	Global	Global
No. of ensemble members	65	1	≤4	23 Short to medium range, 1 long range	51	73
Forecast range (days)	15	10	10	Sub-hourly to several years	45	15
Spatial resolution	5 km, Regular grid	~15 km, Irregular grid, varies by Basin	~10 km	Varies by Basin	10 km, Regular grid	10 km, 50 km, Regular grid
Forecast frequency	12-h	Daily	6–12-h	Sub-daily to daily	Daily	6-h
NWP input	ECMWF ENS, ECMWF deterministic, DWD Deterministic, COSMO-LEPS	ECMWF deterministic	BoM ACCESS global, regional, city-scale and relocatable deterministic forecasts	RFC deterministic, WPC deterministic, GEFS, CFS, historical observations	ECMWF ENS	ECMWF ENS, GEFS, GFS, historical forcing
Rainfall-runoff model	Lisflood Europe	HYPE	GR4J (daily), GR4H (hourly), URBS	Suite of models (see Figure 8)	HTESSEL (within ECMWF IFS)	PCR-GLOBWB, W3RA
Routing model	Lisflood Europe	HYPE	Muskingum channel routing	Suite of Models (see Figure 8)	Lisflood global	Deltares wflow
River Network	JRC Dataset	HydroSHEDS, HYDRO1K	CatchmentSIM	Suite of Models (see Figure 8)	HydroSHEDS, HYDRO1K	PCR-GLOBWB, SRTM90m, HydroSHEDS
Organization	JRC, ECMWF	SMHI	BoM	National Weather Service	JRC, ECMWF	Deltares
Website	www.efas.eu	e-hypeweb.smhi.se	www.bom.gov.au/water/floods	water.weather.gov/ahps/forecasts.php	www.globalfloods.eu	
Corresponding figure number	Figure 2	Figure 5	Figure 6	Figure 8	Figure 10	Figure 12

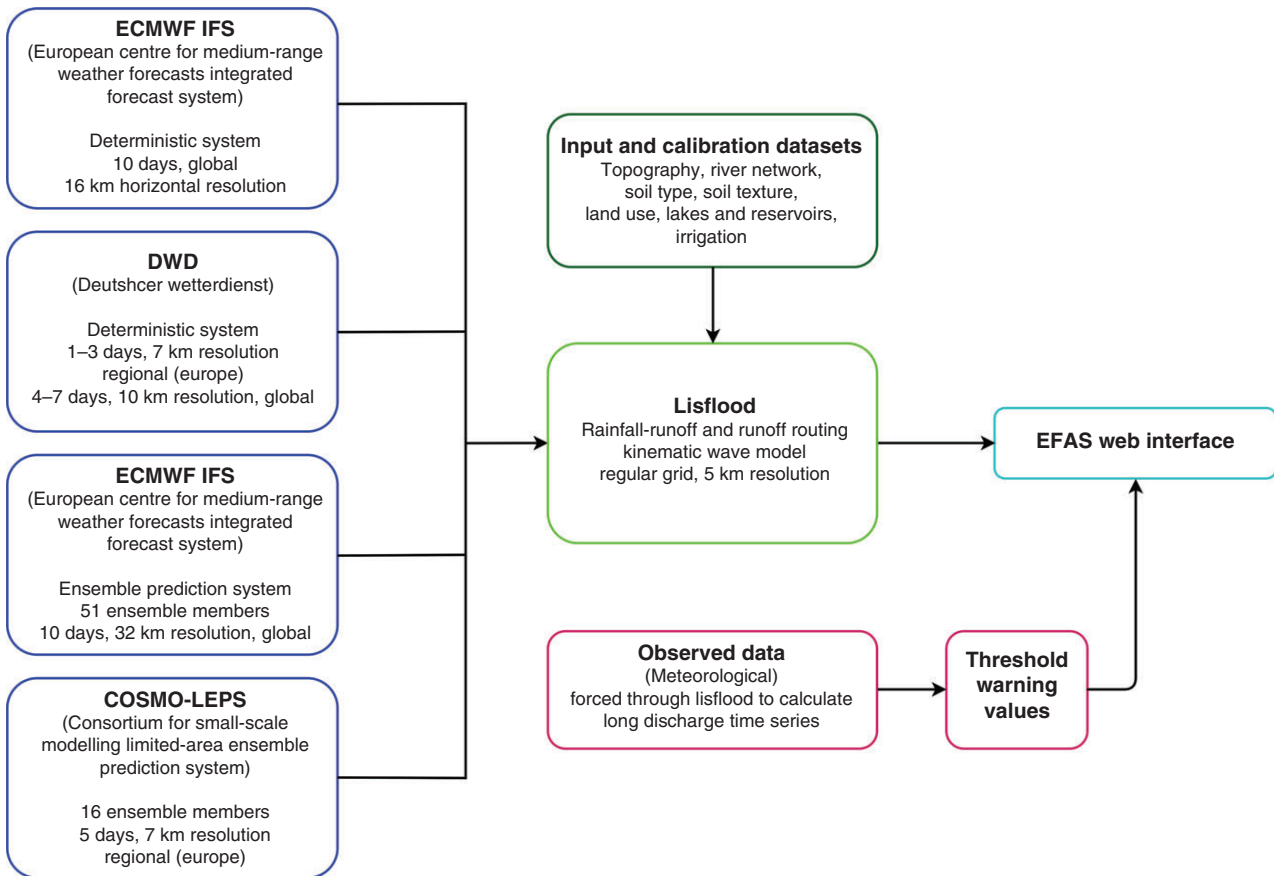**FIGURE 1** | A conceptual large-scale hydrometeorological flood forecasting system.

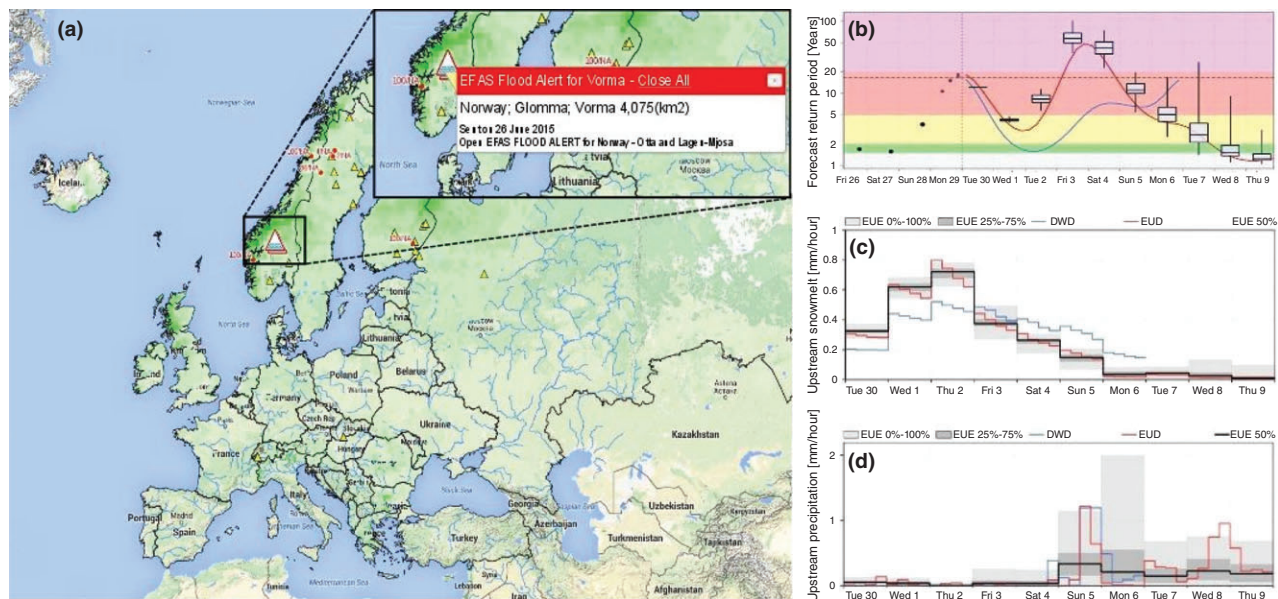**FIGURE 2** | Components of the European Flood Awareness System (EFAS).

FIGURE 3 | The European Flood Awareness System (EFAS) showing (a) the main interface with high (red) and medium (yellow) reporting points, flood alerts (warning triangles), and probability (% likelihood) of exceeding 50mm of precipitation (green shading) during the forecast period (10 days); (inset a) the flood alert displayed when the alert point is clicked on; (b) the return period hydrograph with return period thresholds (1.5 years green, 2 years yellow, 5 years red, 20 years purple); (c) upstream snow melt forecast; (d) upstream precipitation forecast.

Forecast Visualisation

Alongside warnings for each forecast point, the EFAS interface (e.g. Figure 3) provides ensemble hydrographs, which allow the interpretation of the spread of the ensemble and the uncertainty in the forecast. Persistence diagrams showing information about the previous four forecasts also give the user additional information on the forecast uncertainty as NWP models should be able to pick up large-scale synoptic weather systems that typically produce severe events in advance, therefore showing a flood risk consistently in each forecast run.[42] The EFAS interface provides a map of Europe, with all points forecasting a flood event designated by a color responding to the warning threshold; this allows an overview of forecast flood events across the continent. The information and visualization within EFAS are designed to give clear, concise, and unambiguous early warning results.[42]

Warning Dissemination

Copernicus is the European Emergency Management Service, and EFAS is the operational flood early warning system designed to disseminate warnings for Europe under the Copernicus initiative. According to the WMO Executive Council (EC-LVII-Annex VII),[43] National Meteorological and Hydrological Services

(NMHS) constitute the single authoritative voice on weather warnings in their respective countries. Therefore, in order to respect the single voice principle with regard to floods, EFAS real-time information is provided only to hydro-meteorological authorities signing a 'Condition of Access' document. EFAS sends warning emails to these national authorities responsible for flood forecasting, designed to bring awareness of an upcoming flood event, with further details accessed through the interface. There are four types of warning emails provided. *Flood Alerts* are issued when a river basin has a probability of exceeding critical flood thresholds more than 2 days ahead; *Flood Watches* are issued when there is a probability of a river basin exceeding critical thresholds, but the event does not satisfy the conditions for a *Flood Alert* (such as river basin size or warning lead time); and *Flash Flood Watches* are issued when there is a >60% probability of exceeding the flash flood high alert threshold. An example of an EFAS *Flood Alert* is given in Box 1. The 2-day lead time criteria is specified as the forecasting systems used by the national authorities have usually issued a national warning with a lead time of up to 2 days. Additionally, daily overviews are sent to the Emergency Response Coordination Centre (ERCC) of the EC, containing information on ongoing floods in Europe, as reported by the national services and EFAS warnings.

BOX 1

EXAMPLE OF AN EFAS FLOOD ALERT,
SENT TO EFAS PARTNERS AND
NATIONAL AND REGIONAL SERVICES

EFAS FLOOD ALERT REPORT

Dear Partner,

EFAS predicts a high probability of flooding
for Norway—Otta and Lagen-Mjosa tributaries
(Glomma basin) from Monday June
29 onwards.

According to the latest forecasts (2015-06-25
12 UTC), up to 100% EPS (VAREPS) are exceeding
the high threshold (>5 year simulated
return period) and up to 86% EPS (VAREPS)
exceeding the severe threshold (>20-year simulated
return period).

Compared to the VAREPS mean, the ECMWF
deterministic forecast is comparable and the
DWD deterministic forecast is lower.

The earliest flood peak is expected for
Saturday, July 4, 2015.

Please monitor the event on the EFAS-IS
interface (http://www.efas.eu)

Forecast Verification

EFAS also undergoes forecast verification, with two
methods used for this system. First, the hits, false
alarms, and misses are assessed for each flood event,
with events evaluated through feedback reports and
news media. Secondly, skill scores are calculated and
reported regularly through EFAS bulletins, available
via the website (see Table 2).

Operational Applications

EFAS is integrated in the daily forecasting procedures
of many national hydrological services across Europe,
providing operational early warnings and additional
information that is used for decision-making purposes
at national and local scales. Additionally, EFAS is
used by the ERCC to compile reports on the flood situation
and outlook and for the coordination of emergency
response at the continental scale.

The European HYdrological Predictions for the Environment Model

E-HYPE is a multipurpose model based on open data
(Table 3), which is used for various applications such
as water management, research experiments, and
flood forecasting.[45] The E-HYPE Water in Europe

Today (WET) tool (Figure 4) compares the current
hydrological situation with climatological data and
past modeled events. The tool was originally designed
to alert water managers to flow that is predicted to be
outside the normal range (based on the 75th and
25th percentiles) and has evolved to provide information
to many end users. Another setup of the HYPE
model, EFAS-HYPE, uses further restricted datasets
and is currently being tested as an additional model
within EFAS. This section focuses on the river flow
forecasts produced by the WET tool.

Model Components

In contrast to other systems, E-HYPE currently uses
only deterministic NWP input to drive the hydrological
model component, although ensemble forecasting
is intended for future system developments. The
HYPE model[45,46] is a distributed rainfall-runoff
model developed at SMHI, which divides catchments
into sub-basins rather than a regular grid. Each sub-basin
is further divided into classes based on land
use, soil type, and elevation.[44] Alongside processes
such as snow accumulation and melting, evapotranspiration,
and groundwater recharge,[46] HYPE also
takes into account anthropogenic influences including
irrigation and hydropower.[44]

Forecast Visualization

Within the WET tool, forecasts of river flow are compared
to climatology based on the ECMWF ERA-Interim
reanalysis and evaluation datasets (Figure 5)
in order to produce an overview of river flow that is
under or above the normal range. This information is
displayed on a color-coded map of the sub-basins
within the E-HYPE model (Figure 4).

Forecast Verification

Through the E-HYPE and WET interface, various
model performance statistics are available. The model
is verified against observed discharge from river
gauges and allows the user to quickly evaluate the
performance of the model with regard to timing, variability,
and volume error for the point of interest or
across a larger region. The overall model performance
in terms of mean annual discharge is also presented.
Donnelly et al.[45] present a new method for evaluating
the performance of a multi-basin model, and results
from this evaluation of the historical model indicated
that the model is suitable for predictions in ungauged
basins as it captures the spatial variability of flow.
While the model performs well in terms of long-term
means and seasonality, the performance is less effective
in terms of daily variability, particularly in

TABLE 3 | Databases used within the flood forecasting systems. Due to the alternative set-up of the BoM FFWS, this information was not available

| Data Type | Data Source | | | | GLOFFIS | |
	EFAS	E-HYPE[44]	HEFS	GloFAS	PCRGLOB-WB	W3RA
Topography/ routing	SRTM/CCM2	HydroSHEDS & HYDRO1K	NED & NHDPlus	HydroSHEDS & HYDRO1K	HydroSHEDS, HYDRO1K & NASA SRTM	HydroSHEDS, HYDRO1K & NASA SRTM
Land cover	CORINE	CORINE and Globcover 2000	NLCD, MODIS, AVHRR	CORINE and Globcover 2000	GLCC, MIRCA	MODIS
Urban areas	European Soil Data Centre (ESDAC)	Euroland SoilSealing 2009	NA	Harmonized World Soil Database	GLCC	NA
Lake area and spatial distribution	GLWD (Global Lake and Wetland Database)	GLWD (Global Lake and Wetland Database)	NHDPlus	GLWD (Global Lake and Wetland Database)	GLWD, GRaND (Global Reservoir and Dams Database)	NA
Lakes and reservoirs	GLWD, GRaND (Global Reservoir and Dams Database)	GLWD, ERMOBST, FLAKE-Global, International Water Power & Dam, ILEC World Lake Database, LEGOS, SMHI	USGS & Federal state and local water management authorities (e.g. USACE, Reclamation)	GLWD, Global Reservoir and Dams Database GRAND	GLWD, FLAKE-Global, GRaND (Global Reservoir and Dams Database)	NA
Soil Type	European Soil Data Centre (ESDAC)	Based on Land Use and Elevation	SSURGO	Harmonized World Soil Database	FAO DSW	NA
Crop Types	NA	CAPRI, MIRCA-2000	NA	NA	MIRCA	NA
Irrigation	EIM (European Irrigation Map), GMIA (Global Map of Irrigation Areas)	EIM (European Irrigation Map), GMIA (Global Map of Irrigation Areas)	NHDPlus, Local water authorities	GMIA (Global map of Irrigation Areas)	MIRCA	NA

E-HYPE, European HYdrological Predictions for the Environment; GloFAS, Global Flood Awareness System; GLOFFIS, Global Flood Forecasting and Information System; HEFS, Hydrologic Ensemble Forecasting Service; NA, not applicable.
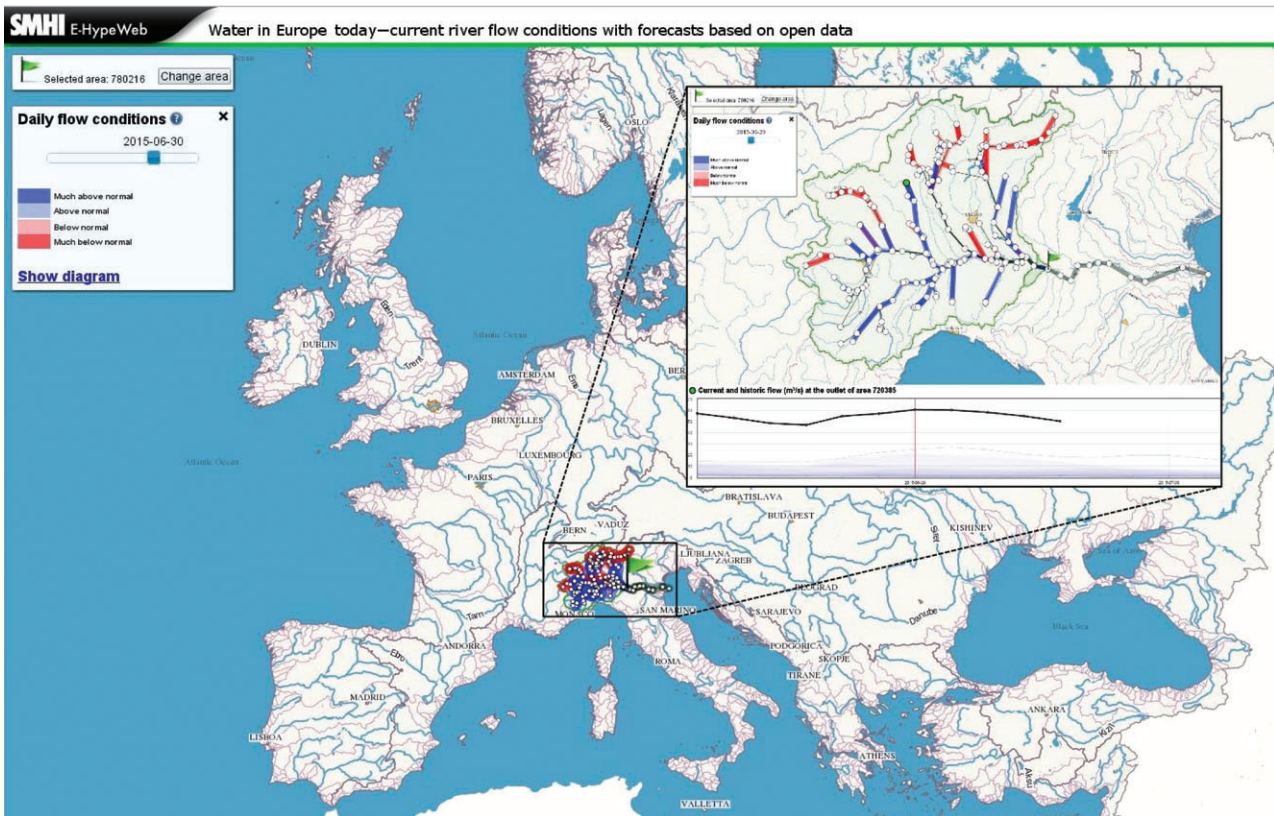
FIGURE 4 | The Water in Europe Today (WET) tool interface with example forecast (inset) showing above-normal (blue shading) and below-normal (red shading) forecast river flow. The hydrograph shows current conditions and forecast river flow (black line) compared to climatology (blue shading). Forecasts are available at hypeweb.smhi.se/europehype/forecasts.

Mediterranean and mountainous areas, and in regions of the most anthropogenic influence.

Operational Applications

E-HYPE is currently being used in several applications across Europe, such as seasonal flow forecasting for the EU European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales (EUPORIAS) project, which aims to help societies deal with climate variability, and providing data for use in oceanography models and as part of the Sharing Water-related Information to Tackle Changes in the Hydrosphere - for Operational Needs (SWITCH-ON) EU project. The WET tool is also used by various other smaller companies around Europe to provide water forecasts, for example, soil-water forecasts for gardening companies.

The Australian Flood Forecasting and Warning Service

The Australian BoM has been producing flood forecasts operationally for several decades, with the technology and systems used to produce these forecasts continually evolving. More recently, the BoM has

introduced short-term (up to 7 days ahead) continuous streamflow forecasting using deterministic NWP models within the Hydrological Forecasting System (HyFS) production environment [based on the Deltares Flood Early Warning System (FEWS) forecasting framework] alongside event-based hydrological modeling and nowcasting using radar rainfall estimates. The BoM services also rely on forecasters for the dissemination and communication of flood warnings and local information regarding river conditions.

Model Components

The NWP forecasts used to force the rainfall-runoff models are produced by the BoM's Australian Community Climate and Earth-System Simulator (ACCESS) NWP model. ACCESS has four components running at different spatial scales and resolutions (Figure 6). In addition to the NWP model output, forecasters and hydrologists at the BoM can produce 'What If' precipitation scenarios, which can force the hydrological models.

Alongside the semi-distributed GR (Génie Rural á 4 Paramétres) hydrological models, event-based
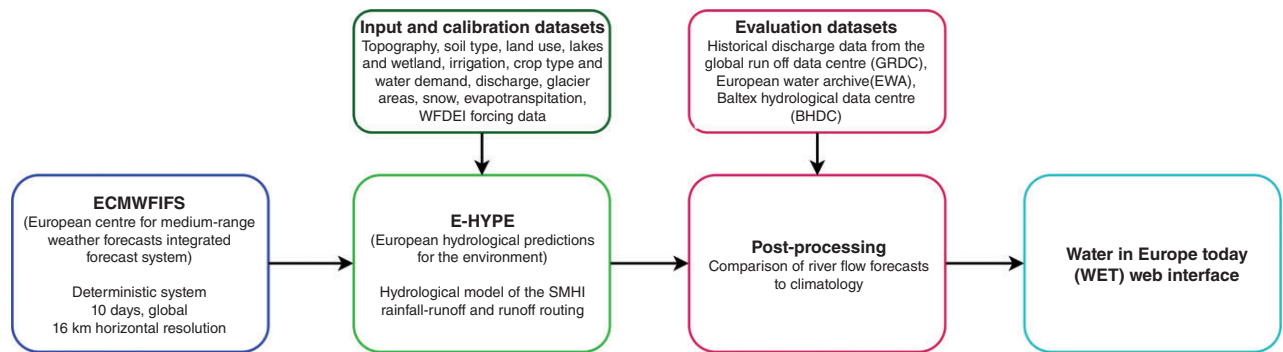
FIGURE 5 | Components of the European Hydrological Predictions for the Environment (E-HYPE) Water in Europe Today (WET) tool.
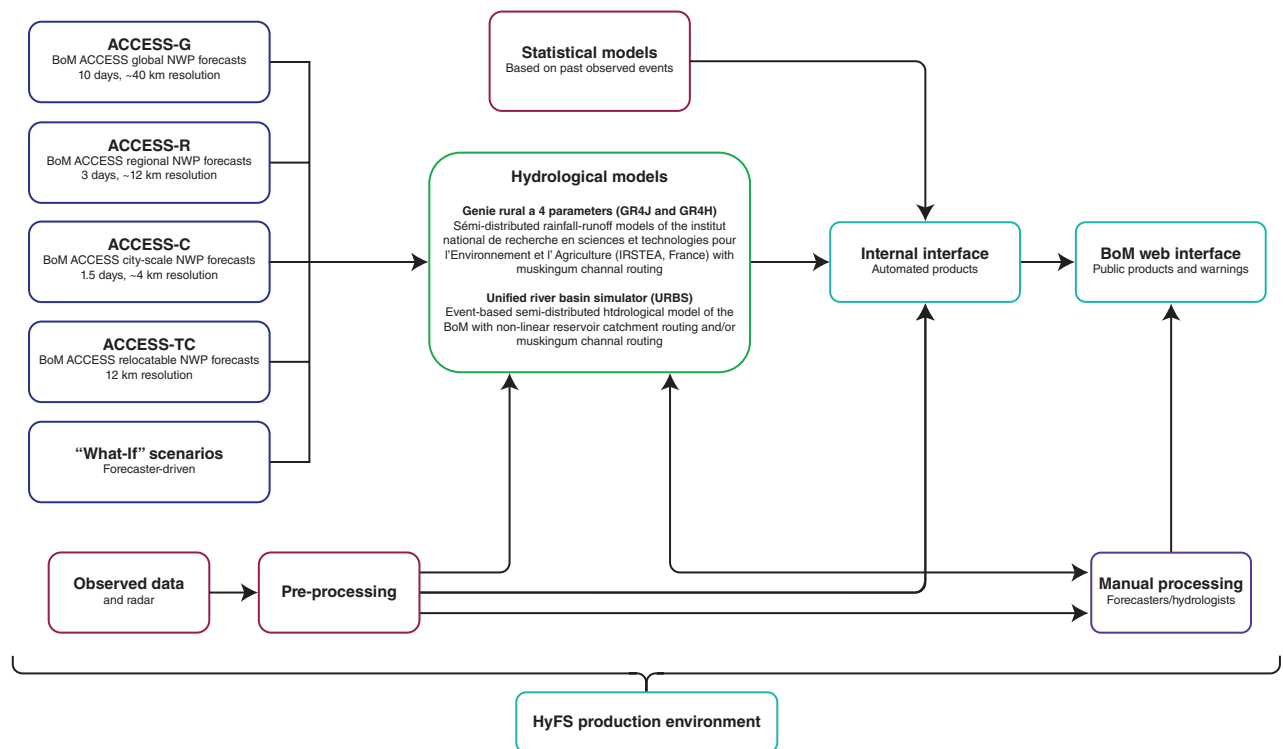
FIGURE 6 | Components of the Australian Flood Forecasting and Warning Service (FFWS).

forecasting is used extensively; for this, local models are used in support of the continental scale system. The resulting river discharge estimations from both model versions are used, alongside observed data and statistical models, to produce automated graphical products such as maps, bulletins, warnings, and alerts.

Role of the Forecaster

Whilst the other systems presented in this paper are almost entirely automated and model-based, the BoM system also relies on the input of expert meteorologists and hydrologists. In addition to producing 'What If' scenarios to feed into the hydrological models, the forecasters are able to manually post-process the forecasts and observed data to produce further products and visualizations and assess the quality of the data and forecasts in real time. The forecasters are also able to produce additional warnings on the fly, for example, if a reservoir is seen to fill or their experience alerts them to an alternative possible scenario to those produced by the hydrological models. The hydrologists at the BoM are also responsible for

dissemination and communication of the forecasts and warnings.

A further reason for the input of forecasters is due to the challenges of producing operational flood forecasts for a large continent with an unevenly distributed population. Metropolitan areas have a dense observation network for both rainfall and river discharge; however, there are large areas of Australia that have no flowing rivers, such as in the Northern Territory where there is an average of one river gauge every 13,360 km^2.

Warning Dissemination

The final products delivered to the end users include flood watches and warnings and information on current river levels and precipitation, which are disseminated to various users at specified stages in the evolution of a flood event through a dedicated web interface, email, fax, and telephone. These are usually text forecasts, an example of which is given in Box 2 for a minor flood event, written by the hydrologists based on the output of the HyFS but can also include automated alerts and bulletins for certain users. Figure 7 shows the corresponding publicly available graphics for this flood event, while the BoM hydrologists also have access to more sophisticated graphical products produced by the automated component of the HyFS, such as ensemble hydrographs.

Forecast Verification

Currently, the BoM uses a manual verification approach, sampling 10% of the warnings issued, based on specifications set out for each forecast point such as a minimum lead time of 6 h or a peak forecast accuracy of ± 0.5 m. With updates to the Flood Forecasting and Warning Service (FFWS), verification software will be introduced, which will automatically compute statistics analyzing the accuracy of the forecast river levels, peak, and timing based on a comparison with observed river levels. The lead time provided for warnings will also be analyzed and compared to the accuracy specifications, providing a measure of performance for a much greater sample of events, which will, in turn, drive further system improvement. Additionally, the HyFS continuous short-term forecasts are verified using a 15-day moving average climatology to calculate the mean absolute error skill score.

Operational Applications

At the BoM, the continuous short-term streamflow forecasts are used across Australia to provide an early indication of an upcoming flood event in order

---

BOX 2

EXAMPLE OF A FLOOD WARNING WRITTEN BY HYDROLOGISTS AT THE BUREAU OF METEOROLOGY

MINOR FLOOD WARNING FOR THE SNOWY RIVER Issued at 9:58 am EST on Wednesday, July 15, 2015

River levels at Orbost are currently around the Minor Flood Level (4.2 m) and rising. A peak of around 4.3–4.4 m is expected during Wednesday afternoon [15/07/2015].

In the interests of community safety, the SES suggests the following precautions:
Don't walk, ride or drive through floodwater, Don't allow children to play in floodwater, Stay away from waterways and stormwater drains, and Keep well clear of fallen power lines

Current Emergency Information is available at http://www.ses.vic.gov.au For emergency assistance, call the SES on telephone number 132 500. For life threatening emergencies, call 000 immediately.

The SES advises that rainfall run-off into waterways in recent fire-affected areas may contain debris such as soil, ash, trees and rocks. People in fire-affected areas should be alert to the potential for landslide and debris on roads.

Weather Forecast:
For the latest weather forecast see www.bom.gov.au/nsw/forecasts/

Next Issue:
The next warning will be issued by 10:00 am Thursday [16/07/2015].

Latest River Heights:
Snowy R. at Basin Creek 4.33 m falling 09:16 AM WED 15/07/15 Buchan R. at Buchan 1.65m falling 08:45 AM WED 15/07/15 Snowy R. at Jarrahmond 4.35 m rising 09:00 AM WED 15/07/15 Snowy R. at Orbost 4.18 m rising 09:00 AM WED 15/07/15

For latest rainfall and river level information see www.bom/gov.au/nsw/flood/

---

to start making arrangements and decisions. These forecasts are then used as a 'heads-up' to start running event-based models at the local scale to provide official, public flood warnings. This is an excellent example of the use of large-scale flood forecasting systems to enhance and supplement existing, local-scale forecasting capabilities.
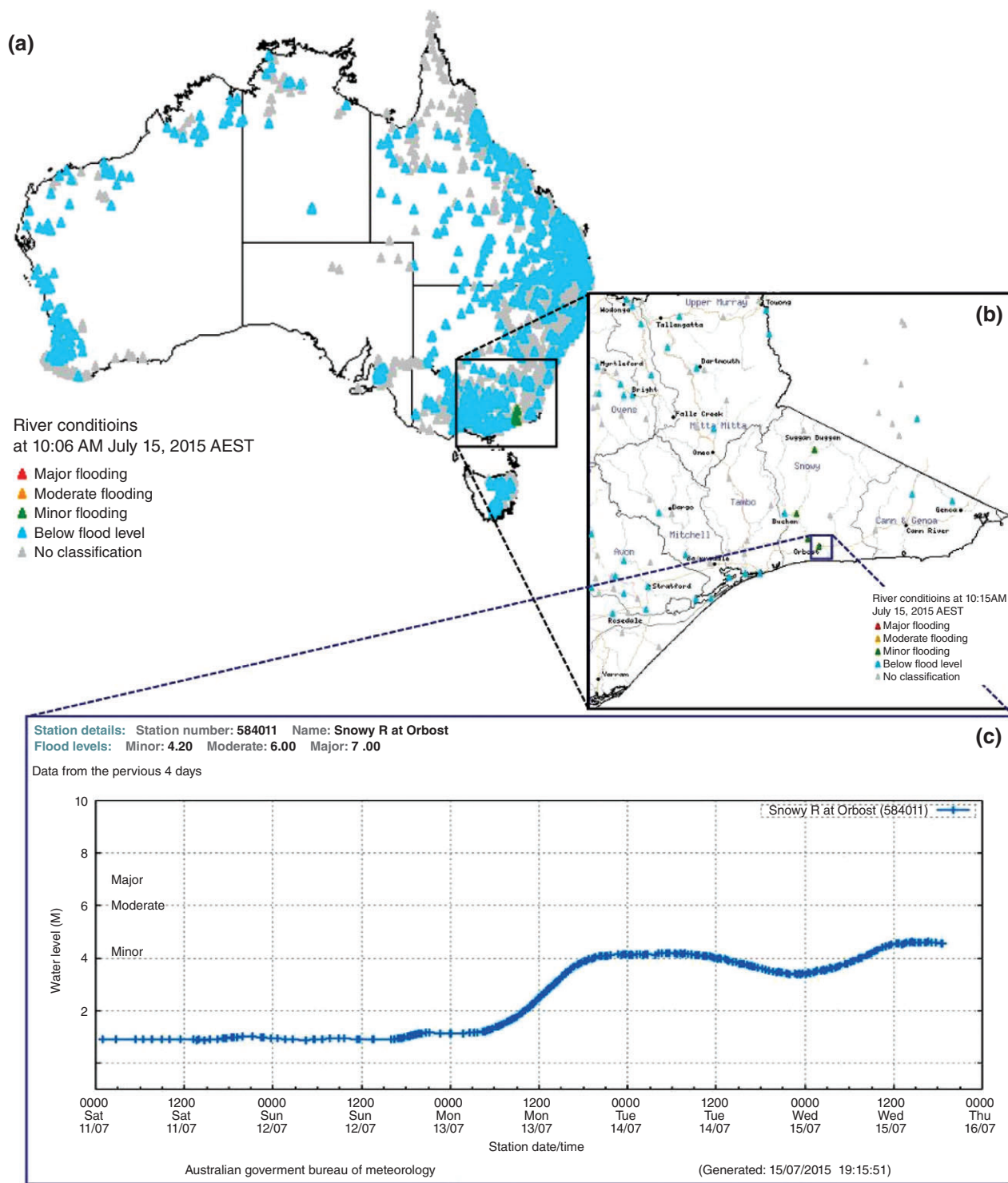

FIGURE 7 | The BoM publicly available flood warnings showing (a) warnings and river conditions across Australia; (b) warnings and river conditions for a particular region; (c) current river levels at a specific warning point where flow is above the minor flood level.

The U.S. Hydrologic Ensemble Forecast Service

The HEFS is run by the NWS and, for river basins across the U.S., provides 'uncertainty- quantified forecast and verification products'.[40] From the late 1990s, NWS service assessments, alongside feedback from end users and the US National Academies,[47] began to confirm the need for probabilistic river forecasts for flood forecasting and water resources. In 2012, the HEFS began to run experimentally at several regional River Forecast Centres (RFCs), each of which forecasts streamflow for hundreds of river locations, and is currently being rolled out operationally at all 13 RFCs. The HEFS aims to produce ensemble streamflow forecasts that seamlessly span lead times from less than 1 h up to several years and that are spatially and temporally consistent, calibrated (i.e., unbiased with an accurate spread), and verified.

Model Components

The HEFS consists of five main components,[40] detailed in Figure 8, and has been implemented to run as part of each RFC's configuration of the Flood Early Warning System (FEWS)-based Community Hydrologic Prediction System (CHPS), which has been the software platform used to run the traditional deterministic flood forecasts and long-range ESP forecasts since 2010. The system is designed to be driven with four meteorological forecast inputs, two of which (GEFS and CFSv2) are the output of NWP models, while the RFC forecasts and climatologies are created by meteorologists for the spatial units of the RFCs' watershed models using predictions from the NCEP Weather Prediction Center (WPC), local NWS Weather Forecast Offices (WFOs), and other sources.[48]

Each RFC may use different combinations of the 19 components within the Hydrological Processor (HP) suite, but the majority of RFC operations center on a lumped implementation of the SAC-SMA[49] and SNOW-17[50] models. The pre-processing step within the HEFS (MEFP, Figure 8) creates an ensemble of seamless hours-to-seasons, calibrated weather and climate forcings, which are fed into the HP. Notably, through use of the MEFP and EnsPost pre- and post-processing components, both the uncertainties in the meteorological input and the hydrology are taken into account.

Forecast Visualization

The graphics generator (Figure 8) uses the resulting ensemble hydrographs to produce visualizations of the forecasts that can be communicated to a range of end users for the purpose of decision making and warning dissemination. These final forecast products include spaghetti plots, exceedance probabilities in the form of bar graphs and probability distribution plots using comparisons with historical simulations (reanalysis datasets), and an expected value chart describing the ensemble distribution. Graphics from the HEFS are currently operational at only a handful of RFCs and are currently being rolled out at the remaining RFCs. An example of an HEFS hydrograph for one river location, alongside the public web interface, is shown in Figure 9. The forecast data associated with the graphical products are also typically available from the RFCs, and many users can access the data directly to drive local decision support models.

Warning Dissemination

NWS product requirements are codified through NWS Directives,[41] and the RFCs generally issue
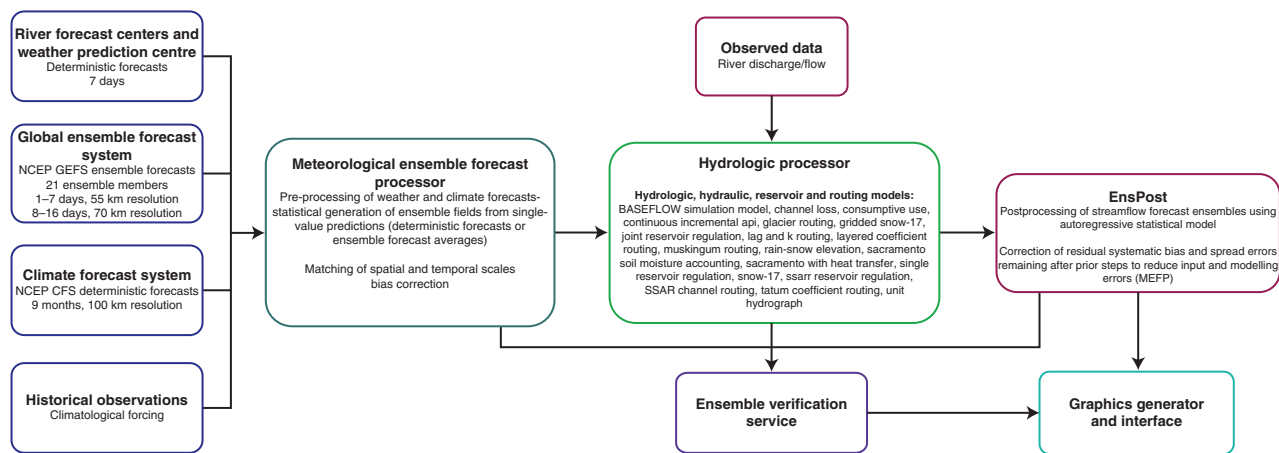

FIGURE 8 | Components of the U.S. Hydrologic Ensemble Forecast System (HEFS).[40,41]
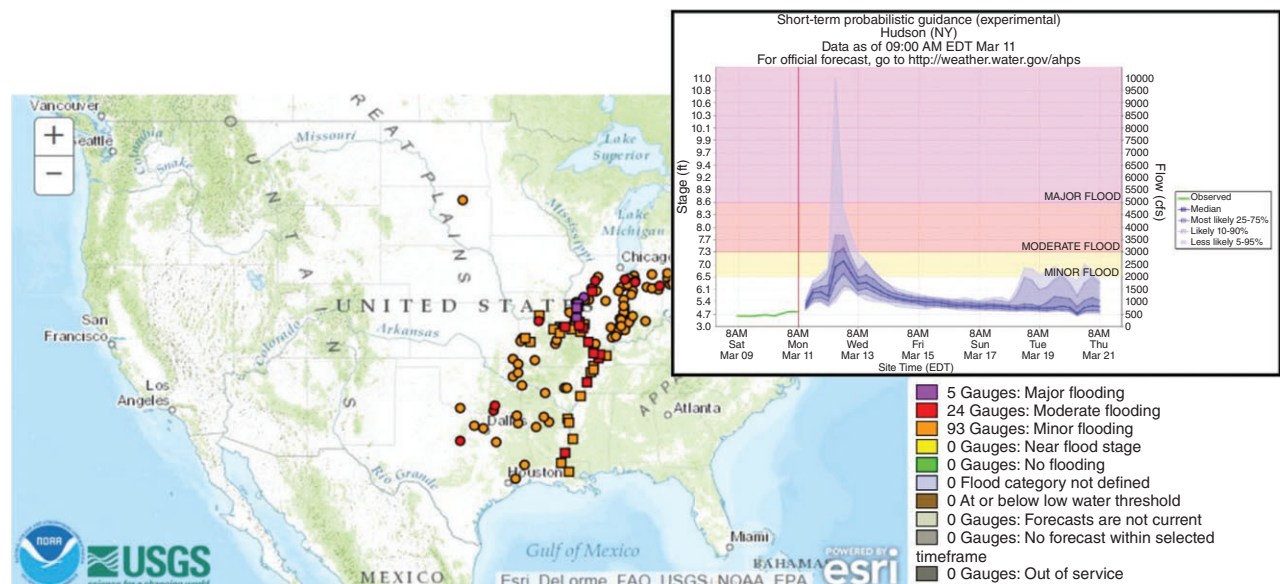
FIGURE 9 | The U.S. Hydrologic Ensemble Forecast System (HEFS) overview map of locations forecasting floods, with color representing flood severity. An ensemble hydrograph is shown for a flood event at one river location, including observed stage and flow (green), forecast stage and flow (purple) in terms of probabilities, and colors indicating the forecast severity based on flood stage data (minor flood, yellow; moderate flood, red; major flood, pink). Forecasts are available at water.weather.gov/ahps/forecasts.php.

products based on hydrometeorological analyzes and long-range predictions that are not time critical and inform non-hazard-related user activities and decisions, such as the *Streamflow Guidance*. The NWS Weather Forecast Offices (WFOs), in contrast, issue the primary hazard-centered alerts related to flooding, including products such as a *Hydrologic Outlook* ('hydrometeorological conditions that could cause flooding or impact water supply'), *Flood Watch* (flooding is likely), or *Flood Warning* (flooding is imminent or occurring). The WFO hydrological products are based primarily on RFC analyzes and predictions; for instance, an RFC forecast exceeding a flood threshold triggers a recommendation to the WFO to release a flood warning that is reviewed by the WFO forecaster. Protocols for linking the newer HEFS ensemble forecasts to alerts are still in development.

Forecast Verification

An additional component of the HEFS shown in Figure 8 is the Ensemble Verification System (EVS), which produces statistics such as the bias in the forecast probabilities, the skill relative to a 'baseline' forecasting system, and the ability to discriminate between events.[46] EVS runs within HEFS and is also freely available as a stand-alone application. The verification statistics are provided as graphical and

textual products. They are used to guide research and development of the HEFS and to improve the configuration of the HEFS for operational forecasting. Studies by Brown et al.[51,52] found that the skill of the precipitation forecasts used for the HEFS are the greatest at lead times of up to 1 week for moderate precipitation and in the wet season (December to March), with limitations in the summer season due to difficulties in forecasting convection. The studies also showed that the skill of the streamflow forecasts, for both the HEFS and traditional RFC deterministic forecasts, is substantially increased through the use of the EnsPost component.

Operational Applications

The HEFS is currently being implemented by all 13 NWS RFCs, with existing or proposed applications ranging from flood forecasting to river navigation, reservoir operation, and long-term planning and management of water resources. For example, reforecasts and operational forecasts from the HEFS are being used by the New York City Department of Environmental Protection (NYCDEP) to improve the management of water supply to NYC by optimizing the quantity and quality of water stored in the NYC reservoirs while avoiding unnecessary infrastructure costs.

GLOBAL-SCALE FLOOD FORECASTING SYSTEMS

At present, there are just two flood forecasting systems that are operational at the global scale, the Global Flood Awareness System (GloFAS) of the ECMWF and EC and the Global Flood Forecasting and Information System (GLOFFIS) run by Deltares. There also exists a Global Flood Monitoring System[53,54] (GFMS) developed by the National Aeronautics and Space Administration (NASA) and the University of Maryland, which uses satellite precipitation as input to a hydrological model to produce real-time global maps of flood events. Global flood monitoring is an important aspect of disaster risk reduction and has many potential applications across the globe; however, the GFMS is not an operational hydrometeorological flood forecasting system and, as such, is not discussed in detail in this review. The reader is referred to the GFMS website[55] and publications[53,54] for further information on the GFMS. This section discusses the components of GloFAS and GLOFFIS along with the products and warnings provided to end users and verification techniques used to assess the performance of these systems.

The Global Flood Awareness System

GloFAS has been producing probabilistic flood forecasts with up to 2 weeks lead time in a pre-operational environment since 2011[9]; this environment enables continuous research, development, and testing in order to produce an operational tool that is independent of administrative and political boundaries. GloFAS can provide downstream countries with early warnings and information on upstream river conditions alongside global overviews of upcoming flood events in large river basins for decision makers ranging from water authorities and hydropower companies to civil protection and international humanitarian aid organizations.

Model Components

In contrast to the other systems presented in this paper, GloFAS uses surface and sub-surface runoff forecasts produced by the NWP model rather than a separate rainfall-runoff component (Figure 1). The Hydrology Tiled ECMWF Scheme for Surface Exchange over Land (HTESSEL) is contained within the IFS and is used as forcing for the Lisflood river routing model. Figure 10 details the components of GloFAS. Although Lisflood global[55] is also a rainfall-runoff model, it is used here to simulate the routing processes and the groundwater processes after re-sampling the runoff forecasts from the IFS to the 0.1° resolution of Lisflood. Additionally, GloFAS contains a loss function to account for water loss within the channel reaches in arid areas, which also simulates the river–aquifer and river–floodplain interaction and the influence of evaporation from large rivers.

Runoff from the ECMWF ERA-Interim reanalysis archive has also been run through Lisflood off-line, producing a deterministic climatology of river flow that is used to compute return periods for the global river network.

Forecast Visualization

Forecasts and warnings produced by GloFAS are provided through a password-protected interface (Figure 11) where users can register to see a global overview of warning points, forecast precipitation accumulations, ensemble hydrographs including
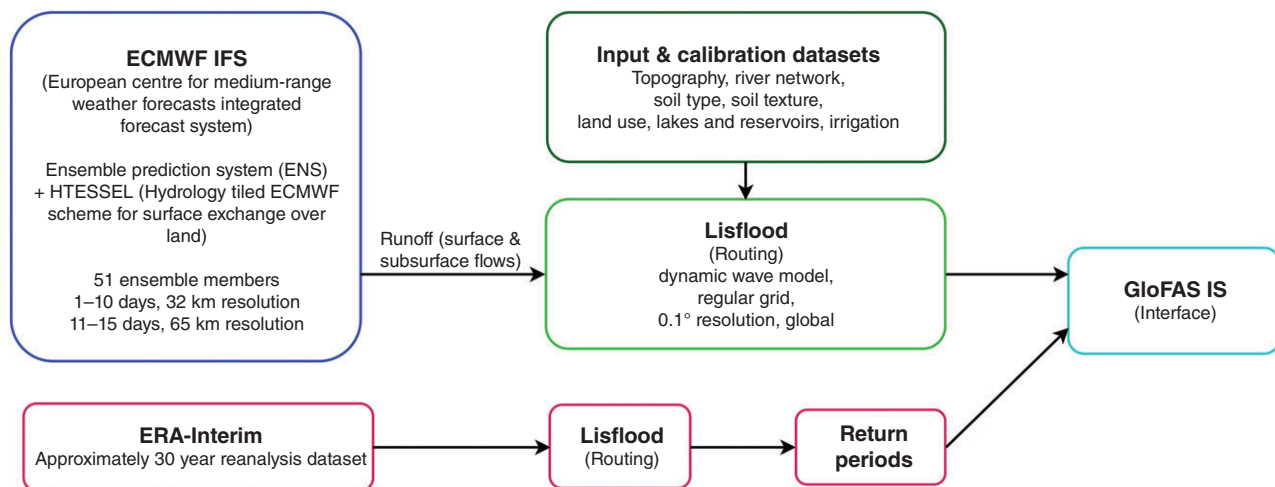
FIGURE 10 | Components of the Global Flood Awareness System (GloFAS).
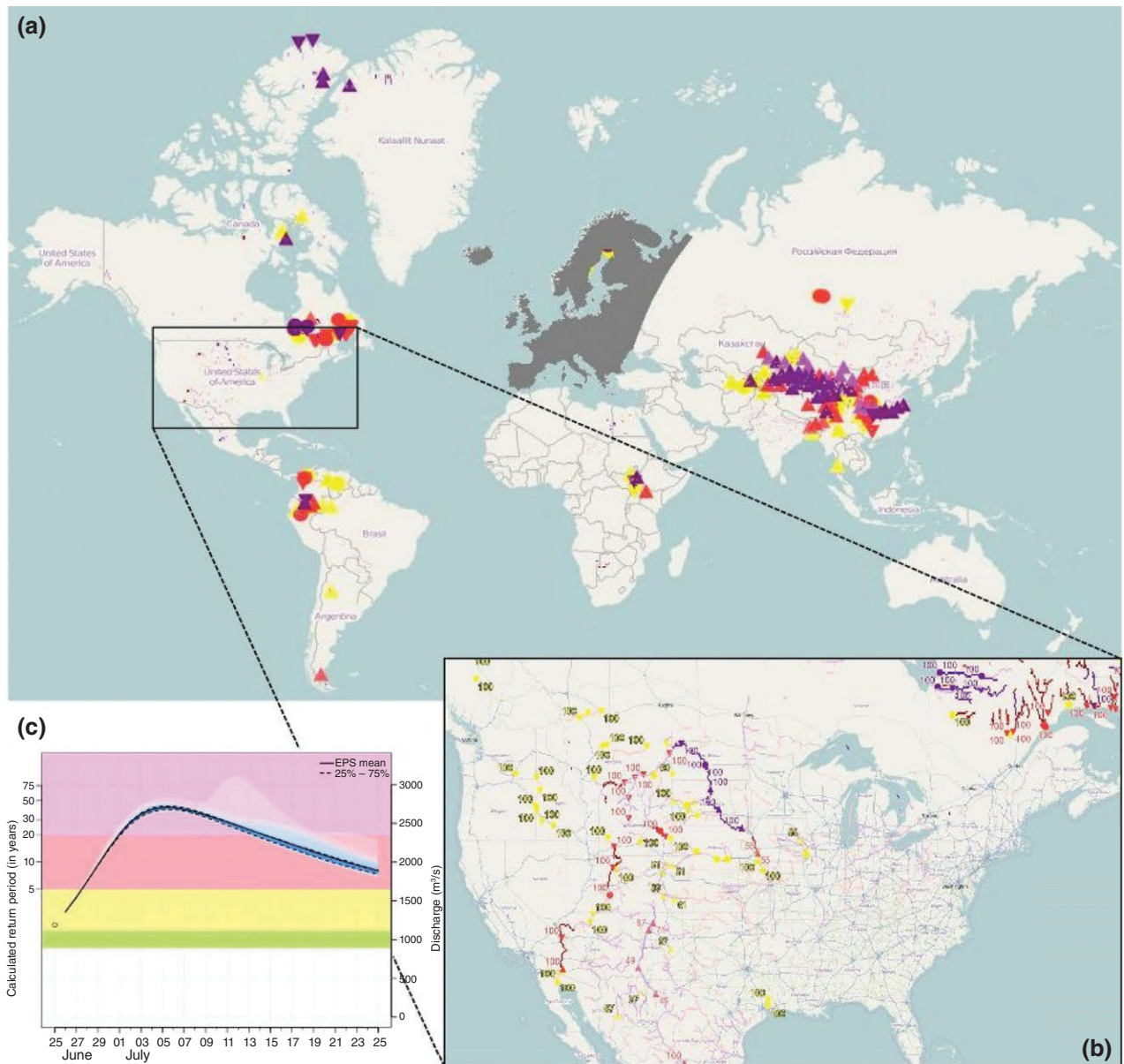
FIGURE 11 | The Global Flood Awareness System (GloFAS) interface showing (a) a global overview of severe (purple), high (red), and medium (yellow) reporting points; (b) a more detailed view of warning points in the U.S.A.; (c) the return period hydrograph with return period thresholds (1.5, green; 2, yellow; 5, red; and 20 years, purple) for one point in the U.S.A. Forecasts are available at www.globalfloods.eu.

return period threshold exceedances and warnings, and persistence diagrams. The ECMWF and JRC do not directly disseminate flood warnings as each country has national procedures to follow, but anyone is able to access and analyze the forecasts for decision-making purposes and research. It is noted that due to the forecast and warning responsibilities within Europe, all countries for which EFAS produces forecasts are removed from the GloFAS interface as these are not publicly available.

Forecast Verification

Alfieri et al.[9] analyzed the performance of GloFAS and found that forecasts were skilful at 58% of stations, which increased to 71% when model bias was removed. Evaluation of the early warning system[9] found that the longest lead times, exceeding 25 days in some regions, are found in large river basins in South America, Africa, and South Asia, while smaller basins have a maximum lead time of 20 days and, in some cases, 10 days. The least skilful forecasts were

for stations in arid and semi-arid regions, such as Australia, Mexico, and the Sahel. Other discrepancies were found in relation to the modeling of snow accumulation and melting processes in HTESSEL and therefore the timing of the peak discharge during spring in snowmelt regions. Evaluation of GloFAS is updated regularly to reflect its continued and ongoing development.

Operational Applications

As of the September 14, 2015, GloFAS has 177 registered users from governmental or other public authorities (~28%), non-governmental organizations (NGOs, ~7%), the private sector (~10%), and from academic/training and/or research institutions (~55%). As with EFAS, GloFAS is used by national services to provide additional early flood information and is used by, for example, civil protection and humanitarian aid organizations who benefit from a global overview of flood events and may have no other source of information for the region of interest. GloFAS is also used by the ERCC for the purpose of compiling reports on natural hazards and flood risk across the globe.

The Global Flood Forecasting Information System

The Global Flood Forecasting Information System (GLOFFIS) is a research-oriented operational system based on Delft-FEWS.[56] GLOFFIS is one of three global systems run by Deltares in the Netherlands; also operational are a storm surge model, GLOSSIS,[57] and a water scarcity system, GLOWA-SIS. These three systems belong to an open, experimental information and communications technology facility, IdLab, and are being used to test new ideas around interoperability, hydrological predictability, big data, and visualization.

Model Components

Similar to the approaches taken by many of the continental-scale flood forecasting systems, GLOFFIS uses several meteorological inputs to drive the hydrological component of the system. The idea behind this is to validate, verify, and inter-compare real-time rainfall (alongside temperature and potential evaporation) products as they become available. The initial conditions are derived from historical forcings based on both the GFS and the ECMWF control forecast (also extracted from the TIGGE archives) and a
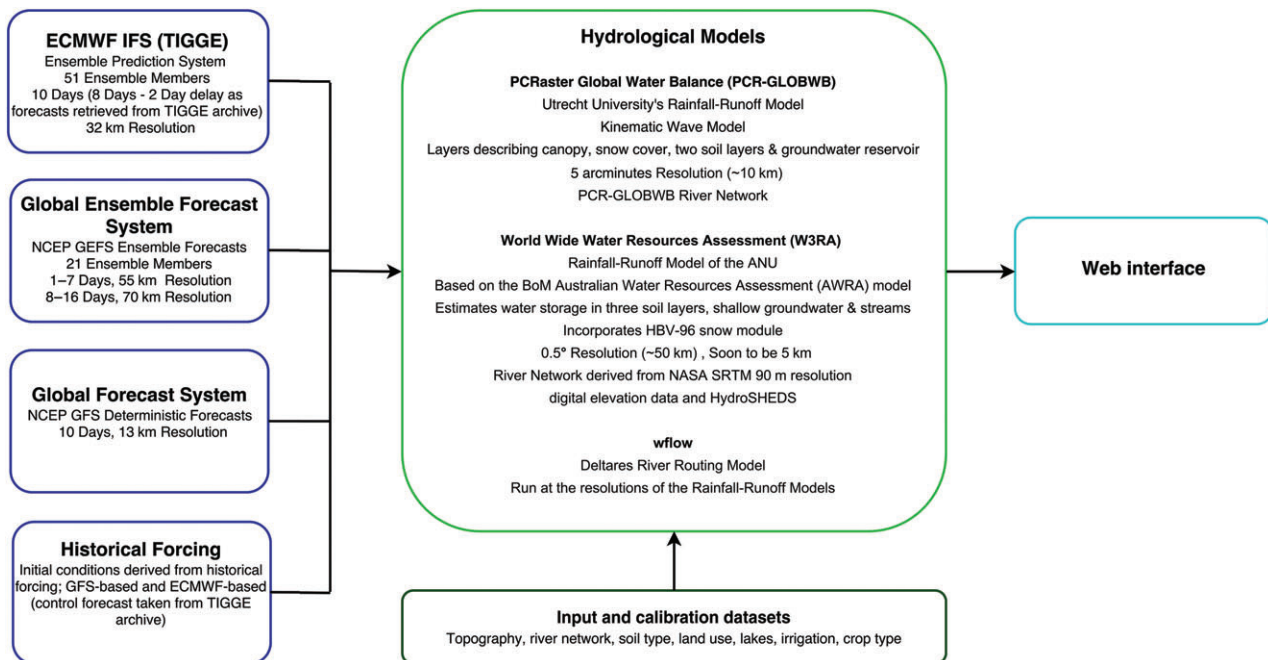

FIGURE 12 | Components of the Global Flood Forecasting Information System (GLOFFIS).

combination of FEWSNET (Africa) and Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Global Daily Precipitation, complimented by GFS temperature and potential evaporation. Each of the NWP inputs are fed into two hydrological models (with multiple initial conditions), PCR-GLOBWB and W3RA, which also incorporate the HBV-96 snow module,[58] to account for snow processes.

The current components and resolution of GLOFFIS are detailed in Figure 12, with plans to update the resolution of the W3RA component to 0.05° (~5km) and implement an improved river network. In the future, the Japan Aerospace Exploration Agency (JAXA) Global Satellite Mapping of Precipitation (GSMaP) and the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) products will also be added as additional datasets from which to derive initial conditions.

Forecast Visualization

As the GLOFFIS and interoperability experiment is a very recent development, many aspects have yet to be implemented. The IdLab is also intended to investigate visualization and data exchange, and for GLOFFIS, multiple visualization and data access and exchange methods will be tested/validated. An example of the Delft-FEWS interface for GLOFFIS is shown in Figure 13. The two forthcoming visualization platforms for GLOFFIS are not yet available, but there is a plan to offer access via a platform similar to the system developed for Guanabara bay[59] and via the Deltares adaguc portal,[60] originally developed by KNMI.[61]

Forecast Verification

Thorough statistical verification of GLOFFIS is underway using available open discharge and meteorological forecast data alongside (real-time) eyeball verification. Real-time discharge data is being collected and can be accessed and compared with the simulated discharge within the Delft-FEWS GLOFFIS platform and reports generated by the system. The verification threshold levels are derived from long historical discharge records and historical simulations, similar to the methods used in other continental- and global-scale forecasting systems.[9]

Operational Applications

Although GLOFFIS is not yet fully implemented, it is being used internally at Deltares and by their customers, with discussions already underway between Deltares and other potential end users of the system. GLOFFIS is intended to be a research tool on predictability and interoperability first and foremost but will be suitable for a variety of applications once fully operational.
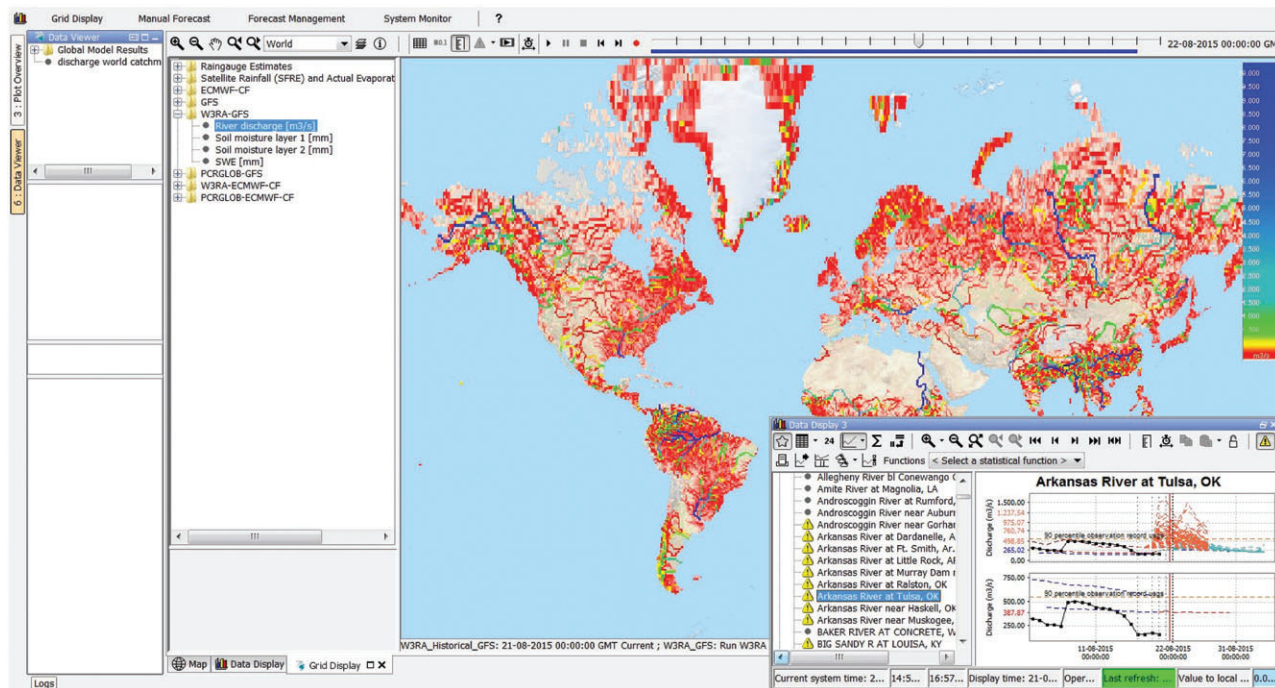

FIGURE 13 | Runoff output of the Global Flood Forecasting Information System (GLOFFIS) W3RA model in the Delft-FEWS forecast platform interface.

THE GRAND CHALLENGES OF GLOBAL-SCALE FLOOD FORECASTING

There are many challenges associated with global-scale flood forecasting. These range from insufficient data and difficulties combining models and computer resource requirements to the cost of running these models and methods of communicating forecasts efficiently. The challenges faced in operational flood forecasting are discussed in detail by Cloke and Pappenberger,[27] Hannah et al.,[62] Wood et al.,[63] Liu et al.,[64] Pappenberger et al.,[65,66] Kauffeldt,[29] Pagano et al.[67] and Bierkens[10]; this section focuses on the current capabilities of the systems reviewed here and discusses some of the grand challenges of global-scale flood forecasting based on the current system's limitations alongside experiences and lessons learned from the development of these systems.

Current Capabilities

Large-scale flood forecasting has only become possible in recent years, and systems such as those outlined in this review are able to produce coarse-scale discharge forecasts at spatial scales covering entire continents or the globe using NWP products and other expertise, comparing these to observed and modeled historic events in order to produce forecasts of flood events in the medium range, typically 7–15 days. Results from EFAS suggest that river flow and flood forecasts driven by meteorological forecasts are able to provide significant added value to the monitoring of European rivers,[68] whilst for GloFAS, results show that the maximum added value is shown '(i) in medium-size river basins, (ii) in those with relatively fast response and (iii) in basins with no definite trend in the seasonal runoff',[9] with lead times of up to 1 month possible in some large river basins.[9] These systems are also capable of producing and disseminating basic forecast, and in some cases, early warning, products in real time and are key in supplementing national and local flood forecasting capabilities while supporting global-scale activities.

A recent study by Pappenberger et al.[66] provides evidence of the economic benefits of large-scale flood early warning systems in addition to the clear benefits of forecasts and early warnings to populations at risk of flooding. The study demonstrates that the monetary benefit of EFAS is ~€400 for every €1 invested, indicating that large-scale flood forecasting systems not only have the capability to provide early awareness of potential severe events but also provide economic benefits through potential avoidance of flood damages.

Improving Data Availability

Grand Challenge: to access data of sufficient quality and length, assimilate new types of observations, and meaningfully incorporate data of inhomogeneous quality.

One of the major challenges in large-scale forecasting lies in the availability of input data of the quality that is required,[62] such as data required for estimation of the initial hydrological state, geographical boundaries of river basins, and large-/global-scale datasets of land use, soil data etc. For example, smaller-scale national flood forecasting systems are often able to assimilate or update discharge information in real time, while continental- and global-scale models are limited by the lack of availability of real-time, open data for this purpose.

Along with the technical challenges associated with accessing and assimilating the data, non-technical data challenges also exist. For example, there are difficulties with retrieving, quality controlling, formatting, archiving, and redistributing the data collected[67] at centers across the globe. This often requires specialized training and staff, for example, at the U.-S. National Weather Service, much of the hydrologists' time is spent processing data and maintaining the infrastructure used to archive and distribute the data, and the stream measurements used in the BoM system are collected by several hundred entities and must be collated before processing.[67]

More international and interdisciplinary data sharing,[62] through institutions such as the Global Runoff Data Centre (GRDC), and cooperation is essential in moving forward with global-scale forecasting efforts and would greatly increase the data available to forecasting centers not just for use in forcing these models but for verification of the forecasts and continuous improvement of forecast accuracy. In order to work towards overcoming this challenge, it is important to contribute to open data policies and ensure that data availability is at the core of all related activities.

Model Parameterization

Grand Challenge: to find regionalization methods and ways to represent sub-grid scale uncertainty on the global scale.

Alongside the problems associated with the data required for forecasting flood events, there are further challenges involved in the parameterization of models and the use of a single model for all catchments across a continent or the globe. Wood et al.[63]

discuss the possibility that much higher resolution forecasting systems will soon be feasible, which would further provide detailed information regarding the storage, movement, and quality of water. In order to implement models of higher resolutions, there are other challenges that must also be addressed; these challenges lie in the parameterization of processes at both current and future spatial resolutions and the 'lack of knowledge involved in evaluating and constraining the uncertainty in those parameters given current and future data availability'.[69]

This challenge could be addressed, for example, by developing scaling theories to represent effective parameterization and associated uncertainties relevant to a global forecasting chain and methods that can incorporate largely varying data and information availability.

Improving Precipitation and Evaporation Forecasts

Grand Challenge: to translate improved precipitation and evaporation forecasts into improved discharge forecasts.

There have been many improvements in NWP and precipitation forecasting thus far, which have enabled global flood forecasting, as discussed earlier in this review. Despite these improvements, there are still limitations in the NWP forecasts that affect the discharge and therefore flood forecasts. Some of these have been discussed, such as difficulties predicting convection[21] and orographic enhancement processes.[23] It is not only precipitation forecasts that need to be further improved but other NWP variables used in hydrometeorological forecasting systems, such as evaporation. The challenge then lies in translating the continuous improvements made to the NWP forecasts into improved discharge forecasts.

Moving forward, it will be important to develop tools and methods, such as satellite measurements, to measure potential evaporation and precipitation on a global level with acceptable accuracy.

Incorporating Anthropogenic Influences

Grand Challenge: to understand which of the anthropogenic influences have a significant impact on hydrological forecasting and therefore need to be included in global forecasting models.

The lack of knowledge of anthropogenic influences on runoff is a major challenge for large-scale flood forecasting.[70] These influences, for which there is currently no global database, include dams and their regulation, reservoirs, weirs, water extraction, irrigation, and river re-routing; some of this activity also goes unreported and unregulated, creating

additional barriers to incorporating information on water management. One of the specific challenges noted by SMHI for Europe is the changes in processes modeled within these systems due to depleted aquifers.

It is also important for these systems to incorporate aspects of anthropogenic influence such as land use and urban areas. Many of the users of these systems require information on potential impacts of the forecast flood events, for example, the number of people likely to be affected and how much agricultural land is threatened. The inclusion of more impact information is one of the current limitations and focuses for the development of EFAS and GloFAS. A further challenge exists in terms of the unevenly distributed global population, which results in sparse data networks in large, unpopulated regions and difficulties in the dissemination and communication of forecasts and warnings; this challenge is specifically mentioned by the BoM for Australia but also exists at the global scale.

In order to account for anthropogenic influences in global flood forecasting systems, one solution would be to map all of these influences and perform a sensitivity analysis to determine which are impacting the forecasts, so that the key anthropogenic influences can be incorporated into the models.

Resources and Costs

Grand Challenge: to quantify, understand, and communicate the values and benefits derived from a global forecast whilst establishing a cost-effective execution of these forecasts.

Thus far, the spatial resolution of global-scale land surface models has largely been constrained by the computational resources required to run global weather models, currently, at best, ~20 km. The monetary costs of producing forecasts using large-scale prediction systems must also be taken into account. While the costs of running these systems are not generally published, the aforementioned study by Pappenberger et al.[66] states that the estimated cost of EFAS (across the four EFAS operational centres, see section The European Flood Awareness System) is €1.8 million per year, with an estimated €20 million in development costs over 10 years. In addition, with each improvement and update to a forecasting system, it also becomes necessary to re-run model climatologies, re-calculate thresholds, and revise decision-making criteria, all of which can be technologically challenging and require significant computational time and resources.[11,20]

As these systems develop, the resources required to run global flood forecasting systems will be reduced, whilst the technology used continues to improve. This will enable more centers to run global models at lower costs and with fewer time constraints in the future.

Effective Communication of Forecasts

Grand Challenge: to communicate uncertainties to a large range of user groups in countries across the globe, some of whom will not be known, and to embed these systems into national warning chains, whilst respecting sensitivities associated with the single voice principle.[43]

A key challenge associated with global-scale flood forecasting stems from the understanding and communication of flood forecasts. For instance, with the move towards ensemble flood forecasting, there is also a need for improved understanding of probabilistic forecasts. Ensemble forecasts produce large amounts of information, and it is vital that the most important information is conveyed appropriately for ease of use and correct interpretation of the forecasts, allowing for well-informed decisions and promoting a common understanding between end users.

One of the current key challenges for EFAS is ensuring that the flood forecast and warning information is easily accessible to a broad range of users from countries across Europe, who interpret the forecasts very differently. This challenge is amplified further when producing forecasts, as with GloFAS and GLOFFIS, for the entire globe and a spectrum of users ranging from experts in the fields of hydrology and meteorology to those with no experience in using these types of products. GloFAS already has a range of partners and end users, from those who are interested in discharge forecasts for specific stations to those who are interested purely in the impact of the floods. An additional consideration is that of the single voice[43] principle, which states that national services constitute the single authoritative voice on weather warnings in their respective countries. As more systems are introduced with the capability of producing forecasts and warnings, the more difficult this principle becomes; in future, it may be that many institutions are able to disseminate warnings and benefit from the wealth of available forecasts and information, and a new challenge of the systems will be to become the trusted source of information.

In order to effectively communicate forecasts and warnings, it is important to co-develop the forecast visualizations and warnings with a large range of users and enable some flexibility for users to customize the interface. International and interdisciplinary cooperation is also key in moving forward with this challenge as issuing forecasts and warnings can be challenging without the existence of a political agreement between upstream and downstream countries for the sharing of information related to floods.[71]

Forecast Evaluation and Intercomparison

Grand Challenge: to find new and novel methods to verify extremes, which are suitable for hydrological forecasting.

Many forecasting systems, including large-scale flood forecasting systems, are moving towards ensemble forecasting methods. While there are many benefits to using a probabilistic approach, a key challenge associated with ensemble flood forecasting is the evaluation of flood forecasts due to the low frequency of occurrence of extreme floods alongside the lack of data from different flood events.[27] The analysis of an ensemble's ability to fully represent the uncertainty is also complex and uncertain in itself.

This relates to a further grand challenge, that of implementing a Flood Forecasting Intercomparison Project to compare various aspects of these large-scale operational flood forecasting systems. This will be a valuable and important project moving forward as these systems become more advanced and widely used for many applications but is currently not undertaken due to the difficulties involved in comparing models of a variety of different scales, with varying system set-ups and interfaces and different objectives and end users. The computational resources required for such a project are also extensive.

To have effective forecast evaluation measures in place, it is important for institutions running these systems to facilitate access to the forecasts so that the forecasts can be evaluated by an unbiased, external entity.

THE FUTURE OF GLOBAL-SCALE FLOOD FORECASTING

Flood forecasting at the large (continental and global) scale is key to providing overviews and early warnings of flood events across the globe, including regions where no alternative local-scale flood forecasts are available. This section outlines aspects of the future of global-scale flood forecasting as we continue to work towards overcoming the grand challenges and move towards ever more valuable multi-hazard forecast and early warning systems.

Adaptive Modeling Strategies

Adaptive modeling strategies involve the idea of adjusting model predictions in real time if discrepancies are observed between the forecast and observations, where discharge measurements are available in real time. This allows the uncertainty in the forecasts to be further constrained. In meteorological applications, this is referred to as data assimilation and is used routinely in weather forecasts and NWP; however, it is often referred to as updating in hydrology and is not widely used at present in applications such as those discussed here.[30] Simple applications of updating require starting new forecasts using available observations (sequential data assimilation[72]), whereas more complex updating involves the adjustment of current predictions to the observations when discrepancies occur, assimilating the new observed data into the model in real time (variational data assimilation[72]). While data assimilation is not used extensively in flood forecasting systems to incorporate observations into the forecasts, this is likely to be increasingly incorporated in future to further improve the accuracy and lead time of large-scale flood forecasts.[63]

An area of research that will be important in moving towards the incorporation of adaptive modeling strategies is the development of data assimilation toolboxes, allowing institutions to use and benefit from data assimilation tools that are otherwise incredibly complex. One example of this is OpenDA, 'an open interface standard for a set of tools to quickly implement data assimilation and calibration for arbitrary numerical models'.[73]

Extended-Range Forecasting

Future advances in global-scale operational flood forecasting are likely to include more long- range forecasting. There already exists an element of river-specific predictability in some large rivers where the movement of a flood wave downstream can take days or weeks, and a flood event is a relatively certain outcome once large amounts of precipitation are recorded upstream. Realistic initial conditions can be beneficial to seasonal prediction; for example, relatively large soil storage capacity leads to long memory of soil moisture, and the accuracy of soil moisture initial conditions may be key in long-range forecasting.[74] The same is true of snow cover and snow pack, particularly in climate zones where snow is the major water resource.[75]

Seasonal forecasts are currently used across a wide range of weather-sensitive sectors, with many operational weather forecasting centers producing seasonal forecasts, which provide 'seasonal-mean estimates' of weather, such as whether the coming season will be wetter or drier than usual.[76] Such forecasts have the potential to aid the forecasting of floods on seasonal time scales, providing crucial information for flood preparedness and mitigation.[77] Seasonal hydrological forecasting has begun to emerge across the globe over the past decade due to the ongoing development of coupled atmosphere–ocean–land general circulation models,[77] while the seasonal water supply forecasts have been used in the U.S. since the 1930s based on snow survey measurements and, later, precipitation data. [78] Yuan et al.[77] highlight several questions related to the future of seasonal hydrological forecasting, from how to combine weather and climate models toward seamless hydrological forecasting to how to improve the prediction of inter-annual variability of variables relevant to hydrological forecasting applications. There also exists the challenge of the effective communication of seasonal flood forecasts and transfer of these forecasts into warnings and actions.[77] The WMO S2S (Sub-seasonal to Seasonal) prediction project[79] aims to improve the understanding and forecast skill of the sub-seasonal and seasonal time scales, with a focus on extreme weather including floods, and will be key in moving towards extended-range flood forecasts.

Flash Flood Forecasting

Flash floods are associated with spatially and/or temporally intense precipitation and can have high societal impacts. For example, 105 out of 139 countries list flash floods as being in the top two of their most important hazards.[80] Despite this, there is currently no global flash flood forecasting system, but continental systems exist in Europe (as part of EFAS),[42,81] northern America,[82] southern Africa,[83] and Australia alongside other national- and basin-scale systems around the globe.[84] These systems often take the form of one or a combination of empirical correlations, unit hydrographs, and hydrological modeling driven by limited area models.[85]

The challenge of creating a global flash flood forecasting system is that global NWP systems typically have a limited resolution of many of the fine spatial scale processes, such as convection, which are responsible for intense precipitation. Increasing the spatial resolution of global NWP systems may reduce this issue and allow for the implementation of a methodology such as that of,[86] which utilizes the surface runoff estimated from HTESSEL to forecast extreme runoff risk. An alternative could be to use

forecasts of parameters that can be used to estimate the likelihood of intense sub-grid scale precipitation arising. For example, the ECMWF NWP model forecasts the convective available potential energy (CAPE) and CAPE-SHEAR parameters that show the atmospheric instability and the ability of supercell formation in the event of deep moisture convection, respectively[87].

With continuous improvements to NWP systems, new continental and global flash flood routines will be developed based on global NWP models.[88] In addition to flash floods, future applications of global flood forecasting and multi-hazard early warning systems will begin to include other types of flooding, for example, coastal storm surges.

Grand Ensemble Techniques

Recent advances in meteorological forecasting and NWP have moved toward multi-model forecasts and grand ensemble techniques. Programs such as TIGGE[89] [The Observing System Research and Predictability EXperiment (THORPEX) Interactive Grand Global Ensemble] have led to advances in ensemble forecasting, predictability, and development of severe weather prediction products in meteorology. In hydrology, combining models for flood forecasting presents an additional challenge (e.g., due to different river networks and climatologies), but despite this, future applications of flood forecasting should move toward the establishment of grand ensemble techniques.[90] In the future, increased access to monthly and sub-seasonal (for example, through the S2S project[79]) forecasts from multiple centers will enable us to push the limits of predictability through use of these grand ensemble techniques.[90]

New Data Possibilities

Alongside the recent and future advances in forecasting systems, other technologies are constantly advancing and will have beneficial impacts on flood forecasting across the globe. For example, new satellites and earth observation technologies for flood observation are being adopted in hydrology to improve flood forecasts.[91,92] García-Pintado et al.[92] discuss several earth observation techniques that have the potential to improve flood detection and forecasting. Improved data from satellites may be able to provide more accurate topographical, land cover, land use, river network and river width information[93]; these are some of the most important data regarding river basin characteristics, and their accuracy is key to flood forecasting systems. Real-time satellite observations of river width during flooding would also serve to improve both forecasts and warnings in real time and verification of the forecasting systems post-event.

Alongside improved databases describing basin and river characteristics, observations of the data used as input to flood forecasting systems and in data assimilation techniques[63] could include snowpack extent, water levels (from altimetry), river discharge, river width, snow, and soil moisture. Continental- and global-scale observations of many of these variables are not currently available, but global coverage from satellites could prove extremely beneficial in large-scale flood forecasting applications, particularly in regions of poor data availability.[69]

CONCLUSIONS

Here, two global- and four continental-scale operational flood forecasting systems have been reviewed, outlining the current state-of-the-art technology in operational large-scale flood forecasting. Producing forecasts at the global scale has only become possible in recent years, with scientific and technological advances and the increasing integration of hydrological and meteorological communities. Due to these recent advances, large-scale flood forecasting systems are able to produce coarse-scale discharge forecasts at spatial scales covering entire continents or the globe using NWP products and other expertise, comparing these to observed and modeled historic events in order to produce medium-range forecasts of flood events.

Many countries are required to prepare for floods that originate outside of their borders. International and interdisciplinary collaboration is necessary in order to overcome many of the challenges involved in transboundary flood forecasting; large-scale forecasting systems have the potential to provide valuable added information about imminent flooding. So far, results from large-scale flood forecasting systems suggest that river flow and flood forecasts are able to provide significant added value to the monitoring of rivers across the globe.[9,67] Many challenges remain for global-scale flood forecasting, from lack of available data of the quality and scale required to the effective communication of forecasts and warnings to varying end users and communities across the globe. Ongoing research aims to overcome these challenges to further improve the accuracy and applicability of large-scale flood forecasting. The systems outlined in this paper are continuously evolving and are already proving to be key in supplementing national and local forecasting capabilities while supporting global-scale activities.

ACKNOWLEDGMENTS

REFERENCES

1. Guha-Sapir G, Below R, Hoyois P. The CRED/EM-DAT International Disaster Database 2015, Université Catholique de Louvain - Brussels – Belgium, 2015. Available at: www.emdat.be. (Accessed January 27, 2015).

2. UNISDR. Sendai framework for disaster risk reduction 2015–2030, 2014. Available at: http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf. (Accessed September 21, 2015).

3. World Meteorological Organisation (WMO). Disaster Risk Reduction (DRR) Programme, Available at: www.wmo.int/pages/prog/drr/projects/Thematic/MHEWS/MHEWS_en.html. (Accessed July 17, 2015).

4. Stephens E, Day JJ, Pappenberger F, Cloke H. Precipitation and floodiness. *Geophys Res Lett* 2015, 42(23): 10316–10323. doi:10.1002/2015GL066779.

5. Braman LS, van Aalst MK, Mason SJ, Suarez P, Ait-Chellouche Y, Tall A. Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008. *Disasters* 2013, 37:144–164. doi:10.1111/j.1467-7717.2012.01297.x.

6. ECMWF. Changes in ECMWF model, Available at: www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model. (Accessed August 5, 2015).

7. Brown A, Milton S, Cullen M, Golding B, Mitchell J, Shelly A. Unified modeling and prediction of weather and climate: a 25-year journey. *Bull Am Meteorol Soc* 2012, 93:1865–1877. doi:10.1175/BAMS-D-12-00018.1.

8. Alfieri L, Salamon P, Pappenberger F, Wetterhall F, Thielen J. Operational early warning systems for water-related hazards in Europe. *Environ Sci Pol* 2012, 21:35–49.

9. Alfieri L, Burek P, Dutra E, Krzeminski B, Muraro D, Thielen J, Pappenberger F. GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrol Earth Syst Sci* 2013, 17:1161–1175.

10. Bierkens MFP. Global hydrology 2015: state, trends, and directions. *Water Resour Res* 2015, 51:1–25. doi:10.1002/2015WR017173.

11. Pappenberger F, Cloke H, Balsamo G, Ngo-Duc T, Oki T. Global runoff routing with the hydrological component of the ECMWF NWP system. *Int J Climatol* 2009, 30:2155–2174.

12. Ramos MH, Mathevet T, Thielen J, Pappenberger F. Communicating uncertainty in hydrometeorological forecasts: mission impossible? *Meteorol Appl* 2010, 17:223–235. doi:10.1002/met.202.

13. Novak DR, Bailey C, Brill KF, Burke P, Hogsett WA, Rausch R, Schichtel M. Precipitation and temperature forecast performance at the Weather Prediction Center. *Weather Forecast* 2013, 29:489–504. doi:10.1175/WAF-D-13-00066.1.

14. Mittermaier M, Roberts N, Thompson SA. A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol Appl* 2013, 20:176–186. doi:10.1002/met.296.

15. Liu Y, Duan Q, Zhao L, Ye A, Tao Y, Miao C, Mu X, Schaake JC. Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin. *Hydrol Process* 2013, 27:57–74. doi:10.1002/hyp.9496.

16. Richardson D, Bidlot J, Ferranti L, Ghelli A, Haiden T, Hewson T, Janousek M, Prates F, Vitart F. *Verification Statistics and Evaluations of ECMWF Forecasts in 2011–2012*. Technical Memorandum 688. Berkshire, England: ECMWF; 2012.

17. Bartholmes J, Todini E. Coupling meteorological and hydrological models for flood forecasting. *Hydrol Earth Syst Sci* 2005, 9:333–346.

18. Lorenz E. The predictability of a flow which contains many scales of motion. *Tellus A* 1969, 21:289–307.

19. Cuo L, Pagano TC, Wang QJ. A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. *J Hydrometeorol* 2011, 12:713–728. doi:10.1175/2011JHM1347.1.

20. Simmons AJ, Hollingsworth A. Some aspects of the improvement in skill of numerical weather prediction. *Q J R Meteorol Soc* 2002, 128:647–677. doi:10.1256/003590002321042135.

21. Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S. Improved weather and seasonal

climate forecasts from multimodel superensemble. *Science* 1999, 285:1548–1550. doi:10.1126/science.285.5433.1548.

22. Olson DA, Junker NW, Korty B. Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Weather Forecast* 1995, 10:498–511. doi:10.1175/1520-0434(1995)010h0498:EOYOQPi2.0.CO;2.

23. Arduino G, Reggiani P, Todini E. Recent advances in flood forecasting and flood risk assessment. *Hydrol Earth Syst Sci* 2005, 9:280–284.

24. Haiden T, Janousek M, Bauer P, Bidlot J, Ferranti L, Hewson T, Prates F, Richardson D, Vitart F. *Evaluation of ECMWF Forecasts, Including 2013–2014 Upgrades*. Technical Memorandum 742. Berkshire, England: ECMWF; 2014.

25. Ebert EE, McBride JL. Verification of precipitation in weather systems: determination of systematic errors. *J Hydrol* 2000, 239:179–202. doi:10.1016/S0022-1694(00)00343-7.

26. Tang Y, Lean HW, Bornemann J. The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorol Appl* 2013, 20:417–426. doi:10.1002/met.1300.

27. Cloke H, Pappenberger F. Ensemble flood forecasting: a review. *J Hydrol* 2009, 375:613–626.

28. Demeritt D, Cloke H, Pappenberger F, Thielen J, Bartholmes J, Ramos MH. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environ Hazards* 2007, 7:115–127. doi:10.1016/j.envhaz.2007.05.001.

29. Kauffeldt A. *Disinformative and Uncertain Data in Global Hydrology: Challenges for Modelling and Regionalisation [dissertation]*. Uppsala: Uppsala Universitet; 2014, 79 p.

30. Shaw E, Beven K, Chappell NA, Lamb R. *Hydrology in Practice*. 4th ed. Oxfordshire, England: Spon Press; 2011, 543 p.

31. Wood AW, Lettenmaier DP. An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys Res Lett* 2008, 35:1–5. doi:10.1029/2008GL034648.

32. Pagano TC, Shrestha DL, Wang QJ, Robertson D, Hapuarachchi P. Ensemble dressing for hydrological applications. *Hydrol Process* 2013, 27:106–116. doi:10.1002/hyp.9313.

33. Buizza R, Houtekamer PL, Pellerin G, Toth Z, Zhu Y, Wei M. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon Weather Rev* 2005, 133:1076–1097. doi:10.1175/MWR2905.1.

34. Leutbecher M, Palmer T. Ensemble forecasting. *J Comput Phys* 2008, 227:2515–3539.

35. Buizza R, Milleer M, Palmer T. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q J R Meteorol Soc* 1999, 125:2887–2908.

36. Clark MP, Hay LE. Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J Hydrometeorol* 2004, 5:15–32. doi:10.1175/1525-7541(2004)005h0015:UOMNWPi2.0.CO;2.

37. Hydrological processes special issue, edited by Hannah L. Cloke, Florian Pappenberger, Florian Pappenberger, Schalk Jan van Andel, Jutta Thielen, Maria-Helena Ramos, Hydrological ensemble prediction systems. *Hydrol Process* 2013, 27:1–4. doi:10.1002/hyp.9679.

38. Stephens E, Cloke H. Improving flood forecasts for better flood preparedness in the UK (and beyond). *Geogr J* 2014, 180:310–316. doi:10.1111/geoj.12103.

39. Bierkens MFP, Bell VA, Burek P, Chaney N, Condon LE, David CH, de Roo A, Doll P, Drost N, Famiglietti JS, et al. Hyper-resolution global hydrological modelling: what is next? "Everywhere and locally relevant". *Hydrol Process* 2015, 29:310–320.

40. Demargne J, Wu L, Regonda SK, Brown JD, Lee H, He M, Seo DJ, Hartman R, Herr HD, Fresch M, et al. The science of NOAA's operational hydrologic ensemble forecast system. *Bull Am Meteorol Soc* 2014, 95:79–98.

41. NOAA (National Oceanic and Atmospheric Administration). NWS Directives System; Operations and Services, http://www.nws.noaa.gov/directives/010/010.htm. (Accessed September 9, 2015).

42. Thielen J, Bartholmes J, Ramos MH, de Roo A. The European Flood Alert System - part 1: concept and development. *Hydrol Earth Syst Sci* 2009, 13:125–140.

43. World Meteorological Organization (WMO). EC statement on the role and operation of National Meteorological and Hydrological Services, Available at: www.wmo.int/pages/governance/policy/ec_statement_nmhs_en.html. (Accessed September 30, 2015)

44. SMHI. About E-HYPE, hypeweb.smhi.se/europehype/about/. (Accessed May 29, 2015).

45. Donnelly C, Andersson JCM, Arheimer B. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol Sci J* 2015. doi:10.1080/02626667.2015.1027710.

46. Lindstrom G, Pers C, Rosberg J, Stromqvist J, Arheimer B. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrol Res* 2010, 41:3–4.

47. *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts; Board on Atmospheric Sciences and Climate; Division on Earth and Life Studies;

National Research Council. Washington, DC: The National Academies Press; 2006, 112 p.

48. NOAA. CHPS Documentation, http://www.nws.noaa.gov/oh/hrl/ general/indexdoc.htm\#hefs. (Accessed June 12, 2015)

49. Burnash R, Ferral R, McGuire R, McGuire R. *A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers*. Sacramento, CA: U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources; 1973, 204 p.

50. Anderson E. *National Weather Service River Forecast System-Snow Accumulation and Ablation Model*. Technical Memorandum NWS HYDRO-17. Washington, DC: US Department of Commerce; 1973.

51. Brown JD, Wu L, He M, Regonda S, Lee H, Seo DJ. Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 1. Experimental design and forcing verification. *J Hydrol* 2014, 519:2869–2889. doi:10.1016/j.jhydrol.2014.05.028.

52. Brown JD, He M, Regonda S, Wu L, Lee H, Seo DJ. Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 2. Streamflow verification. *J Hydrol* 2014, 519:2847–2868. doi:10.1016/j.jhydrol.2014.05.030.

53. Wu H, Adler RF, Tian Y, Huffman GJ, Li H, Wang J. Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour Res* 2014, 50:2693–2717. doi:10.1002/2013WR014710.

54. Yilmaz KK, Adler RF, Tian Y, Hong Y, Pierce HF. Evaluation of a satellite-based global flood monitoring system. *Int J Remote Sens* 2010, 31:3763–3782. doi:10.1080/01431161.2010.483489.

55. NASA. Global Flood and Landslide Monitoring, http://pmm.nasa.gov/trmm/flood-and-landslide-monitoring. (Accessed December 9, 2015).

56. van der Knijff J, Younis J, de Roo A. Lisflood: a GIS-based distributed model for river basin scale water balance and flood simulation. *Int J Geogr Inf Sci* 2010, 24:189–212.

57. Werner M, Schellekens J, Gijsbers P, van Dijk M, van den Akker O, Heynert K. The Delft - FEWS flow forecasting system. *Environ Model Softw* 2013, 40:65–77. doi:10.1016/j.envsoft.2012.07.010.

58. Verlaan M, De Kleermaeker S, Buckman L. GLOSSIS: Global storm surge forecasting and information system 2015. In: *Australasian Coasts & Ports Conference*, Auckland, New Zealand, 15–18 September, 2015.

59. Deltares. The wflow_hbv model, http://schj.home.xs4all.nl/html/wflow_hbv.html. (Accessed June 16, 2015).

60. Deltares. Guanabara Limpa Project Interface, http://guanabaralimpa.deltares.nl/. (Accessed August 21, 2015).

61. Deltares. Adaguc Portal, http://adaguc.deltares.nl/. (Accessed August 21, 2015).

62. Hannah DM, Demuth S, van Lanen HAJ, Looser U, Prudhomme C, Rees G, Stahl K, Tallaksen LM. Large-scale river flow archives: importance, current status and future needs. *Hydrol Process* 2011, 25:1191–1200. doi:10.1002/hyp.7794.

63. Wood EF, Roundy JK, Troy TJ, van Beek LPH, Bierkens MFP, Blyth E, de Roo A, Doll P, Ek M, Famiglietti J, et al. Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour Res* 2011, 47:1944–7973. doi:10.1029/2010WR010090.

64. Liu Y, Weerts AH, Clark M, Hendricks Franssen HJ, Kumar S, Moradkhani H, Seo DJ, Schwanenberg D, Smith P, van Dijk AIJM, et al. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol Earth Syst Sci* 2012, 16:3863–3887. doi:10.5194/hess-16-3863-2012.

65. Pappenberger F, Stephens E, Thielen J, Salamon P, Demeritt D, van Andel SJ, Wetterhall F, Alfieri L. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. *Hydrol Process* 2013, 27:132–146.

66. Pappenberger F, Cloke H, Parker D, Wetterhall F, Richardson D, Thielen J. The monetary benefit of early flood warnings in Europe. *Environ Sci Policy* 2015, 51:278–291.

67. Pagano TC, Wood AW, Ramos MH, Cloke HL, Pappenberger F, Clark MP, Cranston M, Kavetski D, Mathevet T, Sorooshian S, et al. Challenges of operational river forecasting. *J Hydrometeorol* 2014, 15:1692–1707. doi:10.1175/jhm-d-13-0188.1.

68. Alfieri L, Pappenberger F, Wetterhall F, Haiden T, Richardson D, Salamon P. Evaluation of ensemble streamflow predictions in Europe. *J Hydrol* 2014, 517:913–922. doi:10.1016/j.jhydrol.2014.06.035.

69. Wanders N, Karssenberg D, de Roo A, de Jong SM, Bierkens MFP. The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol Earth Syst Sci* 2014, 18:2343–2357. doi:10.5194/hess-18-2343-2014.

70. Widen-Nilsson E, Halldin S, Xu C. Global water-balance modelling with WASMODM: parameter estimation and regionalisation. *J Hydrol* 2007, 340:105–118.

71. Hossain F, Katiyar N. Improving flood forecasting in international river basins. *EOS Trans AGU* 2006, 87:49–60.

72. Rakovec O, Weerts AH, Hazenberg P, Torfs PJJF, Uijlenhoet R. State updating of a distributed

hydrological model with ensemble kalman filtering: effects of updating frequency and observation network density on forecast accuracy. *Hydrol Earth Syst Sci* 2012, 16:3435–3449. doi:10.5194/hess-16-3435-2012.

73. Deltares. OpenDA, https://www.deltares.nl/en/software/openda/. (Accessed December 9, 2015).

74. Fennessy MJ, Shukla J. Impact of initial soil wetness on seasonal atmospheric prediction. *J Clim* 1999, 12:3167–3180.

75. Li H, Luo L, Wood E, Schaake J. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J Geophys Res* 2009, 114:1–10.

76. Weisheimer A, Palmer TN. On the reliability of seasonal climate forecasts. *J R Soc Interface* 2014, 11:1–10. doi:10.1098/rsif.2013.1162.

77. Yuan X, Wood EF, Ma Z. A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *Wiley Interdiscip Rev Water* 2015, 2:523–536. doi:10.1002/wat2.1088.

78. Pagano T, Wood A, Werner K, Tama-Sweet R. Western U.S. water supply forecasting: a tradition evolves. *EOS Forum* 2014, 95:28–29. doi:10.1002/2014EO030007.

79. World Meteorological Organisation (WMO). Subseasonal to Seasonal Prediction Project, http://www.wmo.int/pages/prog/arep/wwrp/new/S2S_project_main_page.html. (Accessed September 21, 2015).

80. World Meteorological Organization (WMO). *Capacity Assessment of National Meteorological and Hydrological Services in Support of Disaster Risk Reduction.* Geneva: World Meteorological Organization; 2008, 338 p.

81. Raynaud D, Thielen J, Salamon P, Burek P, Anquetin S, Alfieri L. A dynamic runoff co-efficient to improve flash flood early warning in Europe: evaluation on the 2013 central European floods in Germany. *Meteorol Appl* 2014, 22:410–418.

82. Gourley J, Erlingis J, Hong Y, Wells E. Evaluation of tools used for monitoring and forecasting flash floods in the United States. *Weather Forecast* 2012, 27:158–173.

83. Georgakakos, K, Graham, R, Jubach, R, Modrick, T, Shamir, E, Spencer, C, Sperfslage, J. 2013. Global Flash Flood Guidance System, Phase 1, HRC Technical Report No. 9. San Diego, CA: Hydrologic Research Center; February 28, 2013, 151 p.

84. Hapuarachchi HAP, Wang QJ. *A Review of Methods and Systems Available for Flash Flood Forecasting.* Water for a Healthy Country National Research Flagship. Clayton: CSIRO; 2008, 61 p.

85. Hapuarachchi HAP, Wang QJ, Pagano TC. A review of advances in flash flood forecasting. *Hydrol Process* 2011, 2:2771–2784.

86. Alfieri L, Pappenberger F, Wetterhall F. The extreme runoff index for flood early warning in Europe. *Nat Hazards Earth Syst Sci* 2014, 14:1505–1515.

87. Tsonevsky I. New EFI parameters for forecasting severe convection. *ECMWF Newsl* 2015, 144:27–32.

88. ECMWF. A global approach to predicting flash floods, http://www.ecmwf.int/en/about/media-centre/news/2015/global-approach-predicting-flash-floods. (Accessed December 9, 2015).

89. ECMWF. TIGGE - The THORPEX Interactive Grand Global Ensemble, http://tigge.ecmwf.int/ (Accessed September 9, 2015).

90. Fan FM, Schwanenberg D, Collischonn W, Weerts A. Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil. *J Hydrol Reg Stud* 2015, 4:196–227. doi:10.1016/j.ejrh.2015.05.012.

91. Khan S, Hong Y, Vergara H, Gourley J, Brakenridge G, De Groeve T, Flamig Z, Policelli F, Yong B. Microwave satellite data for hydrologic modeling in ungauged basins. *IEEE Geosci Remote Sens Lett* 2012, 9:663–667. doi:10.1109/LGRS.2011.2177807.

92. García-Pintado J, Mason DC, Dance SL, Cloke HL, Neal JC, Freer J, Bates PD. Satellite-supported flood forecasting in river networks: a real case study. *J Hydrol* 2015, 523:706–724. doi:10.1016/j.jhydrol.2015.01.084.

93. Yamazaki D, O'Loughlin F, Trigg M, Miller Z, Pavelsky T, Bates P. Development of the global width database for large rivers. *Water Resour Res* 2014, 50:3467–3480.

A2: Complex picture for likelihood of ENSO-driven flood hazard

This paper presents the published version of chapter 4 of this thesis, with the following reference:

Emerton, R., H. L. Cloke, E. M. Stephens, E. Zsoter, S. J. Woolnough and F. Pappenberger, 2017: Complex Picture for Likelihood of ENSO-Driven Flood Hazard, *Nature Communications*, **8**, 14796, doi:10.1038/ncomms14796[*]

Appendix

Complex picture for likelihood of ENSO-driven flood hazard

R. Emerton[1,2,3], H.L. Cloke[1,2], E.M. Stephens[1], E. Zsoter[1,3], S.J. Woolnough[4] & F. Pappenberger[3]

El Niño and La Niña events, the extremes of ENSO climate variability, influence river flow and flooding at the global scale. Estimates of the historical probability of extreme (high or low) precipitation are used to provide vital information on the likelihood of adverse impacts during extreme ENSO events. However, the nonlinearity between precipitation and flood magnitude motivates the need for estimation of historical probabilities using analysis of hydrological data sets. Here, this analysis is undertaken using the ERA-20CM-R river flow reconstruction for the twentieth century. Our results show that the likelihood of increased or decreased flood hazard during ENSO events is much more complex than is often perceived and reported; probabilities vary greatly across the globe, with large uncertainties inherent in the data and clear differences when comparing the hydrological analysis to precipitation.

[1] Department of Geography and Environmental Science, University of Reading, Reading RG6 6AB, UK. [2] Department of Meteorology, University of Reading, Reading RG6 6BB, UK. [3] European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, UK. [4] National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading RG6 6BB, UK. Correspondence and requests for materials should be addressed to R.E. (email: r.e.emerton@pgr.reading.ac.uk).

El Niño Southern Oscillation (ENSO) is the most prominent pattern of inter-annual climate variability[1], and is known to influence river flow[2] and flooding[3–5] at the global scale. In the absence of hydrological analyses, products indicating the likelihood of extreme precipitation are often used as an early indicator of flooding during extreme ENSO events[6]. However, the nonlinearity between precipitation and flood magnitude and frequency[7] means that it is important to assess the impact of ENSO not just on precipitation, but on river flow and flooding. This is especially important as, as stated by Chiew and McMahon[2], 'it is likely that the streamflow-ENSO relationship is stronger than the rainfall-ENSO relationship because the variability in rainfall is enhanced in runoff and because streamflow integrates information spatially'.

Here, a global scale hydrological analysis is performed to estimate the historical probability of increased or decreased flood hazard in any given month during El Niño/La Niña events, assessing the added benefit of directly analysing river flow over the use of precipitation as a proxy for flood hazard.

Historical probabilities provide useful information about typical ENSO impacts based on historical evidence[8,9] and are, as stated by Mason and Goddard[8], 'a better estimate of the future climate than the assumption that seasonal conditions will be the same as average'. Nonetheless, there are some key considerations when using such information. One such consideration is that no two El Niño events are the same[8,10]; differences in the peak amplitude, temporal evolution and spatial pattern of warming are likely to affect the timing and magnitude of the resulting impact on river flow. There are many suggested ways to classify ENSO diversity[11], for example, El Niño events are often described as 'East Pacific' (EP) or 'Central Pacific' (CP), dependent on where the peak warming occurs. While this is an over-simplification of the complexity surrounding ENSO diversity, the location of the peak warming can alter the influence on river flow. An additional consideration is the influence of warming ocean temperatures on ENSO events and their related impacts. Recent studies[12,13] suggest that projected changes in the Walker circulation and associated weakening of equatorial Pacific ocean currents are expected to result in more frequent, and more extreme, El Niño and La Niña events[12,14].

In the past, studies have been limited to reanalysis data sets of no longer than ~40 years[3–5], in which there is a sample of ⩽10 El Niño and ⩽13 La Niña events, or observational data with inconsistent coverage, both spatially and temporally[2]. We have created a twentieth century (1901–2010) model reconstruction of river flow in order to obtain a hydrological data set with consistent global coverage over an extended time period. Research by Essou et al.[15] indicates that global meteorological reanalysis data sets 'have good potential to be used as proxies to observations' in order to force hydrological models, particularly in regions where few observations are available. This data set was created by forcing a research version (described in the Methods) of the Global Flood Awareness System[16,17] (GloFAS) with the ERA-20CM[18] meteorological model reconstruction of the European Centre for Medium-Range Weather Forecasts (ECMWF) to produce a 10-member, 0.5° resolution reconstruction of river flow (from here on, ERA-20CM-R) containing 259,200 grid points covering the global river network (Supplementary Fig. 1). Figure 1 depicts a time series of three key variables used in this study, alongside the timing of the 30 El Niño and 33 La Niña events identified in ERA-20CM-R (see Methods).

Previous work by Ward et al.[4] has looked at the influence of El Niño on flood return periods, quantifying the percentage anomaly during El Niño years in comparison with climatology (defined as the long-term average of historical river conditions or meteorological parameters). To ensure accurate estimation of historical probabilities of ENSO-driven flood hazard, this analysis was replicated using the new ERA-20CM-R data set and gives similar results (Supplementary Fig. 2).

In this study, using a climatology of all years and all El Niño/La Niña years, we calculate the percentage of past El Niño/La Niña events during which the river flow fell in the upper (lower) quartile of climatology, defined here as 'abnormally high (low) flow'. Our results show that the likelihood of increased or decreased flood hazard during ENSO events is much more complex than is often perceived and reported; probabilities vary greatly across the globe, with large uncertainties inherent in the data and clear differences when comparing the hydrological analysis to precipitation.

Results

Historical probabilities during El Niño. Figure 2a shows the historical probabilities for February during an El Niño, with the full set of El Niño and La Niña results presented in Supplementary Figs 7 and 8, respectively. El Niño events tend to span two calendar years, evolving in boreal spring and reaching their peak magnitude in winter of the same year, before decaying into the following spring/summer. Shortly after the peak, February sees some of the highest probabilities and extensive spatial coverage of regions influenced by El Niño (where >40% probability of abnormally high or low river flow represents a significant influence); 34.5% of the land surface indicates a significant increase in the probability of abnormally high or low river flow (19.2% for high, 15.3% for low) compared to any given year.

The influence of El Niño on river flow can be seen as early as June (see Supplementary Fig. 7), shortly after ENSO tends to move into the warm phase, with some regions, mostly confined to the tropics, beginning to see up to a 50% probability of high or low river flow in the ensemble mean. In August and September, much of South America, south of the Amazon River, is somewhat likely (~40–60% probability) to observe higher than normal river flow; however, in November, closer to the typical peak of El Niño events, a reversal to drier conditions across much of Brazil is observed. The southern USA has a high probability (up to 70%) of high river flow from December onwards, while Mexico is another region that experiences a reversal in the influence of El Niño, from decreased flood hazard up until September/October, to increased flood hazard from November onwards. Other regions are much more consistent, such as Indonesia, which has a high certainty of abnormally low river flow throughout the evolution, peak and decay of El Niño. However, it is important to note that across the globe, the uncertainty around these probabilities can be high.

Evaluating the uncertainty. Indeed, the historical probabilities themselves give an indication of the uncertainty in the response of the river flow to ENSO events. Here, the 10 ensemble members of ERA-20CM-R also allow interpretation of the uncertainty in the data set, as each ensemble member represents an equally probable reconstruction of the river flow. To provide an indication of this uncertainty, Fig. 2b shows the range of the probability around the mean probability shown in Fig. 2a. The influence of El Niño is much more certain in some locations; for example, in coastal Ecuador/northern Peru, the probabilities vary by only 9%. These locations (darkest shading, 5–10% range) stand out in Fig. 2b; these are the areas where there is potential to use such historical probabilities as an early indicator of increased or decreased flood hazard, as they tend to give high probabilities combined with small uncertainties. However, much of the globe shows a range of
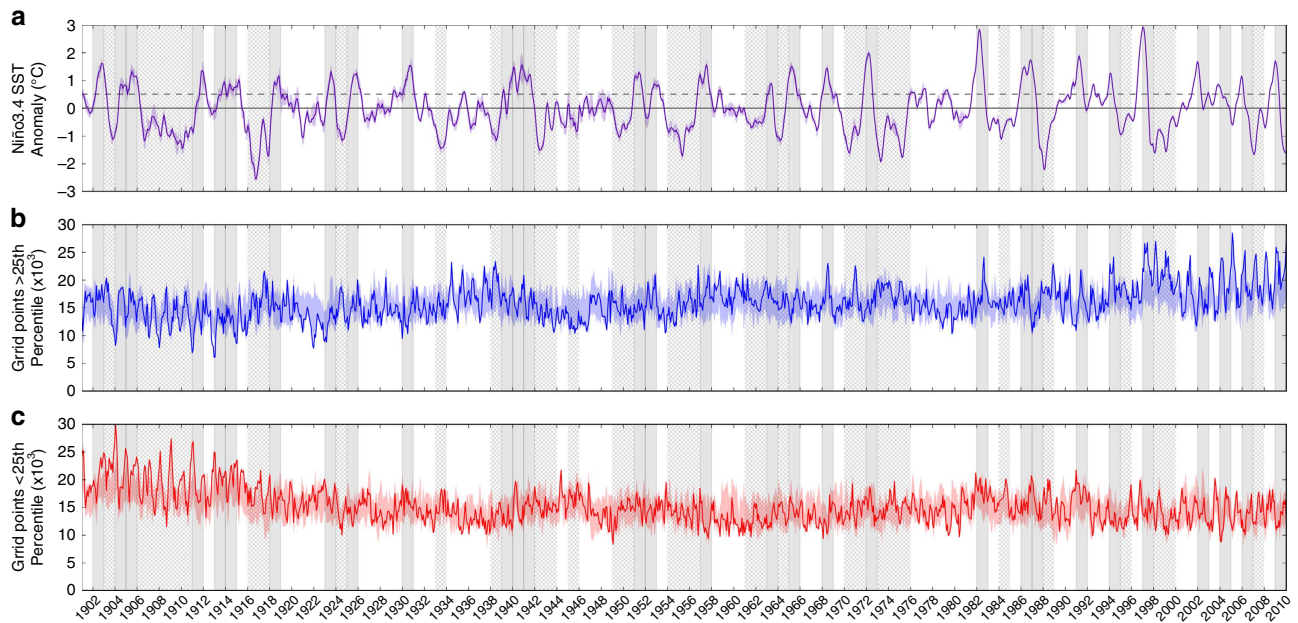
Figure 1 | Time series of three key ERA-20CM-R variables and timing of El Niño and La Niña events. (**a**) Three-month running mean sea surface temperature anomaly in the Niño3.4 region (SSTA3.4), and number of grid points globally in which monthly mean river flow (**b**) exceeds the top 25th percentile and (**c**) falls below the lower 25th percentile. Solid lines show the mean of the 10 ensemble members, while shading indicates the spread of the members. The SSTA3.4 is used to identify El Niño and La Niña years in the data set, highlighted here by the grey shaded and hatched bars, respectively.

20–40%, and some small regions, such as in northwest Spain and eastern Argentina, see a range up to 70% across the ensemble members. The implication is that while some regions see high probabilities of increased flood hazard (e.g., up to 77% in northern Peru), across much of the globe the likelihood is much lower and much more uncertain than might be useful for decision-making purposes.

Importance of the hydrology. Evaluating the historical probabilities of abnormally high or low precipitation, using the ERA-20CM precipitation data set, confirms that there is additional information which can be gained from the hydrological analysis. For example, parts of northern Africa are likely to see high precipitation in February (Supplementary Fig. 3a); however, the River Nile is likely to see dry river conditions (Fig. 2a), indicating that the river is influenced more by upstream rather than local precipitation.

To further highlight the importance of considering the hydrological impacts, Fig. 3 indicates regions, shown in pink (green), where the probability of high river flow is greater (smaller) than that of high precipitation. These differences suggest that the influence of El Niño is more pronounced in the river flow in pink regions, and conversely, green highlights regions where the use of precipitation as a proxy for flood hazard results in an overestimation of the probabilities. This could also indicate that the region is likely to experience a lagged influence of El Niño on river flow. The corresponding results for low flow are presented in Supplementary Fig. 4.

Historical probabilities during La Niña. El Niño events are often followed by a La Niña, the cool phase of ENSO. While La Niña events tend to be less widely discussed in the media, their influence on precipitation is often used as a proxy for flood hazard, as with El Niño. We have therefore extended this analysis to evaluate the probability of increased (or decreased) flood hazard during La Niña years. We find that many regions

influenced by El Niño are likely to observe the opposite response during La Niña. Figure 4 shows these probabilities, again for February, during a La Niña event, with the full set of results shown in Supplementary Fig. 8. It is evident that less of the land surface is significantly influenced by La Niña compared to El Niño during this month (22% of the land surface compared to 34.5%). Probabilities, while still significant, also tend to be lower than for the same month during an El Niño; the highest probability of increased flood hazard shown in Fig. 4a is 67, and 69% for decreased flood hazard. Again, the uncertainty surrounding this mean probability is large (20–40% and in some areas >70%) across much of the globe; this can be seen in Fig. 4b.

Maximum probabilities during El Niño/La Niña. While the monthly maps of historical probabilities give an indicator of the probability of increased (or decreased) flood hazard and when this is likely to occur, it is perhaps useful to consider the event as a whole, as the peak conditions occur at different times across the globe. Figure 5a (b) shows the maximum probability of increased flood hazard during any month of an El Niño (La Niña) event; this provides an overview of whether a region is likely to experience a change in river conditions or not during or following the event. Figure 5 also indicates where the uncertainty surrounding the probabilities is high; this tends to be where the probability is lower, while regions with high probabilities also indicate higher certainty. This analysis further confirms that across much of the globe, such historical probabilities are much more uncertain than is often communicated. The corresponding results for decreased flood hazard are shown in Supplementary Fig. 5.

Comparison with observations. A comparison of the historical probabilities against observed data sets was also undertaken (see Methods and Supplementary Fig. 6). While this proved challenging at the global scale due to a lack of consistent and extensive river flow records in regions of the world where ENSO events

Figure 2 | Historical probability of increased or decreased flood hazard during one month of an El Niño. (**a**) Probability of abnormally high (blue) or low (red) monthly mean river discharge. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology. (**b**) Uncertainty around the probability shown in (a), i.e., the difference between the minimum and maximum of the 10 ensemble members (%). The boxplot (**b**, inset) gives an example graphical representation of the uncertainty range at one grid point, marked on the map by an 'x', where the mean probability indicated in (a) is 63%. The range is given by the difference between the minimum and maximum of the 10 ensemble members; in this case 53 and 81%, giving a 28% range falling in the 20–40% bracket in (b). The month of February is chosen as, occurring shortly after the peak of an El Niño, it sees extensive spatial coverage of land areas influenced by El Niño.

Figure 3 | Comparison of historical probabilities based on precipitation and river flow. Regions where the difference in probability of abnormally high precipitation compared to probability of high river flow, in the month of February during an El Niño, is greater than 10% (based on the ensemble mean). Pink shading indicates that the probability of high precipitation is smaller than the probability of high river flow, while green shading indicates that probabilities are larger for precipitation.

have the most influence, the evaluation suggests a potential overestimation of the probabilities in both the precipitation and river flow reconstructions. This stresses that while these model reconstructions are currently the best available data for such research, there is a need for more extensive river flow observations in regions impacted by ENSO events.

Throughout the results, the complexity and uncertainty surrounding such historical probabilities is evident. Indeed, observations of flooding in February 2016, during the strong 2015–2016 El Niño event, reflect this complex picture of ENSO-driven flood hazard. The expected flooding (based on the results shown in Fig. 2a) in Peru, Bolivia, Argentina and Angola was observed[19]; yet in several other regions, such as Eastern China, Japan and parts of the Middle East, no flood events were recorded. Flooding also occurred in Indonesia despite a high likelihood of dry river conditions. In Kenya and Peru, two examples where flood preparedness actions were taken ahead of El Niño, flooding was much less severe than expected[20,21]. A recent Nature correspondence[22] also highlighted the unexpected winter weather in the USA; California experienced heatwaves rather than prolonged rain events, while Seattle was expecting a worsening drought and instead endured the wettest winter on record (see also Supplementary Fig. 7).

Discussion

We have conducted a global hydrological analysis of ENSO as a predictor of flood hazard based on historical probability estimates using a new, extended-length model reconstruction of river flow. The importance of addressing the hydrology in addition to precipitation is evident in the differences between the probabilities of high river flow and precipitation, and in the ability to further evaluate areas likely to see a lagged influence of El Niño on river flow. We conclude that while it may seem possible to use historical probabilities to evaluate regions across the globe that are more likely to be at risk of flooding during an El Niño/La

Niña, and indeed circle large areas of the globe under one banner of wetter or drier, the reality is much more complex. It is therefore important to undertake research that focuses on the region(s) of interest and consider the impact of ENSO diversity and other drivers of climate variability on the hydrology and flood hazard.

Methods

The new twentieth century river flow data set. For this study, we have created a twentieth century (1901–2010) reconstruction of river discharge, in order to obtain a data set with consistent global coverage over an extended time period. This was achieved by forcing an alternative setup of the GloFAS[16,17] with the 10 ensemble members of the ERA-20CM[18] atmospheric model ensemble of the ECMWF to produce a 10-member ensemble of river discharge for the global river network (ERA-20CM-R).

The operational set-up of GloFAS takes the runoff output from the ECMWF Integrated Forecast System (IFS) and runs this through the Lisflood hydrological routing model[16]. Here, we instead use the Catchment-based Macro-scale Floodplain[23] (CaMa-Flood) routing model to create the river discharge reconstruction at 0.5° resolution from the gridded ERA-20CM runoff data. A map of the CaMa-Flood global river network is given in Supplementary Fig. 1. We note here that the version of GloFAS used in this study is uncalibrated.

While the use of the ERA-20CM model reconstruction allows a consistent analysis at the global scale, and provides a much longer time period over which to study these extreme events, there are limitations that must be considered. ERA-20CM incorporates ENSO and twentieth century climate trends, and assimilates sea-surface temperature and sea ice cover[18]. It does not, however, assimilate atmospheric observations. This is a drawback as the model reconstruction is able to provide a statistical estimate of the climate, but is not able to reproduce synoptic situations. We have therefore undertaken a comparison with the best available precipitation and river discharge observations for the twentieth century and are satisfied that ENSO teleconnections are well-represented in ERA-20CM(-R). Of course, there is further uncertainty introduced when going back as far as the early twentieth century when fewer observations were available; the 10 ensemble members go some way to representing this uncertainty and are a key benefit of this particular data set, and thus are considered throughout this study.

Identifying the El Niño years. To conduct this analysis, we first identified the El Niño/La Niña years in the data set. This was done using the definition that the US National Oceanic and Atmospheric Administration (NOAA) use to declare El Niño (La Niña) conditions operationally[24]. This definition states that the sea
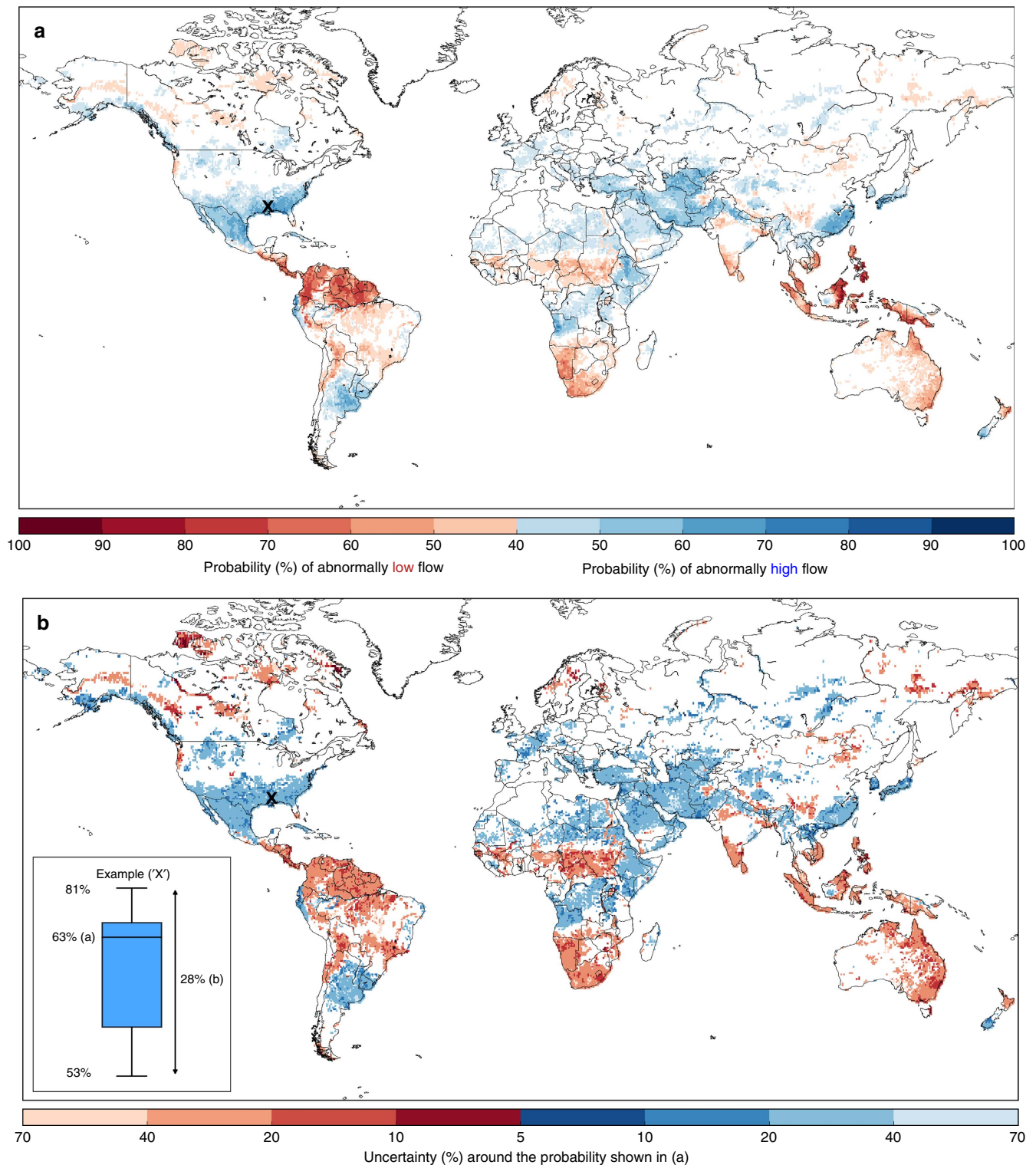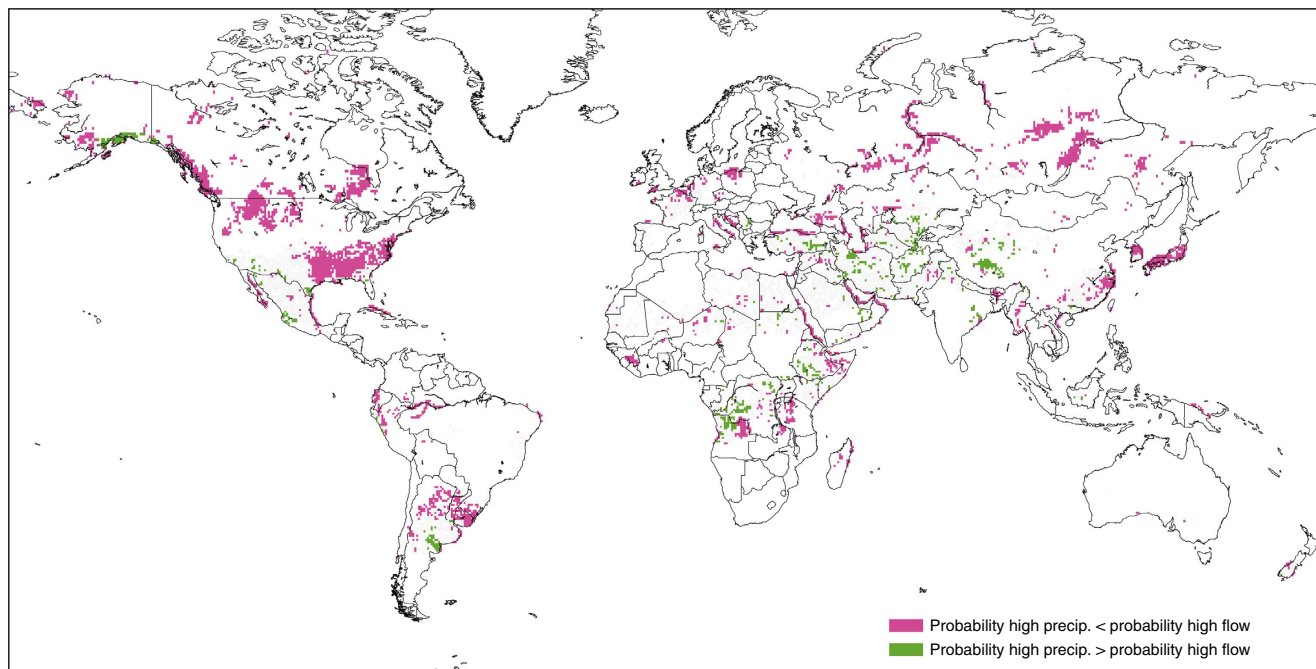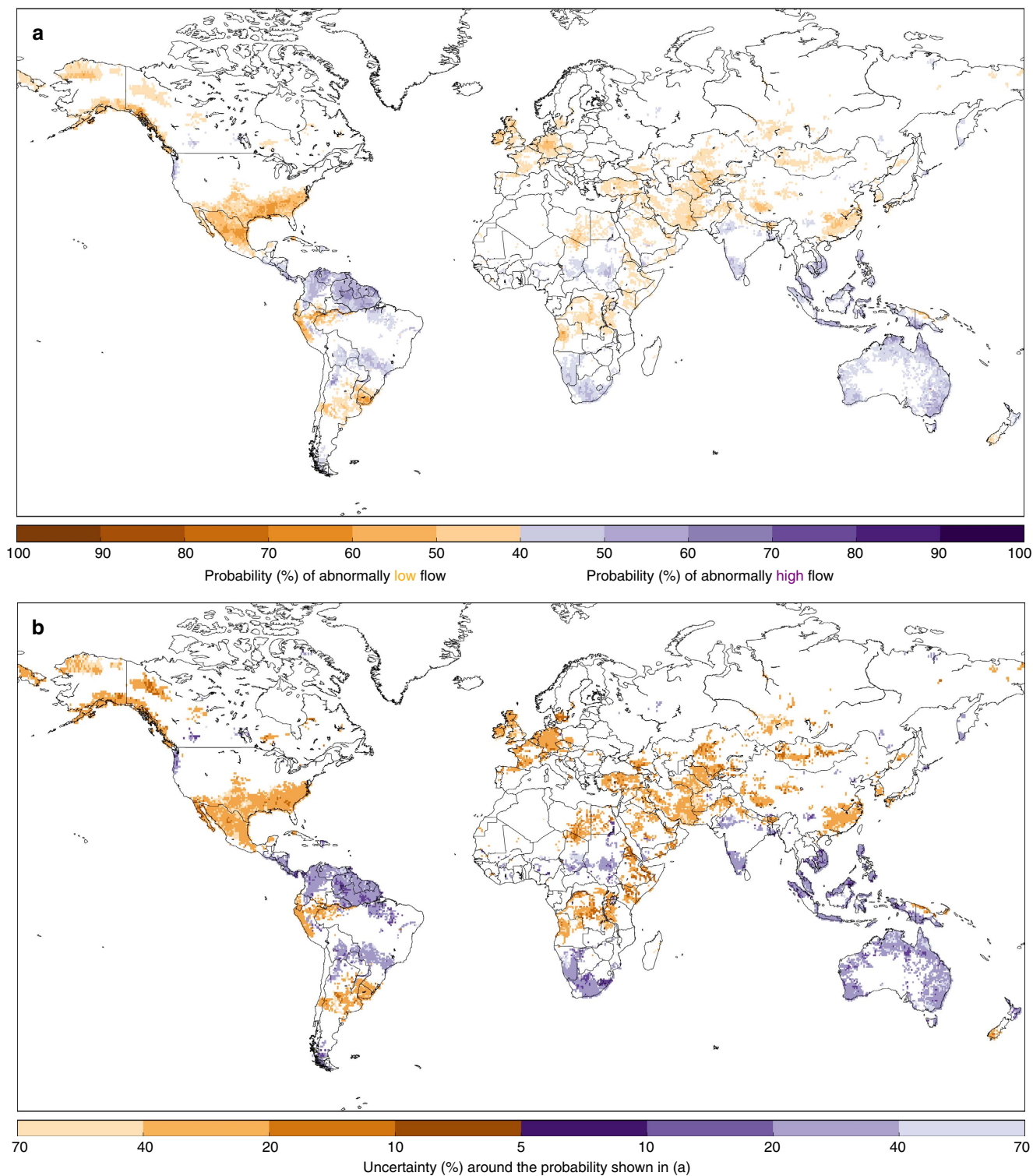
Figure 4 | Historical probability of increased or decreased flood hazard during one month of a La Niña. (a) Probability of abnormally high (purple) or low (orange) monthly mean river discharge in the month of February during a La Niña. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology. **(b)** Uncertainty around the probability shown in (a), i.e., the difference between the maximum and minimum of the 10 ensemble members (%).

surface temperature (SST) anomaly must remain $\geq 0.5\,^\circ\text{C}$ ($\leq 0.5\,^\circ\text{C}$), in the Niño3.4 region in the central Pacific ($5^\circ\,\text{S}$–$5^\circ\,\text{N}$, 170°–$120^\circ\,\text{W}$), for at least five consecutive 3-month periods. Here, we extracted the ERA-20CM SST data and calculated the 3-month running mean SST anomalies for the Niño3.4 region, allowing identification of the 30 (33) years in which El Niño (La Niña) conditions were present from 1901 to 2010. These are listed in Supplementary Table 1, where the El Niño/La Niña year refers to the year in which the event evolves and typically

also reaches its peak, as ENSO events often span 2 years, decaying into the following year. We note that while there is generally a good agreement between the ENSO events identified in ERA-20CM and those published by NOAA[25] for the same period, there are, however, some discrepancies. This is likely due to the different indices/definitions used to identify the ENSO events. For example, in 1977 and 1979, El Niño events are identified by NOAA, using the Multivariate ENSO Index[25], but these are not picked up in this study. In Fig. 1, it is evident that the
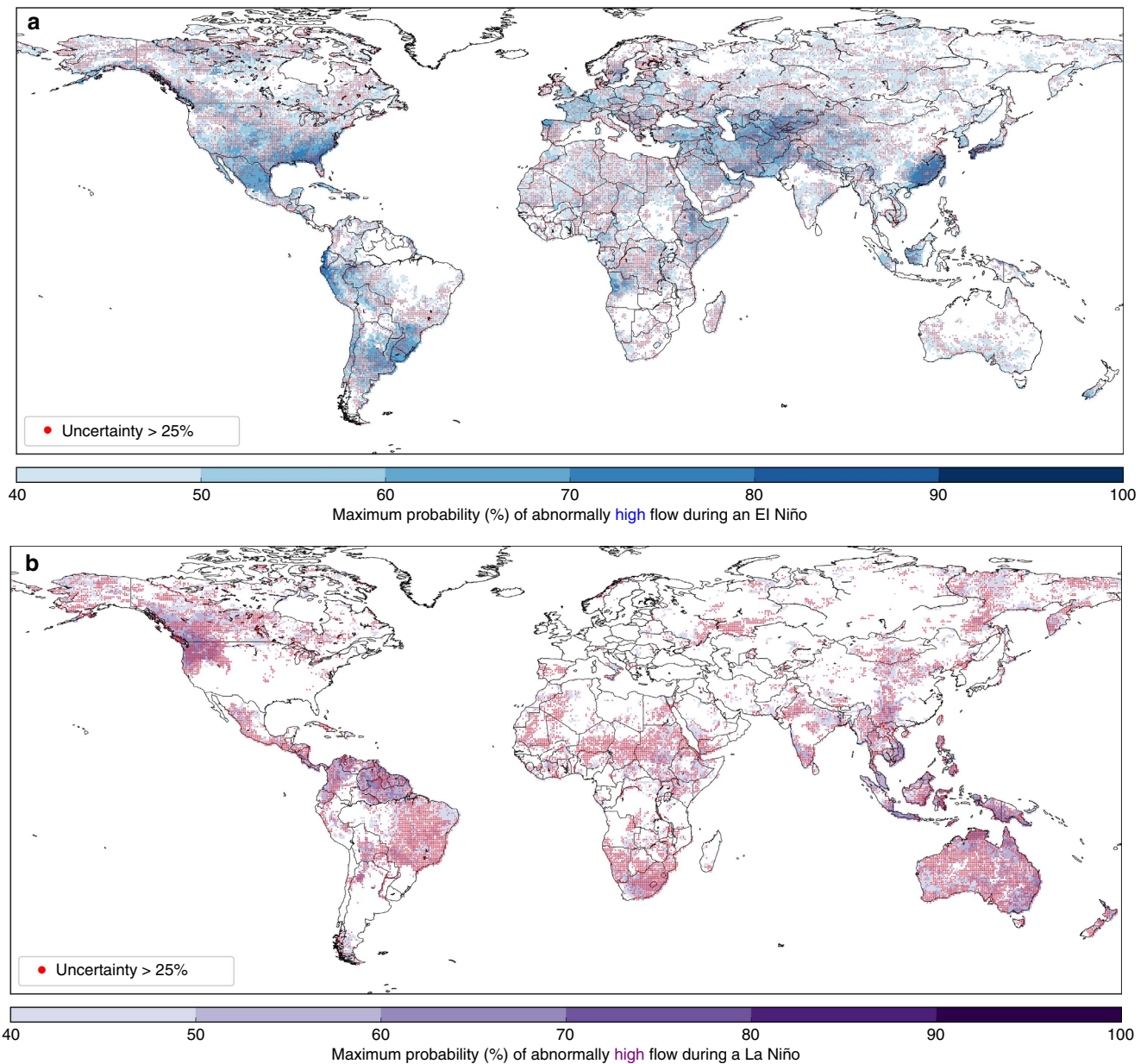
Figure 5 | Maximum probability of increased flood hazard during an ENSO event. Maximum probability of abnormally high river flow in any month during (**a**) an El Niño event and (**b**) a La Niña event. Based on the mean of the 10 ERA-20CM-R ensemble members exceeding the 75th percentile, or falling below the 25th percentile, of the 110-year river discharge climatology during, or shortly after the decay of, an ENSO event. Stippling indicates where the uncertainty surrounding this probability is high, i.e., the range of the ensemble members exceeds 25% probability.

SST did exceed 0.5 °C in ERA-20CM, but this did not persist for long enough to be identified as an event. This is a limitation of the need to use one of the many varying methods of classifying and identifying ENSO events. This method was chosen as it is the most operationally relevant at the time of writing.

Historical probability estimation. For the results presented in this study, the 110-year ERA-20CM-R climatology was used to calculate the upper and lower 25th, 10th and 5th percentiles of river discharge for every grid box. The historical probability of abnormally high or low river flow in any given month was then estimated, through calculation of the percentage of the 30 (33) identified El Niño (La Niña) years in which the river discharge exceeded (high flow) or fell below (low flow) the three percentile thresholds, for each of the 10 ensemble members of ERA-20CM-R. The analysis presented in this paper is based on percentiles so as to avoid potential large errors caused by bias in the data set compared to observations (discussed further below).

Maps of the resulting probabilities were produced based on the mean of the 10 ensemble members. As the number of ENSO events cover a substantial part of the 110-year period, there is a chance of picking up random effects. The maps produced therefore only display results where the probability is significantly greater

than normal, i.e., $\geq 40\%$; an 'event' (occurrence of abnormally high or low flow) with a probability of 40% during one month of an El Niño/La Niña has only a 5% chance of occurring by chance in that month, and thus represents a significant increase in the probability compared to the likelihood of occurring at random.

Additionally, the spread in the ensemble members is designed to reflect the uncertainty in the data set, and can indicate a range of possible outcomes or probabilities. As such, we have further calculated the uncertainty around the mean probability for the whole globe, based on the range across the ensemble members. For each ensemble member, the range between the minimum and maximum ensemble members was calculated for every grid box individually. This allows us to interpret the uncertainty in the probability caused by uncertainty in the data set.

El Niño/La Niña onset tends to occur in boreal spring/early summer and peak in winter[25], before decaying into the following spring. As such, the analysis was undertaken for a period of 2 years; the year of onset, and the following year during which the El Niño/La Niña decays, in order to capture any lagged influence on river flow. Significant influence is shown in the results from June during the El Niño/La Niña year, to the following September (16 months). While it would seem advantageous to summarize the findings by season for simplicity, evaluation of the results shows that the patterns of influence across the globe can change dramatically, in some instances, from one month to the next. Summarizing

these maps into seasons may therefore result in a loss of information for some months.

Difference between river flow and precipitation. A key aim of this paper was to evaluate the added benefit of the hydrological analysis over the use of precipitation as a proxy for flood hazard. To do this, the same method used to estimate the historical probabilities in the river flow reconstruction (ERA-20CM-R) was also applied to the ERA-20CM precipitation reconstruction. The horizontal resolution of the ERA-20CM precipitation data is ∼125 km, while the river flow data is at 0.5° (∼55 km) resolution. To compare these, the results from the precipitation data were remapped to the higher resolution of the river flow data using a simple nearest neighbor remapping algorithm. The difference between the historical precipitation probabilities and river flow probabilities was then calculated for the mean of the 10 ensemble members.

Comparison with observations—precipitation. To evaluate the results shown using the new ERA-20CM(-R) data set, the same method for estimating historical probabilities was also applied to other, related data sets: the Global Precipitation Climatology Centre (GPCC) Full Data reanalysis (GPCC-FD)[26] at 0.5° resolution, and the Global Runoff Data Centre (GRDC) river discharge observations[27]. Again, percentiles are used throughout to allow reliable comparison with observations despite potentially large bias in the model reconstruction values compared to observed values.

The GPCC-FD reanalysis is a global gridded precipitation data set based on interpolated rain gauge data[26]. Comparing the ERA-20CM and GPCC-FD precipitation data sets indicates that the regions influenced by El Niño are well-represented by ERA-20CM (see Supplementary Fig. 3b), and in line with well-known ENSO-sensitive regions, such as Australia, Indonesia, Argentina (the Rio de la Plata delta) and the southern USA—which have been shown to be well-represented in the GPCC-FD[28]. However, the strength of this link appears to be overestimated compared to observations, as the ERA-20CM data show higher probabilities of abnormally high or low precipitation than the GPCC-FD. Some of this overestimation may be caused by the use of the ensemble mean to produce the ERA-20CM maps, as averaging across the 10 ensemble members likely results in a reduction of the variance and we therefore pick up the forced part of the signal.

Comparison with observations—river discharge. As no gridded observational data set of river discharge exists for the global river network, archived station data from the GRDC were used. Criteria for data suitability were chosen to identify those stations which could be of use in this study. Firstly, only stations with at least a 75-year record of observations between 1901 and 2010 were included; these could be stations recording on a daily or monthly basis. Of these, any stations with more than 50% of the data missing were removed. In total, 1287 stations fit the criteria (232 monthly, 1,055 daily), of which the majority have <30% of the data missing. Each of these stations were manually checked to ensure that they correspond to the correct river point (taking into account location and upstream area) on the model river network. A key limitation of using the GRDC observations for this study is that many of these stations lie in river basins outside of the tropics and subtropics—the regions that tend to be most strongly influenced by ENSO events. This highlights the need for more consistent global river flow observations, but in their absence, model reconstructions and reanalyses present the best available data for regional and global scale research based on historical evidence.

To compare the results based on observations with ERA-20CM-R, we produced a reliability diagram (Supplementary Fig. 5) for the historical probability of abnormally high river flow, comparing the forecast (historical) probability of an event (in this case, river flow exceeding a given percentile) with the observed frequency of the event. This was achieved by first locating all grid points in the ERA-20CM-R data set that contain a GRDC station that fit the criteria outlined above. For each percentage band (in 10% bins, as displayed on the maps shown in the Results) of the 'forecast', the observed frequency of river flow exceeding the upper 25th, 10th and 5th percentiles of the 110-year climatology was calculated for each GRDC station, before taking the mean across all stations, and all 16 months used in the analysis (June to the following September). This allows comparison of the predicted probability with the observed frequency. The reliability diagram (Supplementary Fig. 5) and the discrepancy between forecasted and realized probabilities indicates that there is a potential overestimation of the forecasted probabilities. There are limitations, however, in that we have very few, or no, observation stations with which to compare the results for the higher probabilities (Supplementary Fig. 5, inset), particularly in regions that are most significantly influenced by El Niño/La Niña and where reliability may be better, such as the tropics. This suggests that such a reliability analysis may not be fully representative of the results. Additionally, the data records vary from station to station; therefore, the number of El Niño/La Niña years included in the observational record of each station also varies.

Data availability. The ERA-20CM, GPCC-FD and GRDC data that support the findings of this study are publicly available online at http://www.ecmwf.int/en/research/climate-reanalysis/era-20cm-model-integrations, http://www.dwd.de/EN/ourservices/gpcc/gpcc.html and www.bafg.de/GRDC. The ERA-20CM-R data that

support the findings of this study are available from the corresponding author upon reasonable request.

References

1. McPhaden, M. J., Zebiak, S. E. & Glantz, M. H. ENSO as an integrating concept in Earth science. *Science* **314,** 1740–1745 (2006).
2. Chiew, F. H. S. & McMahon, T. A. Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability. *Hydrolog. Sci.* **47,** 505–522 (2002).
3. Ward, P. J., Eisner, S., Florke, M., Dettinger, M. D. & Kummu, M. Annual flood sensitivities to El Nino-Southern Oscillation at the global scale. *Hydrolog. Earth System Sci.* **18,** 47–66 (2014).
4. Ward, P. J. *et al.* Strong influence of El Nino southern oscillation on flood risk around the world. *Proc. Natl Acad. Sci.* **111,** 15659–15664 (2014).
5. Ward, P. J., Kummu, M. & Lall, U. Flood frequencies and durations and their response to El Niño Southern Oscillation: global analysis. *J. Hydrol.* **539,** 358–378 (2016).
6. IRI. IRI Map Room. Available at iridl.ldeo.columbia.edu/maproom/ENSO/Impacts.html (2016).
7. Stephens, E., Day, J. J., Pappenberger, F. & Cloke, H. Precipitation and Floodiness. *Geophys. Res. Lett.* **43,** 316–323 (2015).
8. Mason, S. J. & Goddard, L. Probabilistic precipitation anomalies associated with ENSO. *Bull. Am. Meteorol. Soc.* **82,** 619–638 (2001).
9. Bradley, R. S., Diaz, H. F., Kiladis, G. N. & Eischeid, J. K. ENSO signal in continental temperature and precipitation records. *Nature* **327,** 533–546 (1987).
10. Davey, M. K., Brookshaw, A. & Ineson, S. The probability of the impact of ENSO on precipitation and near-surface temperature. *Clim. Risk Manage.* **1,** 5–24 (2014).
11. Capotondi, A. *et al.* Understanding ENSO Diversity. *Bull. Am. Meteorol. Soc.* **96,** 921–938 (2015).
12. Cai, W. *et al.* ENSO and greenhouse warming. *Nat. Clim. Change* **5,** 849–859 (2015).
13. Cai, W. *et al.* Increasing frequency of extreme El Nino events due to greenhouse warming. *Nat. Clim. Change* **4,** 111–116 (2014).
14. Cai, W. *et al.* Increased frequency of extreme La Nina events under greenhouse warming. *Nat. Clim. Change* **5,** 132–137 (2015).
15. Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F. & Poulin, A. Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? *J. Hydrometeorol.* **17,** 1929–1950 (2016).
16. Alfieri, L. *et al.* GloFAS—global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* **17,** 1161–1175 (2013).
17. Emerton, R. E. *et al.* Continental and global scale flood forecasting systems. *WIREs Water* **3,** 391–418 (2016).
18. Hersbach, H. *et al.* ERA-20CM: a twentieth-century atmospheric model ensemble. *Q. J. R. Meteorol. Soc.* **141,** 2350–2375 (2015).
19. Davies, R. Exceptionally strong El Nino has peaked but impacts to continue. Available at http://floodlist.com/protection/exceptionally-strong-el-nino-peaked-impacts-continue (2016).
20. Muchangi, J. El Nino rains not over yet. Available at http://www.the-star.co.ke/news/2016/01/06/el-nino-rains-not-yet-over-weatherman_c1270441 (2015).
21. Red Cross/Red Crescent Climate Centre. Forecast-based Financing. Available at http://www.climatecentre.org/programmes-engagement/forecast-based-financing (2015).
22. Cohen, J. Weather forecasting: El Nino dons winter disguise as La Nina. *Nature* **533,** 179 (2016).
23. Yamazaki, D., Kanae, S., Kim, H. & Oki, T. A physically-based description of floodplain inundation dynamics in a global river routing model. *Water. Resour. Res.* **47,** W04501 (2011).
24. NOAA. Understanding El Nino. Available at www.noaa.gov/understanding-el-nino (2016).
25. Trenberth, K. E. The definition of El Nino. *Bull Am. Meteorol. Soc.* **78,** 2771–2777 (1997).
26. Schneider, U. *et al.* in *GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.* doi: 10.5676/DWD_GPCC/FD_M_V7_050. (GPCC via DWD, available at http://www.dwd.de/EN/ourservices/gpcc/gpcc.html) (2015).
27. GRDC. *River Discharge Data* (BfG, The Global Runoff Data Centre, Federal Institute of Hydrology, 56068 Koblenz, Germany, 2016).
28. Becker, A. *et al.* A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. *Earth Syst. Sci. Data* **5,** 71–99 (2013).

Author contributions

E.Z. produced the ERA-20CM-R data set; R.E. conceived and posed the research question, carried out the analysis and wrote the paper; R.E., H.L.C., E.M.S., F.P. and S.J.W. designed the study and interpreted the results. All authors commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Emerton, R. *et al.* Complex picture for likelihood of ENSO-driven flood hazard. *Nat. Commun.* **8,** 14796 doi: 10.1038/ncomms14796 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A3: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0

This paper presents the published version of chapter 5 of this thesis, with the following reference:

Emerton, R., E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E. M. Stephens, P. Salamon and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, *Geoscientific Model Development*, **11**, 3327-3346, <u>doi:10.5194/gmd-11-3327-2018</u>[*]

Geoscientific
Model Development

Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0

Rebecca Emerton[1,2]**, Ervin Zsoter**[2,1]**, Louise Arnal**[1,2]**, Hannah L. Cloke**[1,3]**, Davide Muraro**[6]**, Christel Prudhomme**[2,4,5]**,
Elisabeth M. Stephens**[1]**, Peter Salamon**[7]**, and Florian Pappenberger**[2]

[1]Department of Geography & Environmental Science, University of Reading, Reading, UK
[2]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading,UK
[3] Department of Earth Sciences, Uppsala University, Uppsala, Sweden
[4]Centre for Ecology and Hydrology (CEH), Wallingford, UK
[5]Department of Geography and Environment, University of Loughborough, Loughborough, UK
[6]Image Recognition Integrated Systems (IRIS), Ispra, Italy
[7]European Commission, Joint Research Centre (JRC), Ispra, Italy

Correspondence: Rebecca Emerton (r.e.emerton@pgr.reading.ac.uk)

Abstract. Global overviews of upcoming flood and drought events are key for many applications, including disaster risk reduction initiatives. Seasonal forecasts are designed to provide early indications of such events weeks or even months in advance, but seasonal forecasts for hydrological variables at large or global scales are few and far between. Here, we present the first operational global-scale seasonal hydro-meteorological forecasting system: GloFAS-Seasonal. Developed as an extension of the Global Flood Awareness System (GloFAS), GloFAS-Seasonal couples seasonal meteorological forecasts from ECMWF with a hydrological model to provide openly available probabilistic forecasts of river flow out to 4 months ahead for the global river network. This system has potential benefits not only for disaster risk reduction through early awareness of floods and droughts, but also for water-related sectors such as agriculture and water resources management, in particular for regions where no other forecasting system exists. We describe the key hydro-meteorological components and computational framework of GloFAS-Seasonal, alongside the forecast products available, before discussing initial evaluation results and next steps.

1 Introduction

Seasonal meteorological forecasts simulate the evolution of the atmosphere over the coming months. They are designed to provide an early indication of the likelihood that a given variable, for example precipitation or temperature, will differ from normal conditions weeks or months ahead. Will a particular region be warmer or cooler than normal during the next summer? Or will a river have higher or lower flow than normal next winter? Seasonal forecasts of river flow have the potential to benefit many water-related sectors from agriculture and water resources management to disaster risk reduction and humanitarian aid through earlier indications of floods or droughts.

Many operational forecasting centres produce long-range (seasonal) global forecasts of meteorological variables, such as precipitation (Weisheimer and Palmer, 2014). However, at present, operational seasonal forecasts of hydrological variables, particularly for large or global scales, are few and far between. A number of continental-scale seasonal hydro-meteorological forecasting systems have begun to emerge around the globe over the past decade (Yuan et al., 2015a), using seasonal meteorological forecasts as input to hydrological models to produce forecasts of hydrological variables. These include the European Flood Awareness System (EFAS; Arnal et al., 2018; Cloke et al., 2013), the European Service for Water Indicators in Climate Change Adapta-

tion (SWICCA; Copernicus, 2018b), the Australian Government Bureau of Meteorology Seasonal Streamflow Forecasts (Bennett et al., 2017; BoM, 2018), and the USA's National Hydrologic Ensemble Forecast Service (HEFS; Demargne et al., 2014; Emerton et al., 2016). There are also various ongoing research efforts using seasonal hydro-meteorological forecasting systems for forecast applications and research purposes at regional (Bell et al., 2017; Bennett et al., 2016; Crochemore et al., 2016; Meißner et al., 2017; Mo et al., 2014; Prudhomme et al., 2017; Wood et al., 2002, 2005; Yuan et al., 2013) and global (Candogan Yossef et al., 2017; Yuan et al., 2015b) scales. In addition to the ongoing research into improved seasonal hydro-meteorological forecasts at the global scale, an operational system providing consistent global-scale seasonal forecasts of hydrological variables could be of great benefit in regions where no other forecasting system exists and to organisations operating at the global scale (Coughlan De Perez et al., 2017).

Often, in the absence of hydrological forecasts, seasonal precipitation forecasts are used as a proxy for flooding. It has been shown that forecasts of seasonal total rainfall, the most often used seasonal precipitation forecasts, are not necessarily a good indicator of seasonal floodiness (Stephens et al., 2015), and other measures of rainfall patterns, or seasonal hydrological forecasts, would be better indicators of potential flood hazard (Coughlan De Perez et al., 2017).

While it seems a natural next step to produce global-scale seasonal hydro-meteorological forecasts, this is not a simple task, not only due to the complexities of geographical variations in rainfall–run-off processes and river regimes across the globe, but also due to the computing resources required and huge volumes of data that must be efficiently processed and stored and the challenge of effectively communicating forecasts for the entire globe. Indeed, global-scale forecasting for medium-range timescales has only become possible in recent years due to the integration of meteorological and hydrological modelling capabilities, improvements in data, satellite observations, and land-surface hydrology modelling, and increased resources and computer power (Emerton et al., 2016). In addition to continued improvements in computing capabilities, the recent move towards the development of coupled atmosphere–ocean–land models means that it is now becoming possible to produce seasonal hydro-meteorological forecasts for the global river network.

Despite the chaotic nature of the atmosphere (Lorenz, 1963), which introduces a limit of predictability (generally accepted to be ∼ 2 weeks), seasonal predictions are possible as they rely on components that vary on longer timescales and are themselves predictable to an extent. This "second type predictability" (Lorenz, 1993) for seasonal river flow forecasts comes from the initial conditions and large-scale modes of climate variability. The most prominent pattern of climate variability is the El Niño–Southern Oscillation (ENSO; McPhaden et al., 2006), which is known to affect river flow and flooding across the globe (Chiew and McMa-

hon, 2002; Emerton et al., 2017; Guimarães Nobre et al., 2017; Ward et al., 2014a, b, 2016). Other teleconnections also influence river flow in various regions of the globe, such as the North Atlantic Oscillation (NAO), Southern Oscillation (SOI), Indian Ocean Dipole (IOD), and Pacific Decadal Oscillation (PDO), and contribute to the seasonal predictability of hydrologic variables (Yuan et al., 2015a). Coupled atmosphere–ocean–land models are key in representing these large-scale modes of variability in order to produce seasonal hydro-meteorological forecasts.

This motivates the development of an operational global-scale seasonal hydro-meteorological forecasting system as an extension of the Global Flood Awareness System (GloFAS; Alfieri et al., 2013), with openly available forecast products. GloFAS is developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission Joint Research Centre (JRC) and has been producing probabilistic flood forecasts out to 30 days for the entire globe since 2012. In 2016, work began in collaboration with the University of Reading to implement a seasonal outlook in GloFAS, aiming to provide forecasts of both high and low river flow for the global river network up to several months in advance. On 10 November 2017, the first GloFAS seasonal river flow forecast was released. This paper introduces the modelling system, its implementation, and the available forecast products and provides an initial evaluation of the potential usefulness and reliability of the forecasts.

2   Implementation

The GloFAS seasonal outlooks are produced by driving a hydrological river routing model with meteorological forecasts from ECMWF. The forecasts are run operationally on the ECMWF computing facilities. This section provides an overview of the computing facilities, introduces the key hydro-meteorological components of the modelling platform (the meteorological forecast input, hydrological model, and reference climatology), and describes the computational framework of GloFAS-Seasonal.

2.1   ECMWF High-Performance Computing Facility

ECMWF's current High-Performance Computing Facility (HPCF) has been in operation since June 2016 and is used for both forecast production and research activities. The HPCF comprises two identical Cray XC40 supercomputers, each of which is self-sufficient with their own storage and each with equal access to the storage of the other. Each Cray XC40 consists of 20 cabinets of compute notes and 13 storage nodes. One compute node has two Intel Broadwell processors, each with 18 cores, giving 192 nodes (6912 cores) per cabinet. The Cray Aries interconnect is used to connect the processing power. The majority of the nodes of the HPCF are run using the high-performance Cray Linux Environment, a stripped-

down version of Linux, as reducing the number of operating system tasks is critical for providing a highly scalable environment.

In terms of storage, each Cray XC40 has $\sim 10\,\text{PB}$ of storage, and the data handling system (DHS) also comprises two main applications: the Meteorological Archive and Retrieval System (MARS), which stores and provides access to meteorological data collected or produced by ECMWF, and ECFS, which stores data that are not suitable for storing on MARS. The DHS holds over 210 PB of primary data, and the archive increases by $\sim 233\,\text{TB}$ per day. The reader is referred to the ECMWF website at https://www.ecmwf.int/ for further information on the HPCF and DHS.

In addition to the Cray XC40s, the ECMWF computing facility also includes four Linux clusters consisting of 60 servers and 1 PB of storage. The Linux clusters are currently used to run the river routing model used in GloFAS and to produce the forecast products, while the meteorological forcing and ERA5 reanalysis are produced on the HPCF. All data related to GloFAS-Seasonal are stored on the MARS and ECFS archives.

2.2 Hydro-meteorological components

2.2.1 Meteorological forcing

The first model component of the seasonal outlook is the meteorological forecast input from the ECMWF Integrated Forecast System (IFS, cycle 43r1; ECMWF, 2018b). GloFAS-Seasonal makes use of SEAS5, which is the latest version of ECMWF's long-range ensemble forecasting system made operational in November 2017 (ECMWF, 2017a; Stockdale et al., 2018). SEAS5 consists of 51 ensemble members (50 perturbed members and 1 unperturbed control member) and has a horizontal resolution of $\sim 36\,\text{km}$ ($T_{CO}319$). The system, which comprises a data assimilation system and a global circulation model, is run once a month, producing forecasts out to 7 months ahead. Initial pre-implementation testing of SEAS5 has suggested that in comparison to the previous version (System 4), SEAS5 better simulates sea surface temperatures (SSTs) in the Pacific Ocean, leading to improved forecasts of the El Niño–Southern Oscillation (ENSO; Stockdale et al., 2018), which is closely linked to river flow across the globe and can provide added predictability.

SEAS5 is a configuration of the ECMWF IFS (cycle 43r1), including atmosphere–ocean coupling to the NEMO ocean model. SEAS5 is run operationally on the HPCF. Each ensemble member is a complex, HPC-intensive, massively parallel code written in Fortran (version F90). In addition, further complex scripting systems are required to control, prepare, run, post-process, and archive all IFS forecasts. The data assimilation systems used to prepare the initial conditions for the forecasts also make use of Fortran and run on the HPCF. For further information, the reader is referred to the IFS documentation (ECMWF, 2018b).

2.2.2 Land surface component

Within the IFS, which includes SEAS5, the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land, HTESSEL (Balsamo et al., 2011), is used to compute the land surface response to atmospheric forcing. HTESSEL simulates the evolution of soil temperature, moisture content, and snowpack conditions through the forecast horizon to produce a corresponding forecast of surface and subsurface run-off. This component allows for each grid box to be divided into tiles, with up to six tiles per grid box (bare ground, low and high vegetation, intercepted water, and shaded and exposed snow) describing the land surface. For a given precipitation, the scheme distributes the water as surface run-off and drainage, with dependencies on orography and soil texture. An interception layer accumulates precipitation until saturation is reached, with the remaining precipitation partitioned between surface run-off and infiltration. HTESSEL also accounts for frozen soil, redirecting the rainfall and snowmelt to surface run-off when the uppermost soil layer is frozen, and incorporates a snow scheme. Four soil layers are used to describe the vertical transfer of water and energy, with subsurface water fluxes determined by Darcy's law, and each layer has a sink to account for root extraction in vegetated areas. A detailed description of the hydrology of HTESSEL is provided by Balsamo et al. (2011).

HTESSEL comprises a Fortran library of $\sim 20\,000$ lines of code, using both F77 and F90 Fortran versions, and is implemented modularly. While HTESSEL can be run on diverse architectures from a workstation PC to the HPCF, operationally it is run on the HPCF.

2.2.3 River routing model

As HTESSEL does not simulate water fluxes through the river network, Lisflood (Van Der Knijff et al., 2010), driven by the surface and subsurface run-off output from HTESSEL interpolated to the $0.1°$ ($\sim 10\,\text{km}$) spatial resolution of Lisflood is used to simulate the groundwater (subsurface water storage and transport) processes and routing of the water through the river network. The initial conditions used to start the Lisflood model are taken from the ERA5-R river flow reanalysis (see Sect. 2.2.4).

Lisflood is a spatially distributed hydrological model, including a 1-D channel routing model. Groundwater processes are modelled using two linear reservoirs, the upper zone representing a quick run-off component, including subsurface flow through soil macropores and fast groundwater, and the lower zone representing a slow groundwater component fed by percolation from the upper zone. The routing of surface run-off to the outlet of each grid cell, and the routing of run-off produced by every grid cell from the surface, upper,
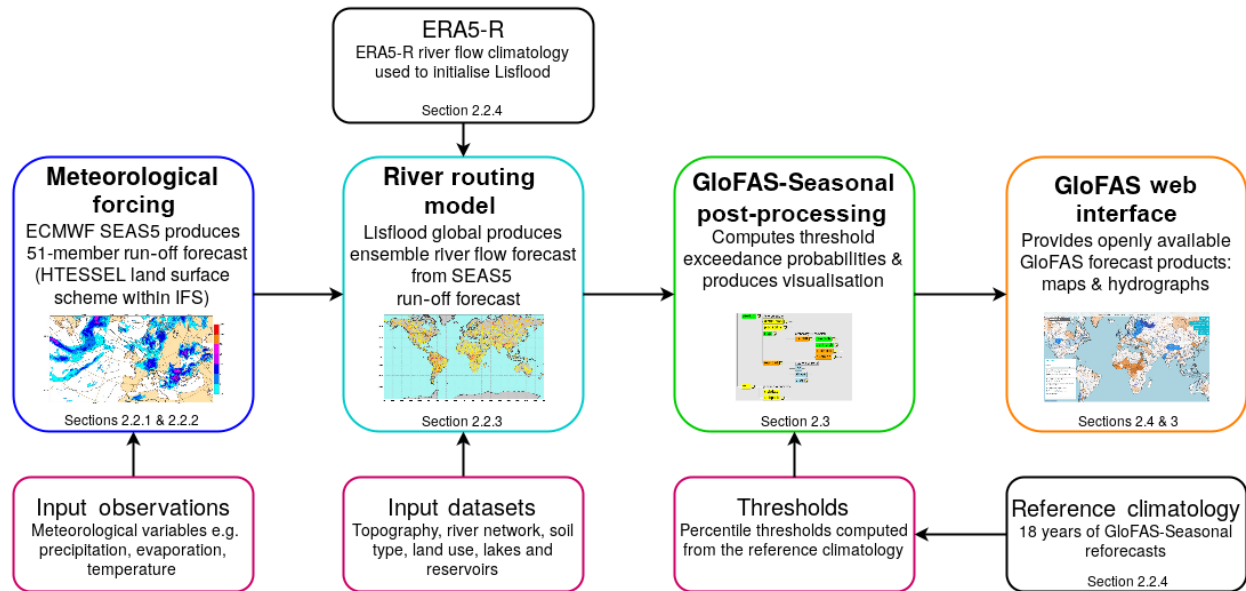
Figure 1. Flowchart depicting the key GloFAS-Seasonal forecasting system components.

and lower groundwater zones through the river network, is done using a four-point implicit finite-difference solution of the kinematic wave equations (Chow et al., 2010). The river network used is that of HydroSHEDS (Lehner et al., 2008), again interpolated to a 0.1° spatial resolution using the approach of Fekete et al. (2001). For a detailed account of the Lisflood model set-up within GloFAS, the reader is referred to Alfieri et al. (2013).

Lisflood is implemented using a combination of PCRaster GIS and Python and is currently run operationally on the Linux cluster at ECMWF.

2.2.4 Generation of reforecasts and reference climatology

In order to generate a reference climatology for GloFAS-Seasonal, the latest of ECMWF's reanalysis products, ERA5, was used. Reanalysis datasets combine historical observations of the atmosphere, ocean, and land surface with a data assimilation system; global models are used to "fill in the gaps" and produce consistent global best estimates of the atmosphere, ocean, and land state. ERA5 represents the current state of the art in terms of reanalysis datasets, providing a much higher spatial and temporal resolution (30 km, hourly) compared to ERA-Interim (79 km, 3-hourly) and better representations of precipitation, evaporation, and soil moisture (ECMWF, 2017b). In order to produce a river flow reanalysis (ERA5-R) for the global river network, the ERA5 surface and subsurface run-off variables were interpolated to 0.1° (∼ 10 km) resolution and used as input to the Lisflood model (see Sect. 2.2.3). ERA5 is currently still in production, and while it will cover the period from 1950 to present

when completed, the full dataset will not be available until 2019. ERA5 is being produced in three "streams" in parallel; at the time of producing the ERA5-R reanalysis, 18 years of ERA5 data were available across the three streams (1990–1992, 2000–2007, and 2010–2016). In addition to the historical climatology, ERA5 is also produced in near real time, with a delay of just ∼ 3 days, allowing its use as initial conditions for the river routing component of the GloFAS-Seasonal forecasts. The ERA5-R reanalysis is thus updated every month prior to producing the forecast. Figure 2 provides an overview of all datasets used in and produced for the development of GloFAS-Seasonal.

Once the ERA5-R reanalysis was obtained, a set of GloFAS-Seasonal reforecasts was produced. From the 25-ensemble-member SEAS5 reforecasts produced by ECMWF, the surface and subsurface run-off variables were used to drive the Lisflood model with initial conditions from ERA5-R. This generated 18 years of seasonal river flow reforecasts (one forecast per month out to 4 months of lead time, with 25 ensemble members at 0.1° resolution). It is the weekly averaged river flow from this reforecast dataset which is used as a reference climatology, including to calculate the high and low flow thresholds used in the real-time forecasts (described in Sect. 2.3).

2.3 GloFAS-Seasonal computational framework

The GloFAS-Seasonal real-time forecasts are implemented and run operationally on the ECMWF computing facilities using ecFlow (Bahra, 2011; ECMWF, 2012), an ECMWF work package used to run large numbers of programmes with dependencies on each other and on time. An ecFlow suite

is a collection of tasks and scheduling instructions with a user interface allowing for the interaction and monitoring of the suite, the code behind it, and the output. The GloFAS-Seasonal suite is run once per month and is used to retrieve the raw SEAS5 forecast data. It runs this through Lisflood and produces the final forecast products and visualisations using the newly developed GloFAS-Seasonal post-processing code.

The GloFAS-Seasonal suite performs tasks (detailed below) such as retrieving data, running Lisflood, computing weekly averages and forecast probabilities from the raw Lisflood river flow forecast data, and producing maps and hydrographs for the interface. It is primarily written in Python (version 2.7), with some elements written in R (version 3.1) and shell scripts incorporating climate data operators (CDOs). The code was developed and tested on OpenSUSE Leap 42 systems.

When a new SEAS5 forecast becomes available (typically on the 5th of the month at 00:00 UTC), the GloFAS-Seasonal ecFlow suite is automatically deployed. The structure of and tasks within the ecFlow suite are shown in Fig. 3. Each "task" represents one script from the GloFAS-Seasonal code. The suite first retrieves the latest raw SEAS5 forecast surface and subsurface variables for all 51 ensemble members (*stagefc* and *getfc* tasks), alongside the river flow reference climatology (see Sect. 2.2.4) for the corresponding month of the forecast (*copywb* task). The Lisflood river routing model (described in Sect. 2.2.3) is then run for each of the 51 ensemble members (*lisflood* task). Lisflood is initialised using the ERA5-R river flow reanalysis (see Sect. 2.2.4) and driven with the SEAS5 surface and subsurface run-off forecast to produce the 4-month ensemble river flow forecast at a daily time step, from which the weekly averaged ensemble river flow forecast is obtained (*average* task). The weekly averages are computed for every Monday–Sunday starting from the first Monday of each month so that the weekly averages correspond from one forecast to the next. While SEAS5 provides forecasts out to 7 months ahead, the first version of GloFAS-Seasonal uses only the first 4 months. This is in order to reduce the data volumes required and to allow for the assessment of the forecast skill out to 4 months ahead before possible extension of the forecasts out to 7 months ahead in the future.

Once the weekly averaging is complete, the *forecast product* section of the suite is deployed, which post-processes the raw forecast output to produce the final forecast products displayed on the web interface. The code behind the *forecast product* section is provided in the Supplement. For a full description of the forecast products, including examples, see Sect. 3. The suite computes the full forecast distribution (*distribution* task), followed by the probability of exceedance for each week of the forecast and for every grid point (*probability* task) based on the number of ensemble members exceeding the high flow threshold or falling below the low flow threshold. The high and low flow thresholds are defined

as the 80th and 20th percentiles of the reference climatology for the week of the year corresponding to the forecast week to use thresholds based on time of year of the forecast. From these weekly exceedance probabilities, the maximum probability of exceedance across the 4-month forecast horizon is calculated for each grid point (*maxprob* task). Basin-averaged maximum probabilities are also produced (*basinprob* task) by calculating the mean maximum probability of exceedance across every grid point at which the upstream area exceeds $1500 \, \mathrm{km}^2$ in each of the 306 major world river basins used in GloFAS-Seasonal (see Sect. 3.1). A minimum upstream area of $1500 \, \mathrm{km}^2$ is chosen, as the current resolution of the global model is such that reliable forecasts for very small rivers are not feasible.

These probabilities are used to produce the forecast visualisation for the web interface (Sect. 3). Firstly, the *map* task produces colour-coded maps of both the river network, again for grid points at which the upstream area exceeds $1500 \, \mathrm{km}^2$, and the major world river basins. The *reppoint* task then produces an ensemble hydrograph and persistence diagrams for a subset of grid points (the "reporting points") across the globe. Further details on the location of reporting points are given in Sect. 3.3. Finally, the *web* task collates and subsequently transfers all data required for the web interface.

This process, from the time a new SEAS5 forecast becomes available, takes $\sim 4\,\mathrm{h}$ on average to complete, with up to 10 tasks running in parallel (for example, running Lisflood for 10 ensemble members at the same time). It is possible to speed up this process by running more ensemble members in parallel; however, the speed is sufficient so that it is not necessary to use further resources to produce the forecast more quickly. GloFAS-Seasonal forecast products are typically produced by the 5th of the month at 05:00 UTC and made available via the web interface on the 10th of the month at 01:00 UTC. This is the earliest that the GloFAS-Seasonal forecasts can be provided publicly under the Copernicus licence agreement. Data are automatically archived at ECMWF as the suite runs in real time; $\sim 285\,\mathrm{GB}$ of data from each SEAS5 forecast are used as input for GloFAS-Seasonal. Each GloFAS-Seasonal forecast run produces an additional $\sim 1.8\,\mathrm{TB}$ of data and makes use of the $\sim 18\,\mathrm{TB}$ reference climatology.

2.4 GloFAS web interface

The GloFAS website is based on a user-centred design (UCD), meaning that user needs are core to the design principles (ISO13407). The website uses Web 2.0 concepts such as simplicity, joy of use, and usability that are synonymous with engaging users. It is a rich internet application (RIA) aiming to provide the same level of interactivity and responsiveness as desktop applications. The website is designed for those engaged in flood forecasting and water resources, as users can browse various aspects of the current forecast or past forecasts in a simple and intuitive way, with spatially distributed
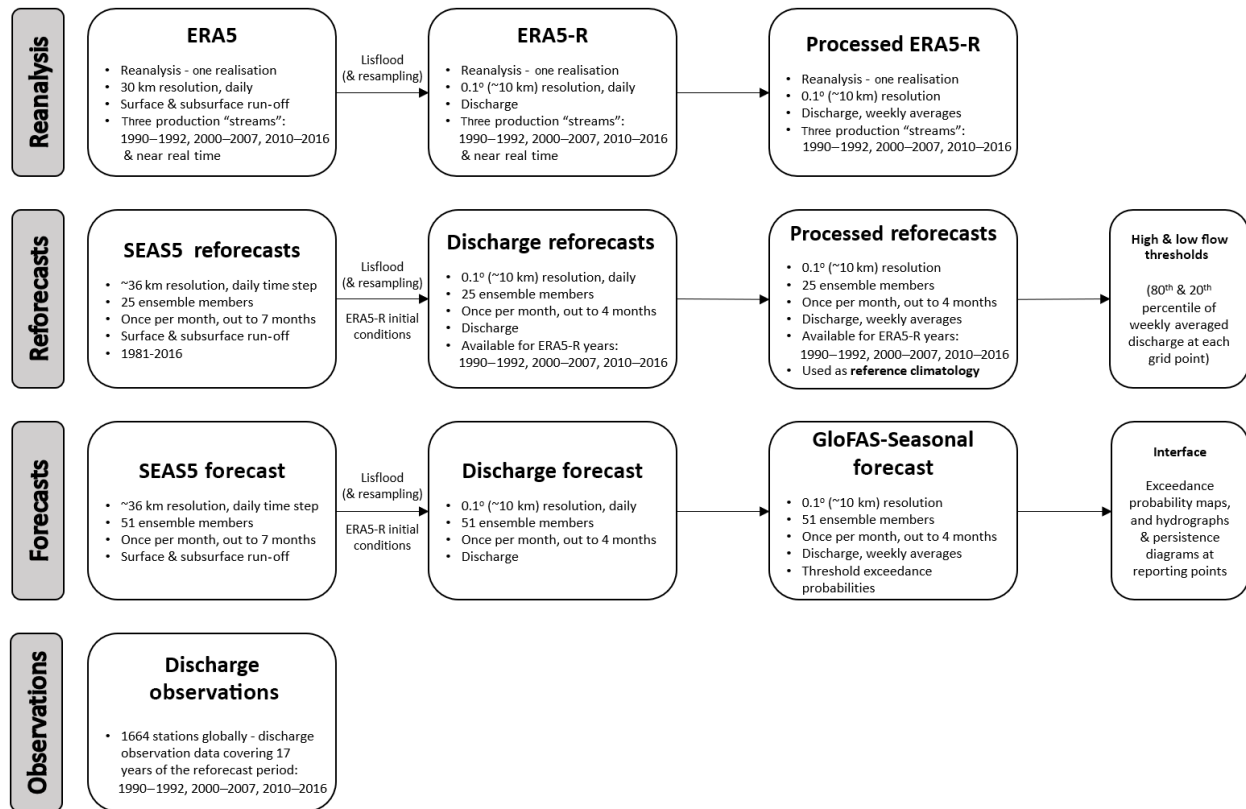
Figure 2. All datasets used and produced for GloFAS-Seasonal, including reanalysis, reforecasts, real-time forecasts, and observations.

information. Map layers containing different information, e.g. flood probabilities for different flood severities, precipitation forecasts, and seasonal outlooks, can be activated and the user can also choose to overlay other information such as land use, urban areas, or flood hazard maps. The interface consists of three principal modules: MapServer, GloFAS Web Map Service Time, and the Forecast Viewer. These are outlined below.

2.4.1 MapServer

MapServer (Open Source Geospatial Foundation, 2016) is an open source development environment for building spatially enabled internet applications developed by the University of Minnesota. MapServer has built-in functionality to support industry standard data formats and spatial databases, which is significant to this project, and the support of popular Open Geospatial Consortium (OGC) standards including WMS. In order to exploit the potential of asynchronous data transfer between server and client, the GloFAS raster data have to be divided into a grid of adequate dimensions and an optimal scale sequence.

2.4.2 GloFAS Web Map Service Time

The OpenGIS Web Map Service (WMS) is a standard protocol for serving geo-referenced map images over the internet. A web map service time (WMS-T) is a web service that produces maps in several raster formats or in vector format that may come simultaneously from multiple remote and heterogeneous sources. A WMS server can provide support to temporal requests (WMS-T) by providing a TIME parameter with a time value in the request.

The WMS specification (OGC, 2015) describes three HTTP requests; *GetCapabilities*, *GetMap*, and *GetFeatureInfo*. *GetCapabilities* returns an XML document describing the map layers available and the server's capabilities (i.e. the image formats, projections, and geographic bounds of the server). *GetMap* returns a raster map image. The request arguments, such as the layer ID and image format, should match those listed as available in the GetCapabilities return document. *GetFeatureInfo* is optional and is designed to provide WMS clients with more information about features in the map images that were returned by earlier GetMap requests. The response should contain data relating to the features nearest to an image coordinate specified in the GetFeatureInfo request. The structure of the data returned is not defined in the specification and is left up to the WMS server
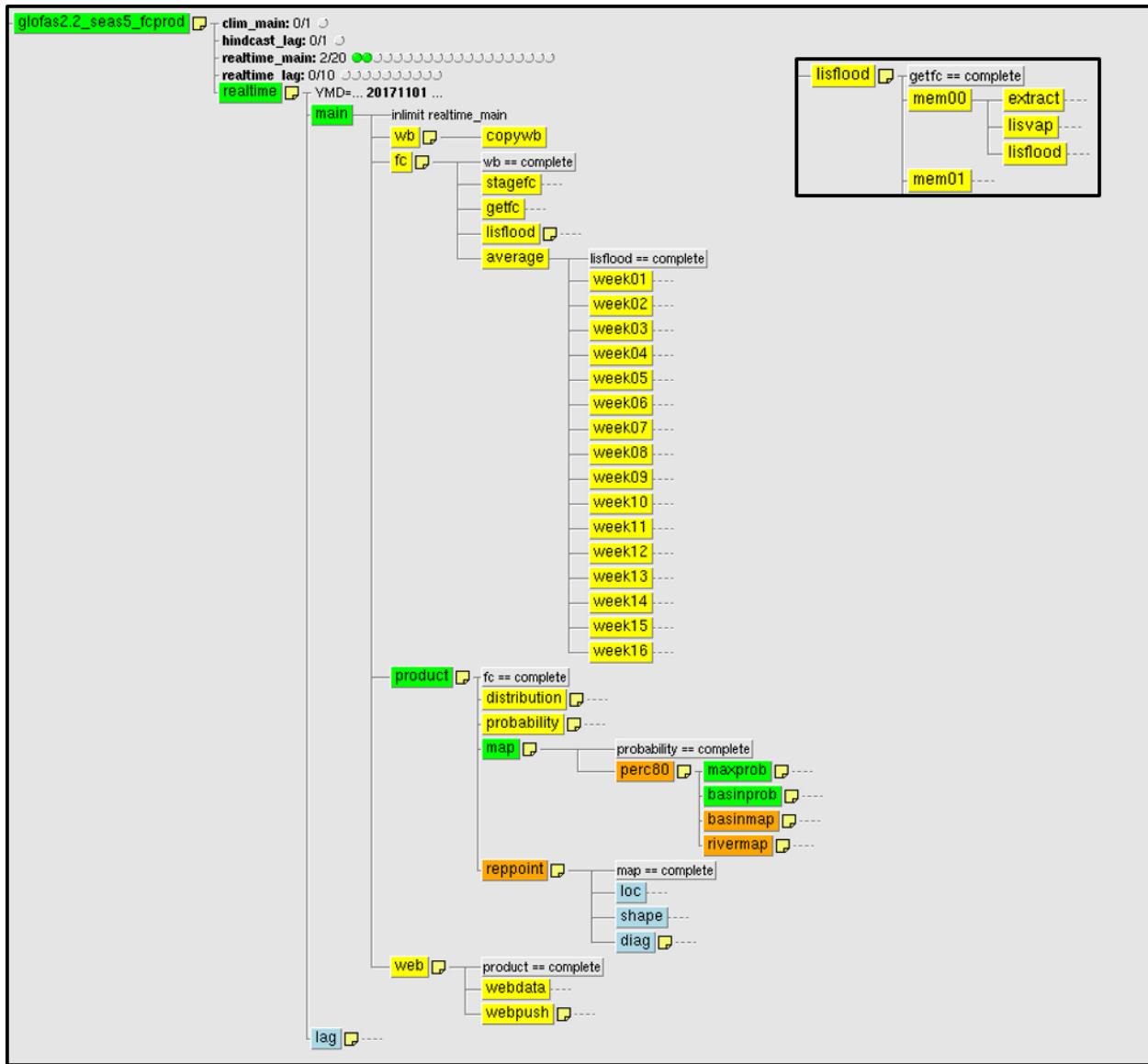
Figure 3. The GloFAS-Seasonal ecFlow suite. The inset image shows the sub-tasks within the Lisflood task for 1 of the 51 ensemble members. Colours indicate the status of each task. Yellow: complete, green: active, orange: suspended, pale blue: waiting, turquoise (not shown): queued, and red (not shown): aborted or failed. Grey boxes indicate dependencies; for example, "lisflood == complete" indicates that the Lisflood task and all Lisflood sub-tasks must have successfully completed in order for the average task to run.

implementation. The GloFAS WMS-T (GloFAS, 2018b) can be freely used, allowing access to the GloFAS layers in any GIS environment, such as QGIS (QGIS Development Team, 2017) or ArcMAP (Environmental Systems Research Institute, 2018). The user manual for the GloFAS WMS-T is available via the GloFAS website (GloFAS, 2018a).

2.4.3 Forecast Viewer

The GloFAS Forecast Viewer is based on the model view controller (MVC) architectural pattern used in software engineering. The pattern isolates "domain logic" (the applica-

tion logic for the user) from input and presentation (user interface, UI), permitting the independent development, testing, and maintenance of each. A fundamental part of this is the AJAX (asynchronous JavaScript and XML) technology used to enhance user-friendly interfaces for web mapping applications. AJAX technologies have a number of benefits; the essential one is removing the need to reload and refresh the whole page after every event. Careful application design and component selection results in a measurably smaller web server load in geodata rendering and publishing, as there is no need to link and send the whole html document, just the relevant part that needs to be changed.

GloFAS uses OpenLayers (OpenLayers, 2018) as a WMS client. OpenLayers is a JavaScript-based web mapping toolkit designed to make it easy to put a dynamic map on any web page. It does not depend on the server technology and can display a set of vector data, such as points, with aerial photographs as backdrop maps from different sources. Closely coupled to the map widget is a layer manager that controls which layers are displayed with facilities for adding, removing, and modifying layers. The new layers associated with GloFAS-Seasonal are described in the following section.

3　Forecast products

The GloFAS seasonal outlook is provided as three new forecast layers in the GloFAS Forecast Viewer: the basin overview, river network, and reporting point layers. Each of the three layers represents a different forecast product described in the following sections. Information on each of the layers is also provided for end users of the forecasts under the dedicated "Seasonal Outlook" page of the GloFAS website.

3.1　Basin overview layer

The first GloFAS seasonal outlook product is designed to provide a quick global overview of areas that are likely to experience unusually high or low river flow over the coming 4 months. The "basin overview" layer displays a map of 306 major world river basins colour coded according to the maximum probability of exceeding the high (blue) or low (orange) flow thresholds (the 80th and 20th percentiles of the reference climatology, respectively) during the 4-month forecast horizon. This value is calculated for each river basin by taking the average of the maximum exceedance probabilities at each grid cell within the basin (using only river pixels with an upstream area $> 1500 \, \mathrm{km}^2$). The three different shades of orange–blue indicate the probability: dark ($>$ 90 %), medium (75 %–90 %), and light (50 %–75 %). Basins that remain white are those in which the probability of unusually high or low flow does not exceed 50 % during the 4-month forecast horizon. An example is shown in Fig. 4.

As mentioned in Sect. 2.2.3, the Lisflood river network is based on HydroSHEDS (Lehner et al., 2008). In order to generate the river basins used in GloFAS-Seasonal, the corresponding HydroBASINS (Lehner and Grill, 2013) data were used. HydroBASINS consists of a suite of polygon layers depicting watershed boundaries at the global scale. These watersheds were manually merged using QGIS (QGIS Development Team, 2017) to create a global polygon layer of major river basins based on the river network used in the model.

3.2　River network layer

The second map layer provides similar information at the sub-basin scale by colour-coding the entire model river net-

work according to the maximum exceedance probability during the 4-month forecast horizon. This allows the user to zoom in to their region of interest and view the forecast maximum exceedance probabilities in more detail. Again, only river pixels with an upstream area $> 1500 \, \mathrm{km}^2$ are shown. The same colour scheme is used for both the basin overview and river network layers, with blue indicating high flow (exceeding the 80th percentile), orange low flow (falling below the 20th percentile), and darker colours indicating higher probabilities. In the river network layer, additional colours also represent areas where the forecast does not exceed 50 % probability of exceeding either the high or low flow threshold (light grey) and where the river pixel lies in a climatologically arid area such that the forecast probability cannot be defined (darker grey–brown). Examples of the river network layer can be seen in both Fig. 4 (globally) and Fig. 5 (zoomed in).

3.3　Reporting points layer

In addition to the two summary map layers, reporting points are provided at both static and dynamic locations throughout the global river network, providing additional forecast information: an ensemble hydrograph and a persistence diagram.

Static points originally consisted of a selection of gauged river stations included in the Global Runoff Data Centre (GRDC; BfG, 2017); this set of points has since been expanded to further include points at locations of particular interest to GloFAS partners. There are now $\sim$ 2200 static reporting points in the GloFAS interface.

Dynamic points are generated to provide the additional forecast information throughout the global river network, including river reaches for which there are no static points. These points are obtained for every new forecast based on a set of selection criteria adapted from the GloFAS flood forecast dynamic point selection criteria (Alfieri et al., 2013).

- The maximum probability of high (low) river flow (exceeding or falling below) the 80th (20th) percentile of the reference climatology) during the 4-month forecast horizon must be ≥ 50 % for at least five contiguous pixels of the river network.

- The upstream area of the selected point must be $\geq$ $4000 \, \mathrm{km}^2$.

- Dynamic reporting points are generated starting from the most downstream river pixel complying with the previous two selection criteria. A new reporting point is then generated every 300 km upstream along the river network, unless a static reporting point already exists within a short distance of the new dynamic point or the forecasts further upstream no longer comply with the previous two criteria.

Reporting points are displayed as black circles in the "reporting points" seasonal outlook layer. An example is shown in
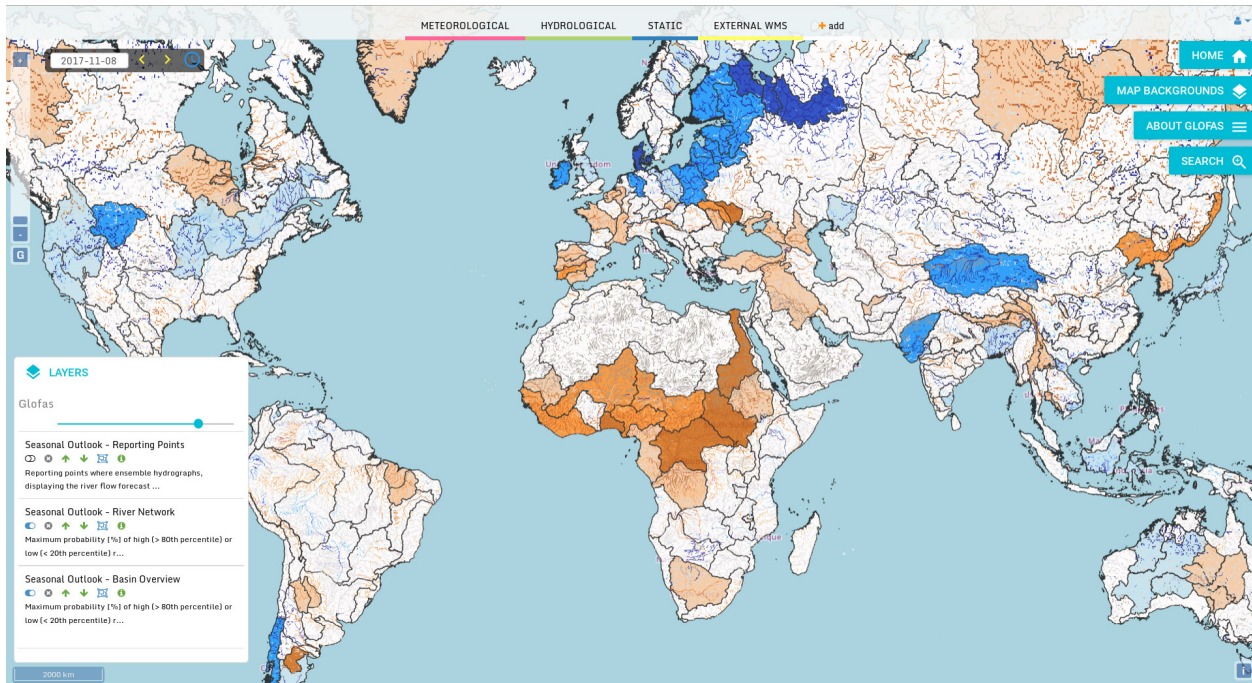
Figure 4. Example screenshot of the seasonal outlook layers in the GloFAS web interface. Shown here are both the "basin overview" layer and "river network" layer, both indicating the maximum probability of unusually high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability: darkest shading indicates > 90 % probability, medium shading indicates 75 %–90 % probability, and light shading indicates 50 %–75 % probability. A white basin or light grey river pixel indicates that the forecast does not exceed 50 % probability of high or low flow during the forecast horizon. Legends providing this information are available for each layer by clicking on the green "i" next to the layer toggle (shown at the bottom left in this example).

Fig. 5. Clicking on a reporting point brings up a new window containing a hydrograph and persistence diagram alongside some basic information about the location, such as the latitude and longitude, and the upstream area of the point in the model river network. The number of dynamic reporting points can vary from one forecast to the next due to the criteria applied; for example, the March 2018 forecast included ∼ 1600 dynamic points in addition to the static points, and thus ∼ 3800 reporting points were available globally.

The ensemble hydrographs (also shown in Fig. 5) display a fan plot of the ensemble forecast of weekly averaged river flow out to 4 months, indicating the spread of the forecast and associated probabilities. Also shown are thresholds based on the reference climatology: the median and the 80th and 20th percentiles. These thresholds are displayed as a 3-week moving average of the weekly averaged river flow for the given threshold for the same months of the climatology as that of the forecast (i.e. a forecast for J–F–M–A also displays thresholds based on the reference climatology for J–F–M–A). This allows for a comparison of the forecast to typical and extreme conditions for the time of year.

Persistence diagrams (see Fig. 5) show the weekly probability of exceeding the high and low flow thresholds for the current forecast (bottom row) and previous three forecasts colour coded to match the probabilities indicated in the map layers. These diagrams are provided in order to highlight the evolution of the forecast, which can indicate whether the forecast is progressing consistently or whether behaviour is variable from month to month.

4 Forecast evaluation

In this section, the GloFAS-Seasonal reforecasts are evaluated using historical river flow observations. Benchmarking a forecasting system is important to evaluate and understand the value of the system and in order to communicate the skill of the forecasts to end users (Pappenberger et al., 2015). This evaluation is designed to measure the ability of the forecasts to predict the correct category of an "event", i.e. the ability of the forecast to predict that weekly averaged river flow will fall in the upper 80th or lower 20th percentile of climatology using a climatology of historical observations as a benchmark. This can be referred to as the potential usefulness of the forecasts and is of particular importance for decision-making purposes (Arnal et al., 2018). Another key aspect of probabilistic forecasts to consider is their reliability, which indicates the agreement between forecast probabilities and the observed frequency of events.
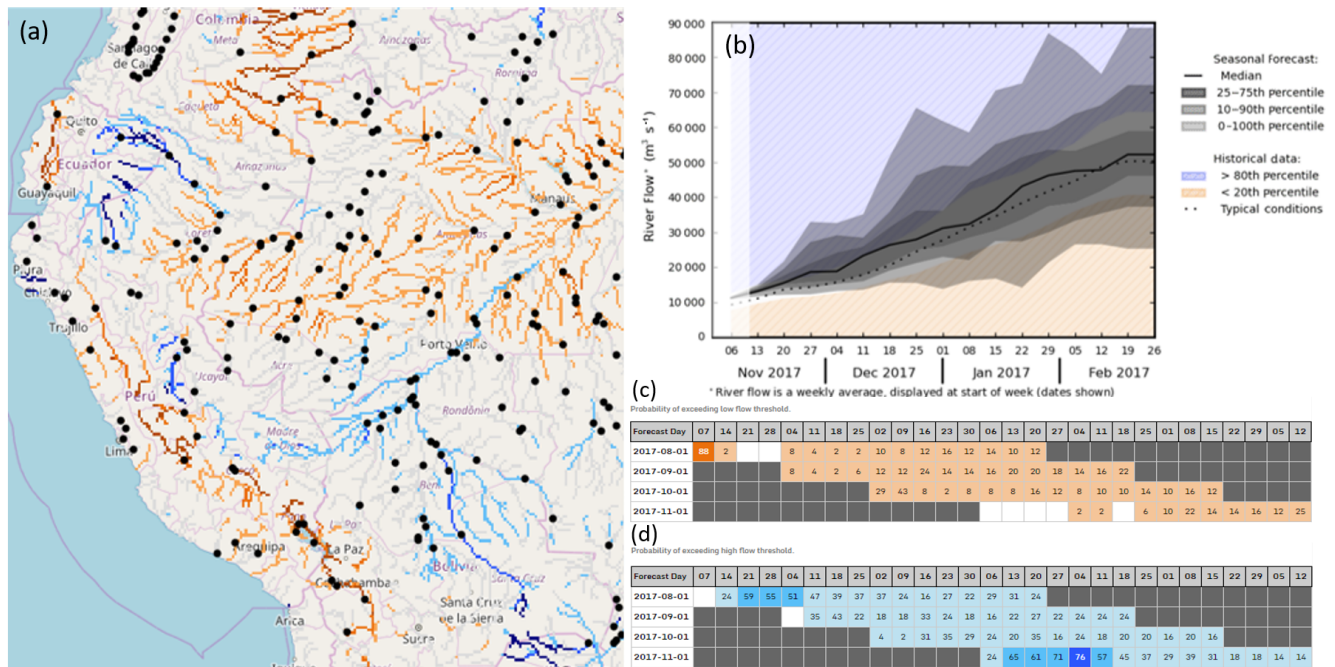
Figure 5. Example of the "reporting points" GloFAS seasonal outlook layer in the web interface **(a)**. Black circles indicate the reporting points, which provide the ensemble hydrograph **(b)** and persistence diagrams for both low flow **(c)** and high flow **(d)**. Also shown is an example section of the "river network" seasonal outlook layer indicating the maximum probability of high (blue) or low (orange) river flow during the 4-month forecast horizon. The darker the colour, the higher the probability.

The potential usefulness is assessed using the relative operating characteristic (ROC) curve, which is based on ratios of the proportion of events (the probability of detection, POD) and non-events (the false alarm rate, FAR) for which warnings were provided (Mason and Graham, 1999); in this case warnings are treated as forecasts of river flow exceeding the 80th or falling below the 20th percentile of the reference climatology (see Sect. 2.2.4). These ratios allow for the estimation of the probability that an event will be predicted.

For each week of the forecast (out to 16 weeks, corresponding to the forecasts provided via the interface; for example, the hydrograph shown in Fig. 5), the POD (Eq. 1) and FAR (Eq. 2) are calculated for both the 80th and 20th percentile events at each observation station:

$$POD = \frac{hits}{hits + misses}, \tag{1}$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms}, \tag{2}$$

where a hit is defined when the forecast correctly exceeded (fell below) the 80th (20th) percentile of the reference climatology during the same week that the observed river flow exceeded (fell below) the 80th (20th) percentile of the observations at that station. It follows that a miss is defined when an event was observed but the forecast did not exceed the threshold, and a false alarm when the forecast exceeded the threshold but no event was observed. From these, the area un-

der the ROC curve (AROC) is calculated, again for both the 80th and 20th percentile events. The AROC ($0 \leq AROC \leq 1$, where 1 is perfect) indicates the skill of the forecasts compared to the long-term average climatology (which has an AROC of 0.5) and is used here to evaluate the potential usefulness of the forecasts. The maximum lead time at which forecasts are more skilful than climatology (AROC > 0.5) is identified; a forecast with an AROC < 0.5 would be less skilful than climatology and thus not useful.

The reliability of the forecasts is assessed using attributes diagrams, which show the relationship between the forecast probability and the observed frequency of the events. While the ROC measures the ability of a forecasting system to predict the correct category of an event, the reliability assesses how closely the forecast probabilities correspond to the actual chance of observing the event. As such, these evaluation metrics are useful to consider together. As with the ROC calculations, the reliability is assessed for each week of the forecast (out to 16 weeks) and for both the 80th and 20th percentile events. The range of forecast probabilities is divided into 10 bins (0 %–10 %, 10 %–20 %, etc.), and the forecast probability is plotted against the frequency at which an event was observed for forecasts in each probability bin. Perfect reliability is exhibited when the forecast probability and the observed frequency are equal; for example, if a forecast predicts that an event will occur with a probability of 60 %, then the event should occur on 60 % of the occasions that this fore-

cast was made. Attributes diagrams can also be used to assess the sharpness and resolution of the forecasts. Forecasts that do not discriminate between events and non-events are said to have no resolution (a forecast of climatology would have no resolution), and forecasts which are capable of predicting events with probabilities that differ from the observed frequency, such as forecasts of high or 0 probability, are said to have sharpness.

The GloFAS-Seasonal reforecasts (of which there are 216 covering 18 years, as described in Sect. 2.2.4 and Fig. 2) are compared to river flow observations that have been made available to GloFAS, covering 17 years of the study period up to the end of 2015 when the data were collated (see Fig. 2). To ensure a large enough sample size for this analysis, alongside the best possible spatial coverage, the following criteria are applied to the data.

- The weekly river flow data record available for each station must contain no more than 53 % (9 years) missing data. The high and low flow thresholds (the 80th and 20th percentile, respectively) are calculated using the observations for each station and for each week across the 17 years of data, so a sample size of 17 is the maximum possible. A threshold of (up to) 53 % missing data allows for a minimum sample size of eight. Selecting a smaller threshold reduced the number of stations and the spatial coverage across the globe significantly. The percentage of missing data is calculated at each station and for each week of the dataset independently, and as such the number of stations used can vary slightly with time.

- The upstream area of the corresponding grid point in the model river network must be at least 1500 km^2.

These criteria allow for the use of 1140±14 stations globally. While the dataset contains 6122 stations, just 1664 of these contain data during the 17-year period, and none have the full 17 years of data available. Data from human-influenced rivers have not been removed, as in this study we are interested in identifying the ability of the forecasting system in its current state to predict observed events rather than the ability of the hydrological model to represent natural flow.

4.1 Potential usefulness

In order to gain an overview of the potential usefulness of the GloFAS-Seasonal forecasts across the globe, we map the maximum lead time at which the forecasts are more skilful than climatology (i.e. AROC > 0.5) at each observation station averaged across all forecast months. These results are shown in Fig. 6, and it is clear that forecasts of both high and low flow events are more skilful than climatology across much of the globe, with potentially useful forecasts at many stations out to 4 months ahead. However, there are regions where the forecasts are (on average across all fore-

cast months) not useful (i.e. AROC < 0.5), such as the western USA and Canada (excluding coastlines), much of Africa, and additionally across parts of Europe for low flow events. As forecasts with an AROC larger than but close to 0.5 could be deemed as only marginally more skilful than climatology, we apply a skill buffer, setting the threshold to AROC > 0.6 for a forecast to be deemed as potentially useful. These results are mapped in Fig. 7 and clearly indicate the reduction in the lead time at which forecasts are potentially useful (for both high and low flow events) at many stations, implying that in some locations, forecasts beyond the first 1–2 months are only marginally more skilful than climatology. There are, however, stations in some rivers with an AROC > 0.6 out to 4 months of lead time and many locations across the globe that still indicate that forecasts are potentially useful 1–2 months ahead for both high and low flow events.

These results can be further broken down by season, indicating whether the forecasts are more potentially useful at certain times of the year. Maps showing the maximum lead time at which AROC > 0.6 for each season (for forecasts started during the season; e.g. DJF indicates the average results for forecasts produced on 1 December, 1 January, and 1 February) are provided for high and low flow events in Figs. S1 and S2 in the Supplement, respectively.

The following paragraphs provide an overview of these results for each continent; for further detail please refer to the maps.

South America. For high flow events, forecasts for the Amazon basin in DJF and MAM are potentially useful out to longer lead times (up to 3–4 months) and at more stations than in JJA and SON, with similar results in MAM for low flow events. In contrast, further south, forecasts are most potentially useful JJA and SON up to 4 months ahead. In the more mountainous regions of western South America, forecasts in JJA and SON are generally less skilful than climatology for high and low flow events. In the north-west, however, for some stations, forecasts started in DJF and MAM are potentially useful up to 3 months ahead.

North America. In eastern North America, JJA and SON forecasts are most potentially useful, with more stations indicating an AROC > 0.6 out to 2–3 months ahead. However, during all seasons there are several stations in the east showing skill out to varying lead times. Much of the western half of the continent (excluding coastal areas) sees forecasts that are less skilful than climatology during all seasons, although some stations do indicate skill up to 4 months ahead for high flow, for forecasts started in MAM and JJA, and for low flow in MAM. At many coastal stations in the west, forecasts of high flow events started in DJF, MAM, and JJA indicate skill out to 3–4 months and out to ∼ 6 weeks in SON.

Europe. Forecasts for European rivers generally perform best for high flow events in SON and DJF, with the exception of some larger rivers in eastern Europe, for which the forecasts are more potentially useful in JJA and SON. In MAM and JJA, the number of stations indicating no skill is gener-
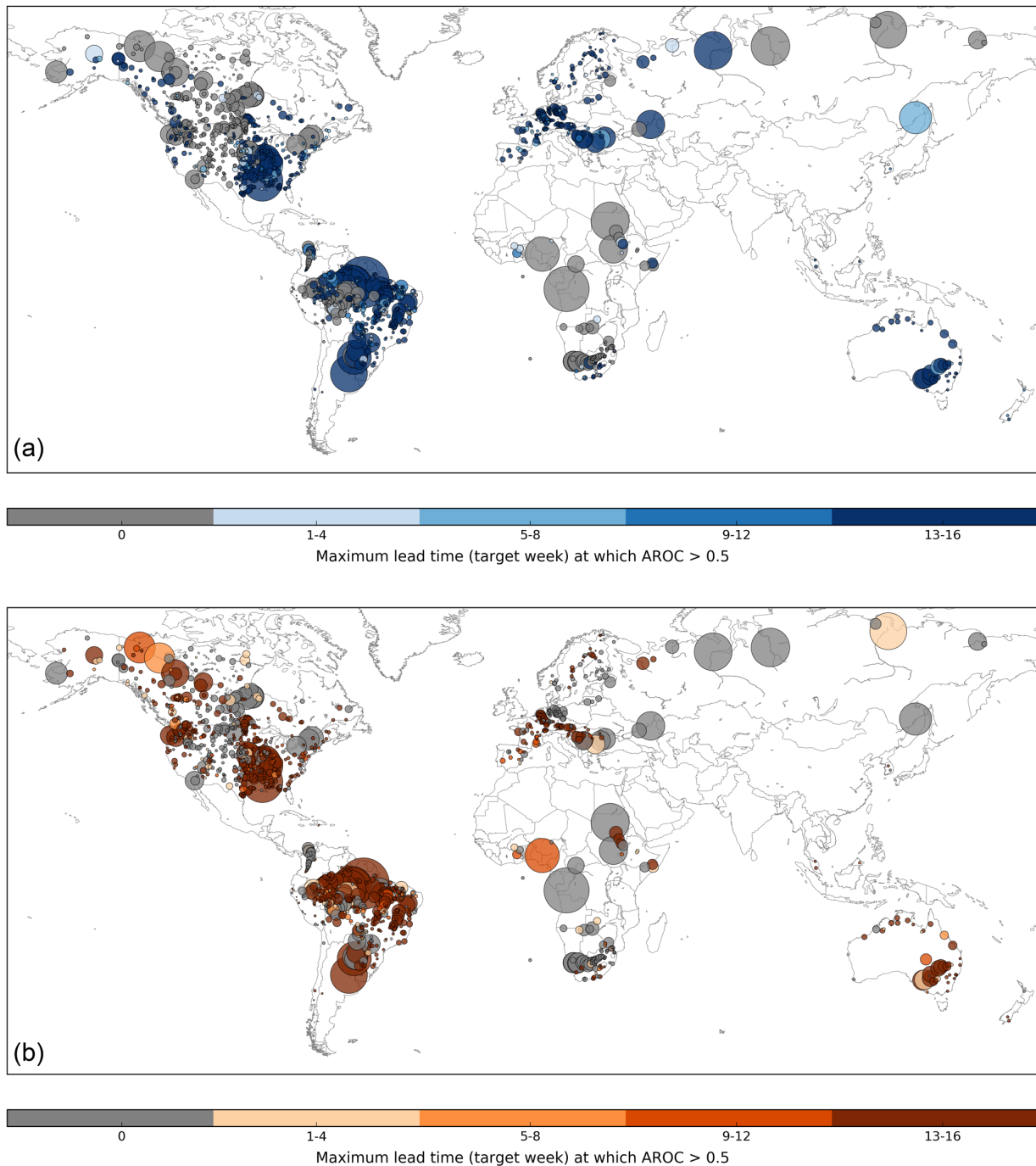
Figure 6. Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.5 **(a)** for high flow events (flow exceeding the 80th percentile of climatology) and **(b)** low flow events (flow below the 20th percentile of climatology) at each observation station. This is used to indicate the maximum lead time at which forecasts are more skilful than the long-term average. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skilful than climatology at all lead times.

ally higher. In contrast, forecasts for low flow events are less skilful than climatology across much of Europe. Particularly in north-east Europe and Scandinavia, forecasts produced in the summer months of JJA have an AROC < 0.6 at all sta-

tions, with only a few stations indicating any skill in other seasons, whereas in central and south-east Europe forecasts of low flow events are most skilful in JJA and SON out to 3–4 months ahead in the larger rivers. These results are similar
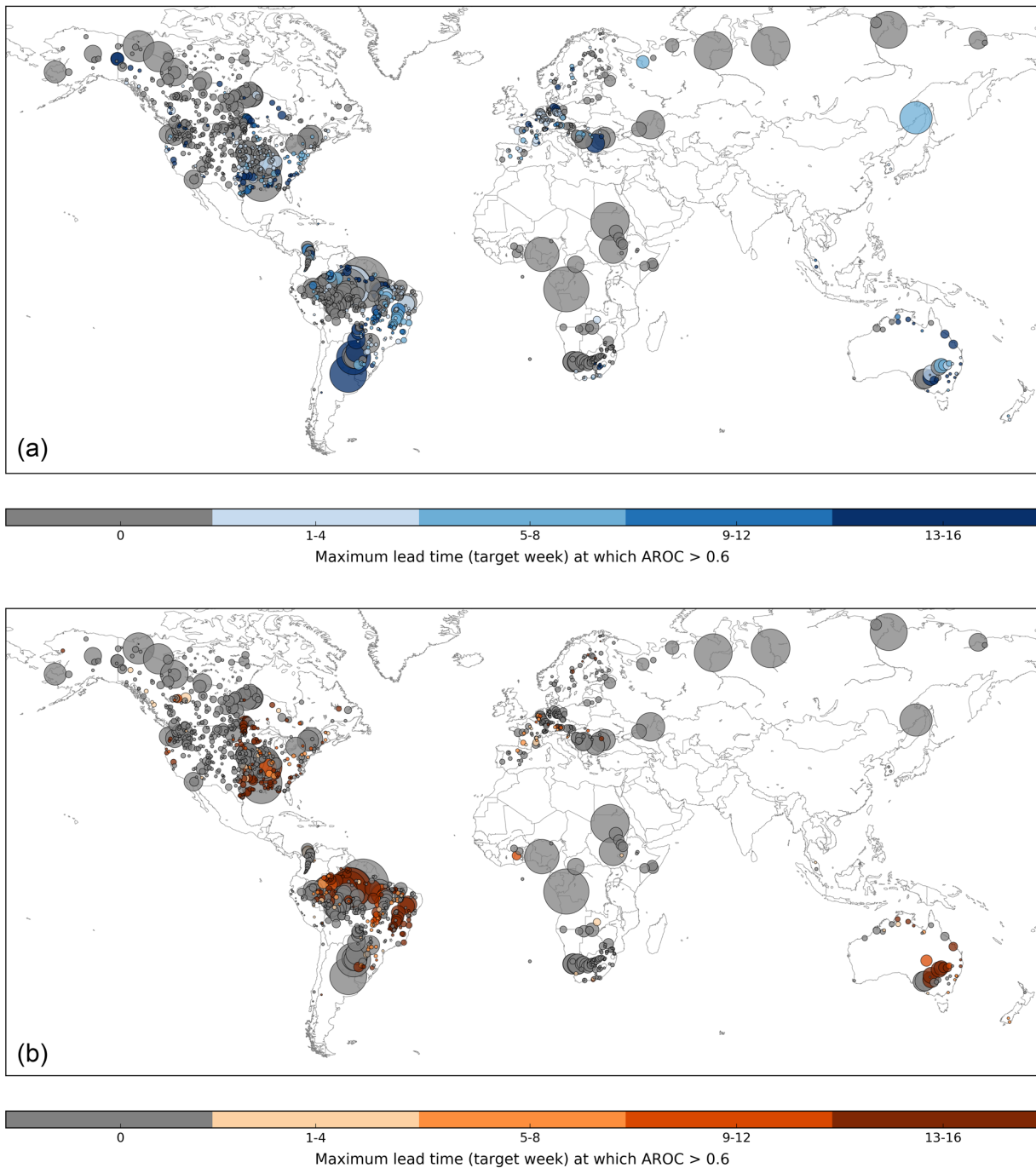
Figure 7. Maximum forecast lead time (target week, averaged across all months) at which the area under the ROC curve (AROC) is greater than 0.6 for **(a)** high flow events (flow exceeding the 80th percentile of climatology) and **(b)** low flow events (flow below the 20th percentile of climatology) at each observation station. This is used to indicate the maximum lead time at which forecasts are deemed skilful. Dot size corresponds to the upstream area of the location – thus larger dots represent larger rivers and vice versa. Grey dots indicate that (on average, across all months) forecasts are less skilful than climatology at all lead times. Maps for each season are provided in the Supplement.

to those of Arnal et al. (2018) for the potential usefulness of the EFAS seasonal outlook.

Asia. Although the number of available stations is very limited, the few stations available in South East Asia indicate

that the forecasts are potentially useful out to 3–4 months ahead, particularly for forecasts started in DJF and MAM preceding the start of the wet season. For low flow events, this skill extends into JJA, whereas forecasts made in SON

towards the end of the wet season tend to be less skilful than climatology.

Australia and New Zealand. Forecasts are most skilful out to longer lead times in the Murray–Darling river basin in the south-east, in particular for forecasts started in JJA and SON during the Southern Hemisphere winter and spring. In northern Australia, forecasts started in DJF and MAM for high flow events and MAM and JJA for low flow events are potentially useful out to 3–4 months ahead. This corresponds with the assessment of the skill of the Bayesian joint probability modelling approach for sub-seasonal to seasonal streamflow forecasting in Australia by Zhao et al. (2016), who found that forecasts in northern Australian catchments tend to be more skilful for the dry season (May to October) than the wet season (December to March). At the three stations in New Zealand, forecasts are only skilful for high flow events during the first month of lead time in DJF and MAM; however, for low flow events forecasts made in SON for the southern stations are potentially useful out to 4 months ahead.

Africa. While the spatial distribution of stations is limited, for high flow events forecasts are seen to be potentially useful at some of the stations in eastern Africa, particularly in SON and to a lesser extent in DJF. In southern Africa, there is skill in DJF and MAM, although the maximum lead time varies significantly from station to station. For low flow, there is little variation between the seasons; forecasts are generally less skilful than climatology across the continent, with some stations in DJF in southern and western Africa indicating skill in the first 1–2 months only.

4.2 Reliability

To provide an overall picture of the reliability of the GloFAS-Seasonal forecasts, attributes diagrams are produced for forecasts aggregated across all observation stations globally for both the 80th and 20th percentile events. In order to assess geographical differences in forecast reliability, attributes diagrams are also produced for forecasts aggregated across the stations within each of the major river basins used in the GloFAS-Seasonal forecast products (see Sect. 3.1). Many of these river basins do not contain a large enough number of stations to produce useful attributes diagrams, and as such the results in this section are presented for one river basin per continent for this initial evaluation. The river basin chosen for each continent is that which contains the largest number of observation stations.

The globally aggregated results (Fig. 8) indicate that, in general, the forecasts have more reliability than a forecast of climatology, though the reliability is less than perfect. It is important to note that the globally aggregated results shown in Fig. 8 mask any variability between river basins. Overall, the reliability appears to be slightly better for forecasts of high flow events than low flow events, and for lower probabilities, indicated by the steeper positive slope showing that as the forecast probability increases, so does the verified

chance of the event. The forecasts for both high and low flow events exhibit sharpness, although more so for high flow events, meaning that they have the ability to forecast probabilities that differ from the climatological average. This is indicated by the histograms inset within the attributes diagrams in Fig. 8; a forecast with sharpness will show a range of forecast probabilities differing from the climatological average (20 %), and a forecast with perfect sharpness will show peaks in the forecast frequency at 0 % and 100 %. Forecasts with no or low sharpness will show a peak in the forecast frequency near the climatological average. A forecast can have sharpness but still be unreliable. Figure 8 also suggests that in general, GloFAS-Seasonal forecasts have a tendency to over-predict the likelihood of an event occurring.

The following paragraphs summarise the forecast reliability for one river basin per continent; for a map of the location of these river basins, please refer to Fig. S3. The attributes diagrams for these river basins for both the 80th and 20th percentile events and for each season are provided in Figs. S4–S8. Each attributes diagram displays the results for forecast weeks 4, 8, 12, and 16, representing the reliability out to 1, 2, 3, and 4 months ahead. There are no river basins in Asia containing enough stations to produce an attributes diagram.

South America, Tocantins River (Fig. S4). For high flow events, forecasts for the Tocantins River indicate good reliability in all seasons, particularly up to 50 % probability. Forecasts in the higher-probability bins tend to over-predict, and this over-prediction worsens with lead time. In MAM and JJA, the forecasts tend to slightly under-predict in the lower-probability bins. The forecasts have sharpness, but it is clear that the sample size of high-probability forecasts is limited. There is a tendency to over-predict the likelihood of low flow events in all seasons, but the forecasts show good reliability for the lower-probability bins, particularly in SON and DJF. In JJA, the resolution of the forecasts is low.

North America, Lower Mississippi River (Fig. S5). For high flow events, the sample size of high-probability forecasts is small, and as such it is difficult to evaluate the reliability of these forecasts. The forecasts at lower probabilities have good reliability, particularly out to 2 months ahead in MAM and JJA. In SON and DJF, forecasts are more reliable at longer lead times. There is a tendency to under-predict at low probabilities and over-predict at high probabilities. For low flow events, the forecasts have a tendency to over-predict in all seasons, and the resolution of the forecasts is lower than for high flow events. At higher probabilities, forecasts of low flow events are more reliable than climatology, but the resolution is particularly low for probabilities up to 50–60 %. The forecasts for both high and low flow events have sharpness.

Europe, River Rhône (Fig. S6). For the River Rhône, the reliability is better than climatology at all lead times for high flow events, although there is a lack of forecasts of higher probabilities, particularly in MAM and JJA, as may be expected in the summer months. In SON, the reliability of forecasts up to 60–70 % is good at all lead times, and in DJF the
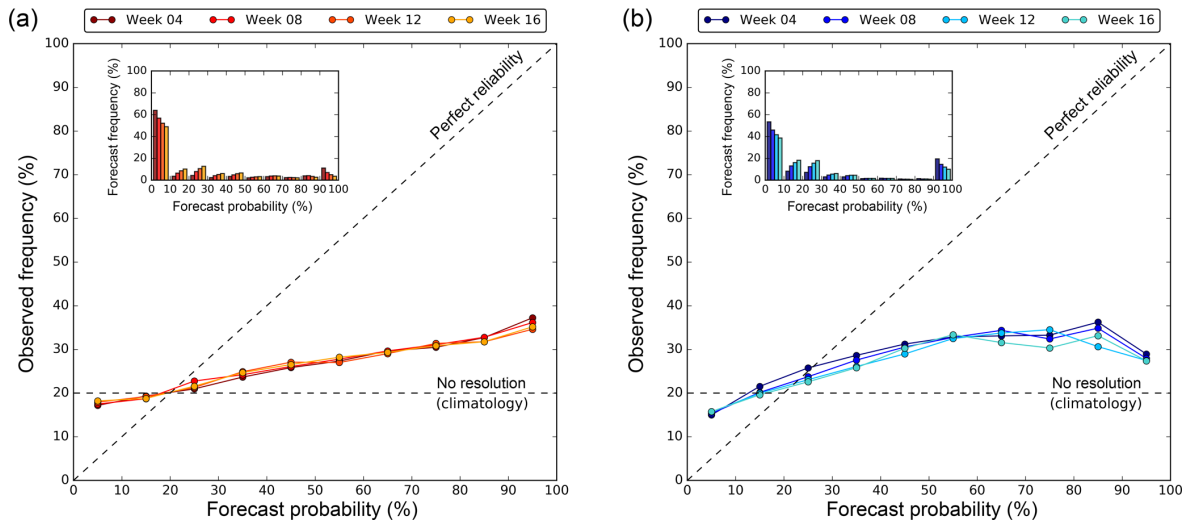
Figure 8. Attributes diagram for forecasts of **(a)** low flow events (flow below the 20th percentile of climatology) and **(b)** high flow events (flow exceeding the 80th percentile of climatology) aggregated across all observation stations globally. Results are shown for lead time weeks 4, 8, 12, and 16 and indicate the reliability of the forecasts. The histograms (inset) show the frequency at which forecasts occur in each probability bin and are used to indicate forecast sharpness. Attributes diagrams for selected river basins are provided in the Supplement.

forecasts are more reliable in the first 2 months of lead time for most probability bins. The reliability is less good for low flow events, but is generally better than climatology, particularly in summer (JJA). In winter (DJF), the resolution and reliability of the forecasts is poor. For all seasons and lead times and for both events, the forecasts have sharpness.

Australia, Murray River (Fig. S7). The attributes diagrams for both high and low flow events indicate that forecasts are often over-confident in this river basin, with probabilities of 0 %–10 % for low flow events and 0 %–30 % and 90 %– 100 % for high flow events, occurring frequently. As such, the sample size of forecasts in several of the bins is low. For high flow events, forecasts tend to over-predict at high probabilities and under-predict at low probabilities. The reliability is very good up to ∼ 30 %, after which the sample size is too small. For low flow events, there is a tendency to under-predict, but based on the forecasts available, the reliability is better than climatology at all lead times. The reliability for low flow events is better in SON and DJF (spring and summer) than MAM and JJA (autumn and winter), and for high flow events there is less differentiation between the seasons.

Africa, Orange River (Fig. S8). For the Orange River, forecasts of high flow events exhibit good reliability for lower probabilities in SON, DJF, and MAM (spring through autumn), particularly at longer lead times in SON and DJF, with a tendency to over-predict at higher probabilities. Resolution and reliability are poor for high flow events in JJA (winter), with probabilities of 90 %–100 % predicted too frequently. For low flow events, forecasts of 0 %–10 % are very frequent, and the forecasts under-predict in all seasons, although the reliability is better than climatology at all lead times (based

on a limited sample of forecasts for most probability bins). Reliability for low flow events is best in DJF (summer).

4.3 Discussion

The results presented provide an initial evaluation of the potential usefulness and reliability of GloFAS-Seasonal forecasts. For decision-making purposes, it is important to measure the ability of a forecasting system to predict the correct category of an event. As such, an event-based evaluation of the forecasts is used to assess whether the forecasts were able to correctly predict observed high and low river flow events over a 17-year period and whether it is able to do so with good reliability. The initial results are promising, indicating that the forecasts are, on average, potentially useful up to 1–2 months ahead in many rivers worldwide and up to 3–4 months ahead in some locations. The GloFAS-Seasonal forecasts have sharpness, i.e. they are able to predict forecasts with probabilities that differ from climatology, and overall have better reliability than a forecast of climatology, but with a tendency to over-predict at higher probabilities. It is also clear that there is a frequency bias in the reliability results, as often there is a small sample of high-probability forecasts. Typically, the reliability is seen to be better when there is a higher forecast frequency on which to base the results. As would be expected, the potential usefulness and reliability of the forecasts vary by region, season, and forecast lead time.

Considering the evaluation results by season allows for further analysis of the times of year in which the forecasts are potentially useful and/or reliable. For example, in southeast Australia, forecasts are seen to be potentially useful up to 4 months ahead in JJA and SON, but for forecasts produced

in DJF the skill only extends to 1 month ahead, and forecasts are less skilful than climatology at several of the stations in MAM. In many rivers across the globe, it is the case that forecasts are potentially useful in some seasons, but not in others, and may be more reliable in certain seasons than others. As such, the maps provided in Figs. S1 and S2 are intended to highlight where and when the forecasts are likely to be useful, information that is key in terms of decision-making.

It is clear that there are regions and seasons in which the forecasts are less skilful than climatology and do not have good reliability, and thus in these rivers it would be more useful to use a long-term average climatology than seasonal hydro-meteorological forecasts of river flow. This lack of skill could be due to several factors, such as certain hydrological regimes that may not be well-represented in the hydrological model or may be difficult to forecast at these lead times (for example, snow-dominated catchments or regions where convective storms produce most of the rainfall in some seasons), poor skill of the meteorological forecast input, poor initial conditions from the ERA5-R reanalysis, extensive management of rivers that cannot be represented by the current model, or the lack of model calibration. While this initial evaluation is designed to provide an overview of whether the forecasts are potentially useful and reliable in predicting high and low flow events, more extensive analysis is required to diagnose the sources of predictability in the forecasts and the potential causes of poor skill. Additionally, it is evident that observations of river flow, particularly covering the reforecast period, are both spatially and temporally limited across large areas of the globe. A more extensive analysis should make use of the globally consistent ERA5-R river flow reanalysis as a benchmark in order to fully assess the forecast skill worldwide, including in regions where no observations are available.

The verification metrics used also require that a high or low flow event is predicted with the correct timing in the same week as that in which it occurred. This is asking a lot of a seasonal forecasting system and for many applications, such as water resources and reservoir management, a forecast of the exact week in which an event is expected at a lead time of several months ahead may not be necessary. That such a system shows real skill despite this being a tough test for the model and is able to successfully predict observed high or low river flow in a specific week, several weeks or months ahead, provides optimism for the future of global-scale seasonal hydro-meteorological forecasting. Further evaluation should aim to assess the skill of the forecasts with a more relaxed constraint on the event timing and also make use of alternative skill measures to cover different aspects of the forecast skill, such as the spread and bias of the forecasts. It will also be important to assess whether the use of weekly averaged river flow is the most appropriate way to display the forecasts. While this is commonly used for applications such as drought early awareness and water resources management, there may be other aspects of decision-making,

such as flood forecasting, for which other measures may be more appropriate, for example daily averages or floodiness (Stephens et al., 2015).

Future development of GloFAS-Seasonal will aim to address these evaluation results and improve the skill and reliability of the current forecasts; it will also aim to overcome some of the grand challenges in operational hydrological forecasting, such as seamless forecasting and the use of data assimilation. Seamless forecasting will be key in the future development of GloFAS; the use of two different meteorological forecast inputs for the medium-range and seasonal versions of the model means that discrepancies can occur between the two timescales, thus producing confusing and inconsistent forecast information for users. Additionally, the use of river flow observations could lead to significant improvements in skill through calibration of the model using historical observations and assimilation of real-time data to adjust the forecasts. This remains a grand challenge due to the lack of openly available river flow data, particularly in real time.

5 Conclusions

In this paper, the development and implementation of a global-scale operational seasonal hydro-meteorological forecasting system, GloFAS-Seasonal, was presented, and an event-based forecast evaluation was carried out using two different but complementary verification metrics to assess the capability of the forecasts to predict high and low river flow events.

GloFAS-Seasonal provides forecasts of high or low river flow out to 4 months ahead for the global river network through three new forecast product layers via the openly available GloFAS web interface at http://www.globalfloods.eu (last access: 16 August 2018). Initial evaluation results are promising, indicating that in many rivers, forecasts are both potentially useful, i.e. more skilful than a long-term average climatology out to several months ahead in some cases, and overall more reliable than a forecast of climatology. Forecast skill and reliability vary significantly by region and by season.

The initial evaluation, however, also indicates a tendency of the forecasts to over-predict in general, and in some regions forecasts are currently less skilful than climatology; future development of the system will aim to improve the forecast skill and reliability with a view to providing potentially useful forecasts across the globe. Development of GloFAS-Seasonal will continue based on results of the forecast evaluation and on feedback from GloFAS partners and users worldwide in order to provide a forecast product that remains state of the art in hydro-meteorological forecasting and caters to the needs of its users. Future versions are likely to address some of the grand challenges in hydro-meteorological forecasting in order to improve forecast skill, such as data assim-

ilation, and will also include more features, such as flexible percentile thresholds and indication of the forecast skill via the interface. A further grand challenge that is important in terms of global-scale hydro-meteorological forecasting, and indeed for the development of GloFAS, is the need for more observed data (Emerton et al., 2016), which is essential not only for providing initial conditions to force the models, but also for evaluation of the forecasts and continuous improvement of forecast accuracy.

While such a forecasting system requires extensive computing resources, the potential for use in decision-making across a range of water-related sectors, and the promising results of the initial evaluation, suggest that it is a worthwhile use of time and resources to develop such global-scale systems. Recent papers have highlighted the fact that seasonal forecasts of precipitation are not necessarily a good indicator of potential floodiness and called for investment in better forecasts of seasonal flood risk (Coughlan De Perez et al., 2017; Stephens et al., 2015). Coughlan de Perez et al. (2017) state that "ultimately, the most informative forecasts of flood hazard at the seasonal scale could be seasonal streamflow forecasts using hydrological models" and that better seasonal forecasts of flood risk could be hugely beneficial for disaster preparedness.

GloFAS-Seasonal represents a first attempt at overcoming the challenges of producing and providing openly available seasonal hydro-meteorological forecast products, which are key for organisations working at the global scale and for regions where no other forecasting system exists. We provide, for the first time, seasonal forecasts of hydrological variables for the global river network by driving a hydrological model with seasonal meteorological forecasts. GloFAS-Seasonal forecasts could be used in addition to other forecast products, such as seasonal rainfall forecasts and short-range forecasts from national hydro-meteorological centres across the globe, to provide useful added information for many water-related applications from water resources management and agriculture to disaster risk reduction.

Code availability. The ECMWF IFS source code is available subject to a licence agreement, and as such access is available to the ECMWF member-state weather services and other approved partners. The IFS code is also available for educational and academic purposes as part of the OpenIFS project (ECMWF, 2011, 2018a), with full forecast capabilities and including the HTESSEL land surface scheme, but without modules for data assimilation. Similarly, the GloFAS river routing component source code is not openly available; however, the "forecast product" code (prior to implementation in ecFlow) that was newly developed for GloFAS-Seasonal and used for a number of tasks such as computing exceedance probabilities and producing the graphics for the interface is provided in the Supplement.

The Supplement related to this article is available online at https://doi.org/10.5194/gmd-11-3327-2018-supplement.

Edited by: Jeffrey Neal
Reviewed by: two anonymous referees

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013.

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, Hydrol. Earth Syst. Sci., 22, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018, 2018.

Bahra, A.: Managing work flows with ecFlow, ECMWF Newsl., 129, 30–32 available from: https://www.ecmwf.int/sites/default/files/elibrary/2011/14594-newsletter-no129-autumn-2011.pdf (last access: 18 April 2018), 2011.

Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P., and van den Hurk, B.: A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, Hydrol. Process., 25, 1046–1054, https://doi.org/10.1002/hyp.7808, 2011.

BDHI: Base de Donnees Historiques sur les Inondations, available at: http://bdhi.fr/appli/web/welcome, last access: 23 April 2018.

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, Hydrol. Earth Syst. Sci., 21, 4681–4691, https://doi.org/10.5194/hess-21-4681-2017, 2017.

Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E., and Schepen, A.: Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, Water Resour. Res., 52, 8238–8259, https://doi.org/10.1002/2016WR019193, 2016.

Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., and Michael, K.: Assessment of an ensemble seasonal streamflow forecasting system for Australia, Hydrol. Earth Syst. Sci., 21, 6007–6030, https://doi.org/10.5194/hess-21-6007-2017, 2017.

BfG: The GRDC, available at: http://www.bafg.de/GRDC/EN/Home/homepage_node.html (last accessed: 23 April 2018), 2017.

BoM: Seasonal Streamflow Forecasts: Water Information: Bureau of Meteorology, available at: http://www.bom.gov.au/water/ssf/about.shtml, last access: 24 April 2018.

Candogan Yossef, N., van Beek, R., Weerts, A., Winsemius, H., and Bierkens, M. F. P.: Skill of a global forecasting system in seasonal ensemble streamflow prediction, Hydrol. Earth Syst. Sci., 21, 4103–4114, https://doi.org/10.5194/hess-21-4103-2017, 2017.

Cemaden: Cemaden – Centro Nacional de Monitoramento e Alertas de Desastres Naturais, available at: http://www.cemaden.gov.br/ (last access: 23 April 2018), 2017.

Chiew, F. H. S. and McMahon, T. A.: Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, Hydrol. Sci. J., 47, 505–522, https://doi.org/10.1080/02626660209492950, 2002.

Chow, V. Te, Maidment, D. R., and Mays, L. W.: Applied hydrology, Tata McGraw-Hill Education, available at: https://books.google.co.uk/books/about/Applied_Hydrology.html?id=RRwidSsBJrEC&redir_esc=y (last access: 17 November 2017), 2010.

Cloke, H., Pappenberger, F., Thielen, J., and Thiemig, V.: Operational European Flood Forecasting, in Environmental Modelling, John Wiley & Sons, Ltd, Chichester, UK, 415–434, 2013.

Copernicus: Copernicus Climate Data Store, available at: https://climate.copernicus.eu/climate-data-store, last access: 23 April 2018a.

Copernicus: SWICCA, Service for Water Indicators in Climate Change Adaptation, available at: http://swicca.climate.copernicus.eu/, last access: 12 January 2018b.

Coughlan de Perez, E., Stephens, E., Bischiniotis, K., van Aalst, M., van den Hurk, B., Mason, S., Nissan, H., and Pappenberger, F.: Should seasonal rainfall forecasts be used for flood preparedness?, Hydrol. Earth Syst. Sci., 21, 4517–4524, https://doi.org/10.5194/hess-21-4517-2017, 2017.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 20, 3601–3618, https://doi.org/10.5194/hess-20-3601-2016, 2016.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., Zhu, Y., Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y.: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, . Am. Meteorol. Soc., 95, 79–98, https://doi.org/10.1175/BAMS-D-12-00081.1, 2014.

DGA: Ministerio de Obras Públicas – Dirección de General de Aguas, available at: http://www.dga.cl/Paginas/default.aspx, last access: 23 April 2018.

DHM: Department of Hydrology and Meteorology, available at: http://www.dhm.gov.np/ (last access: 23 April 2018), 2017.

DWA: Department: WAter and Sanitation, available at: http://www.dwa.gov.za/default.aspx, last access: 23 April 2018.

ECMWF: OpenIFS, available at: https://www.ecmwf.int/en/research/projects/openifs (last access: 16 August 2018), 2011.

ECMWF: ecFlow Documentation, available at: https://software.ecmwf.int/wiki/display/ECFLOW/Documentation (last access: 18 April 2018), 2012.

ECMWF: SEAS5 user guide, available at: https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf (last access: 18 April 2018), 2017a.

ECMWF: What are the changes from ERA-Interim to ERA5? – Copernicus Knowledge Base – ECMWF Confluence Wiki, available at: https://software.ecmwf.int/wiki/pages/viewpage.action?pageId=74764925 (last access: 24 April 2018), 2017b.

ECMWF: About OpenIFS, available at: https://software.ecmwf.int/wiki/display/OIFS/About+OpenIFS, last access: 26 April 2018a.

ECMWF: ECMWF IFS Documentation CY43R1, available at: https://www.ecmwf.int/search/elibrary/IFS?secondary_title=IFSDocumentationCY43R1, last access: 18 April 2018b.

EFAS: European Flood Awareness System (EFAS), available at: https://www.efas.eu/ (last access: 23 April 2018), 2017.

Emerton, R., Cloke, H. L., Stephens, E. M., Zsoter, E., Wool-nough, S. J., and Pappenberger, F.: Complex picture for likelihood of ENSO-driven flood hazard, Nat. Commun., 8, 14796, https://doi.org/10.1038/ncomms14796, 2017.

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., and Cloke, H. L.: Continental and global scale flood forecasting systems, Wiley Interdisc. Rev. Water, 3, 391–418, https://doi.org/10.1002/wat2.1137, 2016.

Environment Canada: Water Level and Flow – Environment Canada, available at: https://wateroffice.ec.gc.ca/ (last access: 23 April 2018), 2014.

Environmental Systems Research Institute: ArcMap, ArcGIS Desktop, available at: http://desktop.arcgis.com/en/arcmap/, last access: 26 April 2018.

Fekete, B. M., Vörösmarty, C. J., and Lammers, R. B.: Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, Water Resour. Res., 37, 1955–1967, https://doi.org/10.1029/2001WR900024, 2001.

GloFAS: GloFAS Web Map Service Time (WMS-T) User Manual, available at: http://www.globalfloods.eu/static/downloads/GloFAS-WMS-T_usermanual.pdf last access: 26 April 2018a.

GloFAS: GloFAS WMS-T, available at: http://globalfloods-ows.ecmwf.int/glofas-ows/?service=WMS&request=GetCapabilities, last access: 16 August 2018b.

IDEAM: IDEAM, available at: http://www.ideam.gov.co/ (last access: 23 April 2018), 2014.

INDRHI: INDRHI – National Institute of Hydraulic Resources, available at: http://indrhi.gob.do/ (last access: 23 April 2018), 2017.

INETER: Ineter, Instituto Nicaragüense de Estudios Territoriales, available at: http://www.ineter.gob.ni/ (last access: 23 April 2018), 2016.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process., 27, 2171–2186, https://doi.org/10.1002/hyp.9740, 2013.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos, Trans. Am. Geophys. Union, 89, 93–94, https://doi.org/10.1029/2008EO100001, 2008.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Lorenz, E. N.: The essence of chaos, University of Washington Press, 1993.

Mason, S. J. and Graham, N. E.: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels, Weather Forecast., 14, 713–725, https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2, 1999.

McPhaden, M. J., Zebiak, S. E., and Glantz, M. H.: ENSO as an integrating concept in earth science., Science, 314, 1740–1745, https://doi.org/10.1126/science.1132588, 2006.

Meißner, D., Klein, B., and Ionita, M.: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, Hydrol. Earth Syst. Sci., 21, 6401–6423, https://doi.org/10.5194/hess-21-6401-2017, 2017.

Mo, K. C., Lettenmaier, D. P., Mo, K. C., and Lettenmaier, D. P.: Hydrologic Prediction over the Conterminous United States Us-ing the National Multi-Model Ensemble, J. Hydrometeorol., 15, 1457–1472, https://doi.org/10.1175/JHM-D-13-0197.1, 2014.

OGC: OGC Web Map Service v1.3.0, https://doi.org/10.3173/air.21.76, 2015.

OpenLayers: OpenLayers, available at: http://openlayers.org/, last access: 18 April 2018.

Open Source Geospatial Foundation: MapServer 7.0.1 documentation, available at: http://mapserver.org/uk/index.html (last access: 26 April 2018), 2016.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, J. Hydrol., 522, 697–713, https://doi.org/10.1016/J.JHYDROL.2015.01.024, 2015.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, Hydrolog. Sci. J., 62, 2753–2768, https://doi.org/10.1080/02626667.2017.1395032, 2017.

QGIS Development Team: Quantum GIS Geographical Information System, available at: https://www.qgis.org/, last access: 4 December 2017.

RCCC: Home – Red Cross Red Crescent Climate Centre, available at: http://www.climatecentre.org/, last access: 23 April 2018.

SHI: "State Hydrological Institute" (SHI), Russian Federal State Budgetary Organization, available at: http://www.hydrology.ru/en, last access: 23 April 2018.

Stephens, E., Day, J. J., Pappenberger, F., and Cloke, H.: Precipitation and floodiness, Geophys. Res. Lett., 42, 10316–10323, https://doi.org/10.1002/2015GL066779, 2015.

Stockdale, T., Johnson, S., Ferranti, L., Balmaseda, M., and Briceag, S.: ECMWF's new long-range forecasting system SEAS5, ECMWF Newsl., 154, 15–20, available at: http://www.ecmwf.int/en/about/news-centre/media-resources, last access: 18 April 2018.

SWALIM: FAO SWALIM: Somalia Water and Land Information Management, available at: http://www.faoswalim.org/, last access: 23 April 2018.

Van Der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci., 24, 189–212, https://doi.org/10.1080/13658810802549154, 2010.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D., and Kummu, M.: Annual flood sensitivities to El Niño–Southern Oscillation at the global scale, Hydrol. Earth Syst. Sci., 18, 47–66, https://doi.org/10.5194/hess-18-47-2014, 2014a.

Ward, P. J., Jongman, B., Kummu, M., Dettinger, M. D., Sperna Weiland, F. C., and Winsemius, H. C.: Strong influence of El Niño Southern Oscillation on flood risk around the world, P. Natl. Acad. Sci. USA, 111, 15659–15664, https://doi.org/10.1073/pnas.1409822111, 2014b.

Ward, P. J., Kummu, M., and Lall, U.: Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis, J. Hydrol., 539, 358–378, https://doi.org/10.1016/J.JHYDROL.2016.05.045, 2016.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, J. R. Soc. Interface, 11, 20131162, https://doi.org/10.1098/rsif.2013.1162, 2014.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys. Res., 107, 4429, https://doi.org/10.1029/2001JD000659, 2002.

Wood, A. W., Kumar, A., and Lettenmaier, D. P.: A retrospective assessment of National Centers for Environmental Prediction climate model–based ensemble hydrologic forecasting in the western United States, J. Geophys. Res., 110, D04105, https://doi.org/10.1029/2004JD004508, 2005.

Yuan, X., Wood, E. F., Chaney, N. W., Sheffield, J., Kam, J., Liang, M., and Guan, K.: Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, J. Hydrometeorol., 14, 1706–1720, https://doi.org/10.1175/JHM-D-13-054.1, 2013.

Yuan, X., Wood, E. F., and Ma, Z.: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, Wiley Interdiscip. Rev. Water, 2, 523–536, https://doi.org/10.1002/wat2.1088, 2015a.

Yuan, X., Roundy, J. K., Wood, E. F., Sheffield, J., Yuan, X., Roundy, J. K., Wood, E. F., and Sheffield, J.: Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins, B. Am. Meteorol. Soc., 96, 1895–1912, https://doi.org/10.1175/BAMS-D-14-00003.1, 2015b.

Zhao, T., Schepen, A., and Wang, Q. J.: Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, J. Hydrol., 541, 839–849, https://doi.org/10.1016/J.JHYDROL.2016.07.040, 2016.

A4: Building a multimodel flood prediction system with the TIGGE archive

This paper presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

Zsoter, E., F. Pappenberger, P. Smith, **R. E. Emerton,** E. Dutra, F. Wetterhall, D. Richardson, K. Bogner and G. Balsamo, 2016: Building a multimodel flood prediction system with the TIGGE archive, *Journal of Hydrometeorology*, **17**, 2923-2940, doi:10.1175/JHM-D-15-0130.1[*]

R.E. provided some technical support for the analysis presented in this paper, including matching the location of each river flow observation station to the corresponding location on the model river network, and commented on the manuscript before and during publication.

Building a Multimodel Flood Prediction System with the TIGGE Archive

ERVIN ZSÓTÉR,[a,d] FLORIAN PAPPENBERGER,[a,b] PAUL SMITH,[a,c] REBECCA ELIZABETH EMERTON,[a,d]
EMANUEL DUTRA,[a] FREDRIK WETTERHALL,[a] DAVID RICHARDSON,[a] KONRAD BOGNER,[a,e] AND
GIANPAOLO BALSAMO[a]

[a] *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*
[b] *College of Hydrology and Water Resources, Hohai University, Nanjing, China*
[c] *Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom*
[d] *Department of Geography and Environmental Science, University of Reading, Reading, United Kingdom*
[e] *Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

(Manuscript received 28 July 2015, in final form 23 July 2016)

ABSTRACT

In the last decade operational probabilistic ensemble flood forecasts have become common in supporting decision-making processes leading to risk reduction. Ensemble forecasts can assess uncertainty, but they are limited to the uncertainty in a specific modeling system. Many of the current operational flood prediction systems use a multimodel approach to better represent the uncertainty arising from insufficient model structure. This study presents a multimodel approach to building a global flood prediction system using multiple atmospheric reanalysis datasets for river initial conditions and multiple TIGGE forcing inputs to the ECMWF land surface model. A sensitivity study is carried out to clarify the effect of using archive ensemble meteorological predictions and uncoupled land surface models. The probabilistic discharge forecasts derived from the different atmospheric models are compared with those from the multimodel combination. The potential for further improving forecast skill by bias correction and Bayesian model averaging is examined. The results show that the impact of the different TIGGE input variables in the HTESSEL/Catchment-Based Macroscale Floodplain model (CaMa-Flood) setup is rather limited other than for precipitation. This provides a sufficient basis for evaluation of the multimodel discharge predictions. The results also highlight that the three applied reanalysis datasets have different error characteristics that allow for large potential gains with a multimodel combination. It is shown that large improvements to the forecast performance for all models can be achieved through appropriate statistical postprocessing (bias and spread correction). A simple multimodel combination generally improves the forecasts, while a more advanced combination using Bayesian model averaging provides further benefits.

1. Introduction

Operational probabilistic ensemble flood forecasts have become more common in the last decade (Cloke and Pappenberger 2009; Demargne et al. 2014; Olsson and Lindström 2008). Ensemble forecasts are a good way of assessing forecast uncertainty, but they are limited to the uncertainty captured by a specific modeling system. A multimodel approach can address this shortcoming and provide a more complete representation of the uncertainty in the model structure, also potentially reducing the errors (Krishnamurti et al. 1999).

"Multimodel" can refer to systems using multiple meteorological models, hydrological models, or both (Velázquez et al. 2011). According to Emerton et al. (2016), among the many regional-scale operational hydrological ensemble prediction systems across the globe, at present there are six large-scale (continental and global) models: four that run at continental scale over Europe, Australia, and the United States and two that are available globally. The U.S. Hydrologic Ensemble Forecast Service (HEFS), run by the National Weather Service (NWS; Demargne et al. 2014), and the Global Flood Forecasting Information System (GLOFFIS), a

recent development at Deltares in the Netherlands, are examples of systems using different hydrological models as well as multiple meteorological inputs. The European Flood Awareness System (EFAS) developed by the Joint Research Centre (JRC) of the European Commission and ECMWF operates using a single hydrological model with multimodel meteorological input (Thielen et al. 2009). Finally, the European Hydrological Predictions for the Environment (E-HYPE) Water in Europe Today (WET) model of the Swedish Meteorological and Hydrological Institute (SMHI; Donnelly et al. 2015), the Australian Flood Forecasting and Warning Service, and the Global Flood Awareness System (GloFAS; Alfieri et al. 2013), running in collaboration between ECMWF and JRC, all use one main hydrological model and one meteorological model input.

While the multimodel approach has traditionally involved the use of multiple forcing inputs and hydrological models to generate discharge forecasts, it also allows for consideration of multiple initial conditions. In keeping with GloFAS, this paper uses atmospheric reanalysis data to generate the initial conditions of the land surface components of the forecasting system; therefore, a multimodel approach based on three reanalysis datasets is trialed.

The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010) archive is an invaluable source of multimodel meteorological forcing data. The archive has attracted attention among hydrological forecasters and is already being extensively used in hydrological applications. The first published example of a hydrometeorological forecasting application was by Pappenberger et al. (2008). In that paper, the forecasts of nine TIGGE centers were used within the setting of EFAS for a case study of a flood event in Romania in October 2007 and showed that the lead time of flood warnings could be improved by up to 4 days through the use of multiple forecasting models rather than a single model. This study and other subsequent studies using TIGGE multimodel data (e.g., He et al. 2009, 2010; Bao and Zhao 2012) have indicated that combining different models not only increases the skill, but also the lead time at which warnings could be issued. He et al. (2009) highlighted this and further showed that individual systems of the multimodel forecast have systematic errors in time and space that would require temporal and spatial postprocessing. Such postprocessing should carefully maintain spatial, temporal, and intervariable correlations; otherwise, they lead to deteriorating hydrological forecast skill.

The scientific literature contains numerous studies on methods that can lead to significant gain in forecast skill by combining and postprocessing different forecast systems. Statistical ensemble postprocessing techniques target the generation of sharp and reliable probabilistic forecasts from ensemble outputs. Hagedorn et al. (2012) showed, based on TIGGE, that by considering an equal-weight multimodel approach, a selection of best NWP models might be needed to gain skill on the best-performing single model. In addition to this, the calibration of the best single model using a reforecast dataset can lead to comparable or even superior quality to the multimodel prediction. Gneiting and Katzfuss (2014) focus on various methodologies that require weighting of the different contributing forecasts to optimize model error corrections. They recommend the application of well-established techniques in the operational environment such as the nonhomogeneous regression or Bayesian model averaging (BMA). The BMA method generates calibrated and sharp probability density functions (PDFs) from ensemble forecasts (Raftery et al. 2005), where the predictive PDF is a weighted average of the PDFs centered on the bias-corrected forecasts. The weights reflect the relative skill of the individual members over a training period. The BMA has been widely used and proved to be beneficial in hydrological ensemble systems (e.g., Ajami et al. 2007; Cane et al. 2013; Dong et al. 2013; Liang et al. 2013; Todini 2008; Vrugt and Robinson 2007).

Previous studies have used hydrological models, rather than land surface models, to analyze the benefits of multimodel forecasting and have focused on individual catchments. The potential of multimodel forecasts at the regional or continental scale shown in previous studies provides the motivation for building a global multimodel hydrometeorological forecasting system.

In this study we present our experiences in building a multimodel hydrometeorological forecasting system. Global ensemble discharge forecasts with a 10-day horizon are generated using the ECMWF land surface model and a river-routing model. The multimodel approach arises from the use of meteorological forecasts from four models in the TIGGE archive and the derivation of river initial conditions using three global reanalysis datasets. The main focus of our study is the quality of the discharge forecasts derived from the TIGGE data. We analyze the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESSEL)/Catchment-Based Macroscale Floodplain model (CaMa-Flood) setup and the scope for error reduction by applying the multimodel approach and different postprocessing methods on the forecast data. Three sets of experiments are undertaken to test

(i) the sensitivity of the forecasting system to the input variables, (ii) the potential improvements in forecasting historical discharge that can be achieved by a combination of different reanalysis datasets, and (iii) the use of bias correction and model combination to improve the predictive distribution of the forecasts.

In section 2 the datasets, models, and methodology used throughout the paper are described. Section 3 summarizes the discharge experiments we produced and analyzed. In section 4, we provide the results, while section 5 gives conclusions to the paper.

2. System description and datasets

a. HTESSEL land surface model

The hydrological component of this study was the HTESSEL (Balsamo et al. 2009, 2011) land surface model. The HTESSEL scheme follows a mosaic (or tiling) approach where the grid boxes are divided into patches (or tiles), with up to six fractions over land (bare ground, low and high vegetation, intercepted water, and shaded and exposed snow) and two extra tiles over water (open and frozen water) exchanging energy and water with the atmosphere. The model is part of the Integrated Forecast System (IFS) at ECMWF and is used in coupled atmosphere–surface mode on time ranges from medium range to seasonal forecasts. In addition, the model provides a research test bed for applications where the land surface model can run in a stand-alone mode. In this so-called "offline" version the model is forced with near-surface meteorological input (temperature, specific humidity, wind speed, and surface pressure), radiative fluxes (downward solar and thermal radiation), and water fluxes (liquid and solid precipitation). This offline methodology has been explored in various research applications where HTESSEL or other models were applied (e.g., Agustí-Panareda et al. 2010; Dutra et al. 2011; Haddeland et al. 2011).

b. CaMa-Flood river routing

CaMa-Flood (Yamazaki et al. 2011) was used to integrate HTESSEL runoff over the river network into discharge. CaMa-Flood is a distributed global river-routing model that routes runoff to oceans or inland seas using a river network map. A major advantage of CaMa-Flood is the explicit representation of water level and flooded area in addition to river discharge. The relationship between water storage (the only prognostic variable), water level, and flooded area is determined on the basis of the subgrid-scale topographic parameters based on a 1-km digital elevation model.

c. TIGGE forecasts

The atmospheric forcing for the forecast experiments is taken from the TIGGE archive where all variables are available on a standard 6-h forecast frequency. The ensemble systems of ECMWF, the Met Office (UKMO), the National Centers for Environmental Prediction (NCEP), and the China Meteorological Administration (CMA) provide, in the TIGGE archive, meteorological forcing fields from the 0000 UTC runs with 6-h frequency starting from 2006 to 2008 depending on the model. All four models were only available with the complete forcing variable set from August 2008. ECMWF was available with 50 ensemble members on 32-km horizontal resolution (~50 km before January 2010) up to 15 days ahead, UKMO was available with 23 members on ~60-km horizontal resolution (~90 km before March 2010) also up to 15 days ahead, NCEP was available with 20 members on ~110-km horizontal resolution up to 16 days ahead, and finally CMA was available with 14 members on ~60-km horizontal resolution up to 10 days ahead. In testing the sensitivity of the experimental setup to meteorological forcing (see section 4a) the ECMWF control forecasts were used, extracted directly from ECMWF's Meteorological Archival and Retrieval System (MARS), where the meteorological variables are available without the TIGGE restrictions. These have the same resolution as the 50 ensemble members but start from the unperturbed analysis.

d. Reanalysis data

The discharge modeling experiments require reanalysis data, which are used to provide the climate and the initial conditions needed for the HTESSEL land surface model runs and to produce the river initial conditions required in the CaMa-Flood routing part of the TIGGE forecast experiments.

In this study we have used three different reanalysis datasets: two produced by ECMWF, ERA-Interim (hereafter ERAI) and ERA-Interim/Land with Global Precipitation Climatology Project, version 2.2 (GPCP v2.2), precipitation (Huffman et al. 2009) correction (hereafter ERAI-Land; Balsamo et al. 2015), and a third, the Modern-Era Retrospective Analysis for Research and Applications (MERRA) land upgrade (MERRA-Land) produced by NOAA. The combination of these three sources was a proof of concept to potential added value of the multi-initial conditions.

ERAI is ECMWF's global atmospheric reanalysis from 1979 to present produced with an older (2006) version of the ECMWF IFS on a T255 spectral resolution (Dee et al. 2011). ERAI-Land is a version of ERAI
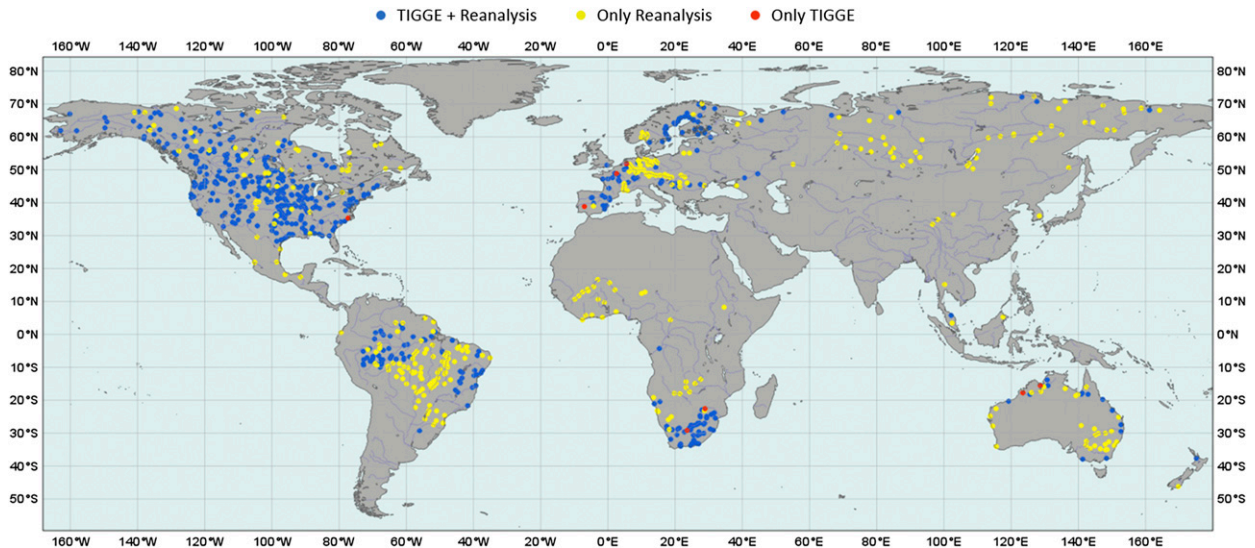
FIG. 1. Location of discharge observing stations that could be processed in the discharge experiments. The blue points are used in both the reanalysis (at least 15 years of data available in 1981–2010) and in the TIGGE forecast experiment (at least 80% of days available from August 2008 to May 2010) evaluation. The yellow points provide enough observation only for the reanalysis while the red points have enough data available only for the TIGGE forecasts.

at the same 80-km spatial resolution with improvements for land surface. It was produced in offline mode with a 2014 version of the HTESSEL land surface model using atmospheric forcing from ERAI, with precipitation adjustments based on GPCP v2.2, where the ERAI 3-hourly precipitation is rescaled to match the monthly accumulated precipitation provided by the GPCP v2.2 product [for more details, please consult Balsamo et al. (2010)].

The MERRA-Land dataset is similar to ERAI-Land in that it is a land-only version of the MERRA land model component, produced also in offline mode, using improved precipitation forcing and an improved version of the catchment land surface model (Reichle et al. 2011).

e. Discharge data

In this study a subset of the observations available in GloFAS was used, mainly originating from the Global Runoff Data Centre (GRDC) archive. The GRDC is the digital worldwide depository of discharge data and associated metadata. It is an international archive of data started in 1812, and it fosters multinational and global long-term hydrological studies.

For the discharge modeling, a dataset of 1121 stations with upstream areas over $10\,000\,km^2$ was available until the end of 2013. GRDC has a gradually decreasing number of stations with data in the archive limiting their use for more recent years. For the forecast discharge, we limited our analyses to the period from August 2008 to May 2010. This period provided the optimal compromise in increasing the sample size between the length of the

period and the number of stations with good data coverage. For the reanalysis discharge experiments and also for generating the observed discharge climate, stations with a minimum of 15 years of available observations were used in the 30-yr period from 1981 to 2010. For the forecast experiments, stations with at least 80% of the observations available were used in the 22-month period from August 2008 to May 2010. Figure 1 shows the observation availability in the reanalysis and TIGGE forecast experiments. It highlights that for the reanalysis the coverage is better globally, with about 850 stations, while the forecast experiments have around 550 stations with large missing areas, mainly in Africa and Asia.

f. Forecasting system setup

To produce runoff from the TIGGE atmospheric ensemble variables (see section 2c), HTESSEL experiments were run with 6-hourly forcing frequency and hourly model time step. For the instantaneous variables (such as 2-m temperature), linear interpolation was used to move from the 6-h to hourly time step used in the HTESSEL simulations. For accumulated variables (such as precipitation), a disaggregation algorithm that conserves the 6-hourly totals was used. The disaggregation algorithm divides into hourly values based on a linear combination of the current and adjacent 6-hourly totals with weights derived from the time differences.

The climate and the initial conditions needed for the HTESSEL land surface model runs to produce runoff were taken from ERAI-Land, the same initial conditions for all models and ensemble members without

perturbations. The other two reanalysis datasets could also be used to initialize HTESSEL, but the variability on the resulting TIGGE runoff (and thus on the TIGGE discharge) would be very small compared with the impact of the TIGGE atmospheric forcing (especially precipitation, see also section 4a) and the impact of the TIGGE forecast routing initialization (see section 3b for further details).

HTESSEL was set to T255 spectral resolution (~80 km). This was the horizontal resolution used in ERAI and was an adequate compromise between the highest (ECMWF mainly ~50 km) and lowest (NCEP with ~110 km) forcing model resolution that also allowed fast enough computations. The TIGGE forcing fields were transformed to T255 using bilinear interpolation.

The TIGGE archive includes variables at the surface and several pressure levels. However, variables are not available on model levels, and as such, temperature, wind, and humidity at the surface (i.e., 2 m for temperature and humidity and 10 m for wind) were used in HTESSEL rather than on the preferred lowest model level (LML).

Similarly, TIGGE contains several radiation variables, but not the downward radiations required by HTESSEL. To run HTESSEL without major technical modifications, we had to use a radiation replacement for all TIGGE models and ensemble members. We used ERAI-Land for this purpose, as it does not favor any of the TIGGE models used in this study. This way, for one daily run the same single radiation forecast was used for all ensemble members and all models. These 10-day radiation forecasts were built from 12-h ERAI-Land short-range predictions. To reduce the possible spinup effects in the first hours of the ERAI-Land forecasts, the 6–18-h radiation fluxes were combined (as 12-h sections) from subsequent 0000 and 1200 UTC runs, following the approach described in Balsamo et al. (2015). The sensitivity to the HTESSEL input variables will be discussed in section 4a.

In this study we were able to process four models out of the 10 global models archived in TIGGE: ECMWF, UKMO, NCEP, and CMA. The other six models do not archive one or more of the forcing variables, in addition to the downward radiation, required for this study.

The runoff produced by HTESSEL for TIGGE was routed over the river network by CaMa-Flood. These relatively short experiments for the TIGGE forecasts required initial river conditions. These were provided by three CaMa-Flood runs for the 1980–2010 period with ERAI, ERAI-Land, and MERRA-Land runoff input.

The discharge forecasts were produced by CaMa-Flood out to 10 days ($T + 240$ h), the longest forecast horizon common to all models. No perturbations were applied on the river initial conditions for the ensemble members. The forecasts were extracted from the CaMa-Flood 15-arc-min (~25 km) model grid for every 24-h similarly to the 24-h reporting frequency of the discharge observations.

3. Experiments

The main focus of the experiments was on the quality of the discharge forecasts derived from the TIGGE data. Three sets of experiments were performed to test the HTESSEL/CaMa-Flood setup and the scope for error reduction by applying the multimodel approach and different postprocessing methods:

- Discharge sensitivity to meteorological forcing: The first experiment (section 4a) tests the sensitivity of the forecasting system to the input variables.
- Reanalysis impact on discharge: The second experiment (section 4b) evaluates the potential improvements on the historical discharge that can be achieved by a combination of different reanalysis datasets.
- Improving the forecast distribution: In the third experiment (section 4c), the use of bias correction and model combination to improve the predictive distribution of the forecast is considered.

a. Discharge sensitivity to meteorological forcing

In section 2f, a number of compromises in the coupling of HTESSEL and forecasts from the TIGGE archive were introduced. Sensitivity experiments were conducted to study the impact of these. Table 1 provides a short description of the experiments.

The baseline for the comparisons is the discharge forecasts generated by HTESSEL and CaMa-Flood driven by ECMWF ensemble control (EC) forecasts. These forecasts were produced weekly (at 0000 UTC) throughout 2008–12 to cover several seasons (~260 forecast runs in total). In the baseline setup, the LML meteorological output for temperature, wind, and humidity was used to drive HTESSEL.

The first sensitivity test (Surf vs LML) was to replace these LML values with the surface values (as 2-m temperature and humidity and 10-m wind) from the same model run. This mirrors the change needed to make use of the TIGGE archive. Because of limitations in the TIGGE archive, the ERAI-Land radiation was used for all forecasts. Substitution of the ECMWF EC radiation in the HTESSEL input by ERAI-Land is the second sensitivity test (Rad). Further to this, substitution of the wind (Wind), temperature, humidity, and surface pressure together (THP), and precipitation (Prec) from ERAI-Land in place of the ECMWF EC run values was also

TABLE 1. Description of the sensitivity experiments with the ECMWF EC forecasts. The baseline is the reference run at the LML for wind, temperature, and humidity forcing. The other experiments are with different changes for the forcing variables. First, the LML is changed to surface (Surf), then different variables of the EC and their combinations are substituted by ERAI-Land data. Roman font means EC forcing input while italicized font denotes substituted ERAI-Land input.

Sensitivity expt	Forcing variable setup			
	Rad	THP	Wind	Prec
Baseline	EC-Surf	EC-LML	EC-LML	EC-Surf
Surf vs LML	EC-Surf	EC-Surf	EC-Surf	EC-Surf
Rad	*ERAI-Surf*	EC-LML	EC-LML	EC-Surf
Wind	EC-Surf	EC-LML	*ERAI-LML*	EC-Surf
THP	EC-Surf	*ERAI-LML*	EC-LML	EC-Surf
Rad + THP	*ERAI-Surf*	*ERAI-LML*	EC-LML	EC-Surf
Rad + THP + Wind	*ERAI-Surf*	*ERAI-LML*	*ERAI-LML*	EC-Surf
Rad + THP + Wind + Prec	*ERAI-Surf*	*ERAI-LML*	*ERAI-LML*	*ERAI-Surf*

evaluated. Temperature and humidity were analyzed together because of the sensitive nature of the balance between these two variables. Although these changes were not applied on the TIGGE data, they give a more complete picture on sensitivity to the forcing variables. This puts into context the discharge errors that we indirectly introduced through the TIGGE–HTESSEL setup changes.

The impact on the errors was compared by evaluating the ratio of the magnitude (absolute value) of the discharges to the baseline experiments, discharge value. These changes in relative discharge were computed for each station as the average of the relative changes over all runs (in the 2008–12 period with weekly runs) and also as a global average of all available stations.

b. Reanalysis impact on discharge

For the forecast of CaMa-Flood routing, the river initial conditions are provided by reanalysis-based simulations (see section 2d). They do not make use of observed river flow and therefore are an estimate of the observed values. The quality of the forecast discharge is expected to be strongly dependent on the skill of this reanalysis-derived historical discharge. This is highlighted in Fig. 2, where ERAI-Land, ERAI, and MERRA-Land are compared for a station in the United States for a 4-yr period.

Each of these reanalyses provides different error characteristics that can potentially be harnessed by using a multimodel approach. For this station ERAI has a tendency to produce occasional high peaks, while MERRA-Land has a strong negative bias. Although Fig. 2 is only a single example, it highlights the large variability between these reanalysis datasets and therefore a potentially severe underestimation of the uncertainties in the subsequent forecast experiments by using only a single initialization dataset.

The impact of the multimodel approach was analyzed by experiments with the historical discharges derived

from ERAI, ERAI-Land, and MERRA-Land inputs. Three sets of CaMa-Flood routing runs were performed for each of the four TIGGE models for the whole 22-month period in 2008–10, each initialized from one of the three reanalysis-derived historical river conditions. The performance of the historical discharge was evaluated independently of the TIGGE forecasts on the period of 1981–2010.

c. Improving the forecast distribution

In the third group of experiments a number of post-processing techniques were applied at each site with the aim of improving the forecast distribution for the observed data. Here we outline the techniques with reference to a single site and forecast origin t. The forecast values available are denoted $f_{m,j,t,i}$, where m indices over the forecast products (ECMWF, UKMO, NCEP, and CMA), j indices over the N_m ensemble members in forecast product m, and $i = 1, \ldots, 10$ indicates the available lead times.
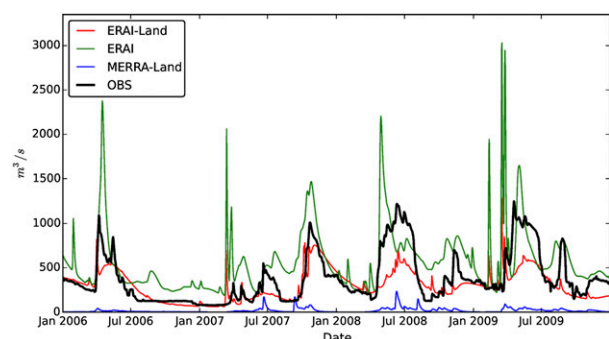

FIG. 2. Example of discharge produced by ERAI-Land (red), ERAI (green), and MERRA-Land (blue) forcing and the corresponding observations (black) for a GRDC station on the Rainy River at Manitou Rapids in the United States.

1) Bias correction

As a first step we analyzed the biases of the data. As described in section 3b, the historical river initial conditions have potentially large errors. In addition, the variability of the discharge in a 10-day forecast horizon is generally much smaller than derived from reanalysis over a long period. Therefore, any timing or magnitude error in the historical discharge provided initial conditions means the forecast errors can be very large and will change only slightly, in relative terms, throughout the 10-day forecast period.

As bias was expected to be a very important aspect of the errors, three methods of computing the bias correction $e_{m,j,t,i}$ to add to the forecast $f_{m,j,t,i}$ were proposed. The first of these is to apply no correction (or uncorrected); that is, $e_{m,j,t,i} = 0$ in all cases. The second method, referred to as 30-day correction, removes the mean bias of the 30-day period preceding the actual forecast run for each forecast product at each specified forecast range. The mean bias is computed as an average error of the ensemble mean over a 30-day period. In this case, given a series of 30 dates $t = 1, \ldots, N_{30}$ and observed discharge data y_t, the bias corrections are given by

$$\frac{\sum_{t=1}^{N_{30}} \sum_{j=1}^{N_m} (y_t - f_{m,j,t-i,i})}{N_{30} N_m}.$$

The third correction method, referred to as initial time correction, focused specifically on the historical discharge-based initial condition errors. The error at initialization of the routing $f_{m,j,t,0}$, that is, the error of the historical discharge, was used as a correction for all forecast ranges. This initial time correction gives

$$e_{m,j,t,i} = y_t - f_{m,j,t,0}.$$

This method therefore uses a specific error correction for each individual forecast run from day 1 to day 10. Because of the common initialization, the initial time correction was the same for all four TIGGE models for all three historical discharge experiments, respectively.

2) Multimodel combination

To investigate the potential further benefits of combining different forecast products, two model combination strategies were trialed. The naïve combination strategy [also referred to as multimodel combination (MM)] was based on utilizing a grand ensemble with each member having equal weight. In this combination, the larger ensembles (the largest being ECMWF with

50 members) get larger weights. In direct analogy to the case of a single forecast product, the cumulative forecast distribution is expressed in terms of the indicator function $\delta(z)$, which takes the value 1 if the statement z is true and 0 otherwise, as

$$\Pr(Y_{t+i} < y) = \frac{\sum_m \sum_{j=1}^{N_m} \delta(f_{m,j,t,i} + e_{m,j,t,i} < y)}{\sum_m N_m}.$$

Here $e_{m,j,t,i}$ indicates one of the three bias corrections we introduced in the previous section.

In the second combination strategy, BMA was used to explore further the effects of weighted combination and a temporally localized bias correction. Since discharge is always positive, the variables were transformed so that their distributions marginalized over time are standard Gaussian. This is achieved using the normal quantile transform (Krzysztofowicz 1997), with the upper and lower tails handled as in Coccia and Todini (2011). The transformed values of the bias-corrected forecasts and observations are denoted $\tilde{f}_{m,j,t,i}$ and $\tilde{y}_t$, respectively.

This study follows the BMA approach proposed by Fraley et al. (2010) for systems with exchangeable members with the weight $w_{m,t,i}$, linear bias correction (with parameters $a_{m,t,i}$ and $b_{m,t,i}$), and nugget variance $\sigma_{m,t,i}^2$ being identical for each ensemble member within a given forecast product. The resulting cumulative forecast distribution in the transformed space is then a weighted combination of standard Gaussian cumulative distributions Φ, specifically,

$$\Pr(\tilde{Y}_{t+i} < y) = \sum_m w_{m,t,i} \sum_{j=1}^{N_m} \Phi\left(\frac{y - a_{m,t,i} - b_{m,t,i}\tilde{f}_{m,j,t,i}}{\sigma_{m,t,i}}\right).$$

As indicated by the origin and lead time subscripts, the BMA parameters were estimated for each forecast origin and lead time. Estimation bias proceeds by first fitting the linear correction using least squares before estimating the weight and variance terms using maximum likelihood (Raftery et al. 2005). A moving window of 30 days of data, before the initialization of the forecasts similarly to the 30-day correction, was utilized for the estimation to mimic operational practice.

As the initial conditions were expected to play an important role, a further forecast was introduced in the context of the BMA analysis. The deterministic persistence forecast is, throughout the 10-day forecast range, the most recent observation available at time of issue, that is,

$$f_{m,1,t,i} = y_t.$$

This persistence forecast was also used as a simple reference to compare our forecasts against.

To aid comparison with the naïve combination strategy a similar-sized ensemble of forecasts was generated from the BMA combination by applying ensemble copula coupling (Schefzik et al. 2013) to a sample generated by taking equally spaced quantiles from the forecast distribution and reversing the transformation.

3) VERIFICATION STATISTICS

The forecast distributions were evaluated using the continuous ranked probability score (CRPS; Candille and Talagrand 2005). The CRPS evaluates the global skill of the ensemble prediction systems by measuring a distance between the predicted and the observed cumulative density functions of scalar variables. For a set of dates $t = 1, \ldots, N$ with observations and probabilistic forecasts issued with the same lead time (which are realizations of the random variables Y_{t+i}), the CRPS can be defined as

$$\mathrm{CRPS} = \frac{1}{N_t} \sum_{t=1}^{N_t} \int_{-\infty}^{\infty} [\mathrm{Pr}(Y_{t+i} < y) - \delta(y \geq y_t)]^2 \, dy.$$

The CRPS has a perfect score of 0 and has the advantage of transforming into the mean absolute error for deterministic forecasts and thus providing a simple way of comparing different types of systems. In this study the method of Hersbach (2000) for computing the CRPS from samples was used. The global CRPS reported for each lead time were produced by pooling the samples from all the stations before computing the scores.

As the CRPS has the unit of the physical quantity (e.g., for discharge $\mathrm{m^3\,s^{-1}}$), comparing scores can be problematic and is only meaningful if two homogeneous sample-based scores are compared. For example, different geographical areas or different seasons cannot really be compared. In this study we ensured that, for any comparison of forecast models and postprocessed versions, the samples were homogeneous. We considered the same days in the verification period at each station specifically, and also the same stations in the global analysis, producing equal sample sizes across all compared products.

To help compare results across different stations and areas, we used the CRPS-based skill score (CRPSS) with the reference system of the observed discharge climate in our verification. We produced the daily observed climate for the 30-yr period of 1981–2010 and pooled observations from a 31-day window centered over each day. Observed climate was produced for stations with at least 10 years of data available in total (310 values) for all days of the year.

Each of the historical discharge experiments produce a time series of discharges (f_t: $t = 1, \ldots, N$), which were compared to the observed data using the mean absolute error (MAE)-based skill score (MAESS) with the observed daily discharge climate (obsclim) as reference,

$$\mathrm{MAESS} = 1 - \frac{\mathrm{MAE}}{\mathrm{MAE}_{\mathrm{obsclim}}},$$

and the sample Pearson correlation coefficient (CORR),

$$\mathrm{CORR} = \frac{\sum_{t=1}^{N} (f_t - \overline{f})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{N} (f_t - \overline{f})^2} \sqrt{\sum_{t=1}^{N} (y_t - \overline{y})^2}},$$

where the bar denotes the temporal average of the variable. The MAE reflects the ability of the systems to match the actual observed discharge, while the correlation highlights the quality of match between the temporal behavior of the historical forecast time series and the observation time series.

4. Results

First, we present the findings of the sensitivity experiments carried out, using the ECMWF EC forecast, on the impact of the HTESSEL coupling with the TIGGE meteorological input. Then we compare the quality of the historical discharge produced from the ERAI, ERAI-Land, and MERRA-Land datasets and the impact of their combination. Finally, from the large number of forecast products described in sections 3b and 3c, we present results that aid interpretation of the discharge forecast skills and errors with focus on the potential multimodel improvements:

- the four uncorrected TIGGE forecasts with ERAI-Land initialization;
- the MM combination of the four uncorrected models with the ERAI, ERAI-Land, and MERRA-Land initializations and the grand ensemble of these three MM combinations (called GMM hereafter);
- the GMM combinations of the 30-day-corrected, the initial-time-corrected, and the combined initial-time- and 30-day-corrected MM forecasts (first initial-time-correct the forecasts, then apply the 30-day correction on these); and
- finally, the GMM of the BMA combined MM forecasts (from all three initializations) with the uncorrected models, the initial-time-corrected models, and also the uncorrected models extended by the persistence as a separate single value model.
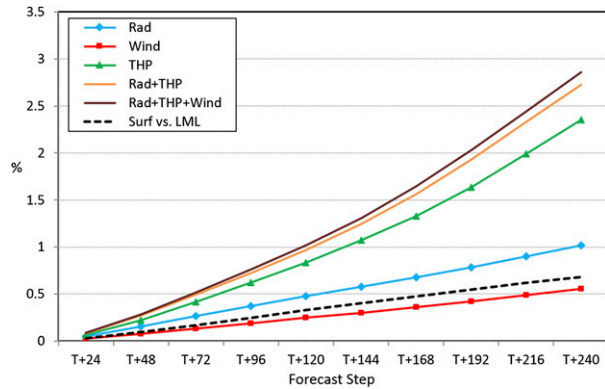
FIG. 3. Impact of different forcing configurations in HTESSEL on the discharge outputs as a relative change compared to baseline. The black dashed line displays the impact of changing the LML to surface forcing (2 m for temperature and humidity, 10 m for wind). The colored lines highlight the impact of replacing different EC forcing variables, either individually or in combination, with ERAI-Land data.

a. Discharge sensitivity to meteorological forcing

The impact of replacing HTESSEL forcing variables other than precipitation (combination of Rad, Wind, and THP tests) with ERAI-Land (Fig. 3) is rather small (~3% by $T + 240$, brown curve in Fig. 3). The least influential is the Wind (red curve), while the biggest contribution comes from the THP (green curve). When all ensemble forcing is replaced, including precipitation, the impact jumps to ~15% by $T + 240$ h, showing that a large majority of the change in the discharge comes from differences in precipitation (not shown).

The analysis of different areas and periods (see Table 2) highlights that larger impacts are seen for the winter period where the contribution of precipitation decreases and the contribution of the other forcing variables, both individually and combined, increases by approximately twofold to fivefold (this is particularly noticeable for THP). This is most likely a consequence of the snow-related processes, with snowmelt being dependent on temperature, radiation, and also wind in the cold seasons.

This also implies that the results are dependent on seasonality, a result that was also found by Liu et al. (2013), who looked at the skill of postprocessed precipitation forecasts using TIGGE data for the Huai River basin in China. In this study, because of the relatively short period we were able to use in the forecasts verification, scores were only computed for the whole verification period and no seasonal differences were analyzed.

Regarding the change from LML to surface forcing for temperature (2 m), wind (10 m), and humidity (2 m), the potential impact can be substantial, as shown by an example for 1–10 January 2012 in Fig. 4. In such cold winter conditions, large erroneous surface runoff values could appear in some parts of Russia when switching to surface forcing in HTESSEL. The representation of dew deposition is a general feature of HTESSEL that can be amplified in stand-alone mode. When coupled to the atmosphere, the deposition is limited in time, as it leads to a decrease of atmospheric humidity. However, in stand-alone mode, since the atmospheric conditions are prescribed, large deposition rates can be generated when the atmospheric forcing is not in balance (e.g., after model grid interpolation or changing from LML to surface forcing).

This demonstrates that with a land surface model such as HTESSEL, particular care needs to be taken in design of the experiments when model imbalances are expected. The use of surface data was an acceptable compromise as the sensitivity experiments highlighted only a small impact caused by the switch from LML to surface forcing (black dashed line in Fig. 3), and similarly by the impact of the Rad test, confirming that the necessary changes in the TIGGE land surface model setup did not have a major impact on the TIGGE discharge.

b. Reanalysis impact on discharge

The quality of the historical river flow that provides initial conditions for the CaMa-Flood TIGGE routing is expected to have a significant impact on the forecast skill. We analyze the discharge performance that is

TABLE 2. Detailed evaluation of the discharge sensitivity experiments at $T + 240$ h range for different areas and periods. Relative discharge differences are shown after replacing EC forcing variables, either individually or in combination, by ERAI-Land, and also the LML with surface forcing (2 m for temperature and humidity, 10 m for wind). The whole globe, the northern extratropics (defined here as 35°–70°N), and the tropics (30°S–30°N) as well as the specific seasons are displayed.

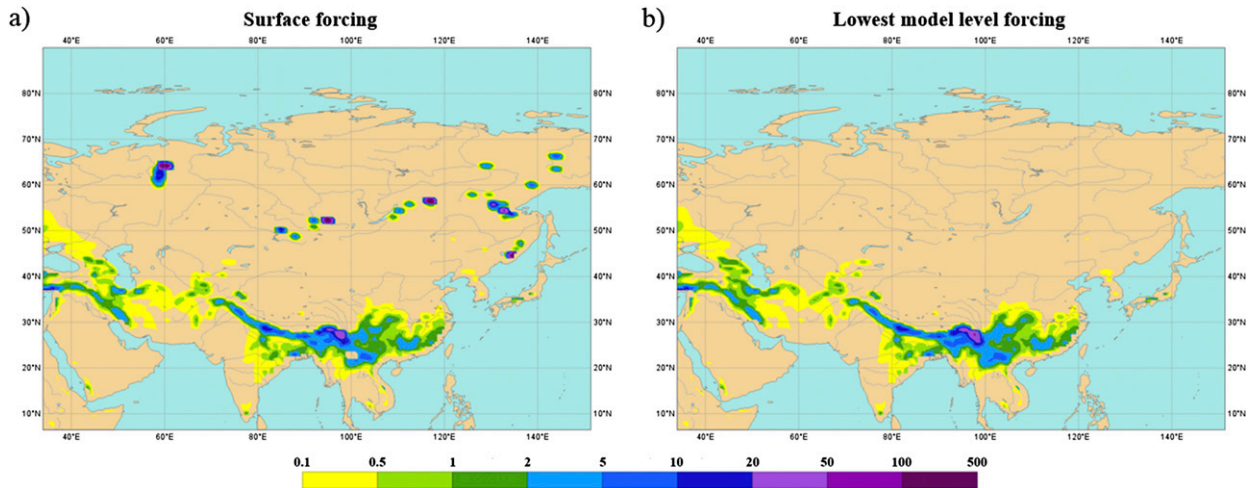| Avg diff (%)	Rad	THP	Wind	Rad + THP + Wind	Rad + THP + Wind + Prec	Surf vs LML
Global	1.0	2.4	0.6	2.9	15.6	0.7
Northern extratropics	1.0	3.1	0.6	3.5	12.8	0.8
Northern extratropics JJA	0.5	0.6	0.2	0.9	13.0	0.3
Northern extratropics DJF	1.1	2.9	0.7	3.6	9.5	0.8
Tropics	1.0	1.1	0.5	1.8	17.6	0.5
Tropics JJA	0.9	0.9	0.4	1.5	15.0	0.5
Tropics DJF	1.2	1.3	0.6	2.1	18.7	0.6

FIG. 4. Surface runoff output of HTESSEL for the period 1–10 Jan 2012 (240-h accumulation) from two EC experiments, using (a) surface forcing and (b) LML forcing, where possible. In (a), very large erroneous surface runoff values appear in very cold winter conditions.

highlighted in Fig. 5. This shows the MAESS and CORR for the ERAI-, ERAI-Land-, and MERRA-Land-simulated historical discharge from 1981 to 2010, and for their equal-weight multimodel average (MMA). The results are provided as continental and also as global averages of the available stations for Europe (~150 stations), North America (~350 stations), South America (~150 stations), Africa (~80 stations), Asia (mainly Russia, 60 stations), and Australia and Indonesia (~50 stations), making ~840 stations globally.

The general quality of these global simulations is quite low. The MAESS averages over the available stations (see Fig. 1) are <0 for all continents, that is, large-scale average performance is worse than the daily observed climatology. The models are closest to the observed climate performance over Europe and Australia and
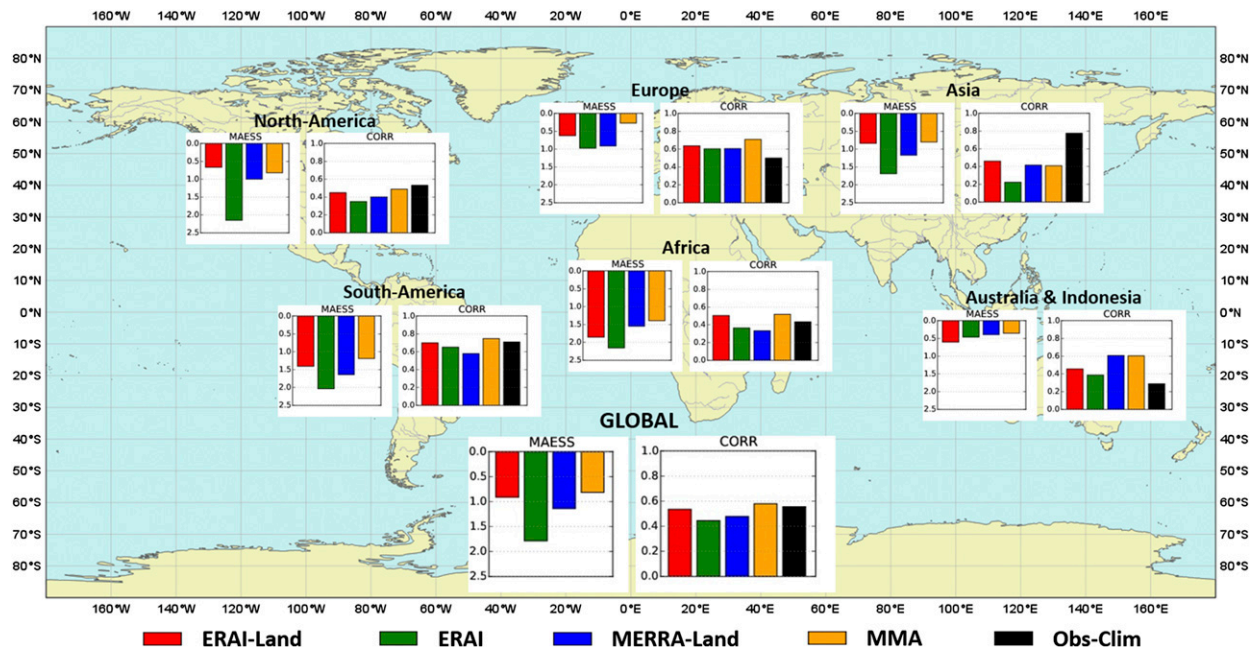
FIG. 5. Historical discharge forecast performance for ERAI-Land, ERAI, MERRA-Land, and their equal-weight MMA. MAESS and CORR are provided for each continent (North America, South America, Europe, Africa, Asia, and Australia and Indonesia). The reference forecast system in the skill score is the observed discharge climate as daily prediction. CORR are also provided for the observed climate. The scores are continental and global averages of the individual scores of the available stations (for station reference, see Fig. 1).
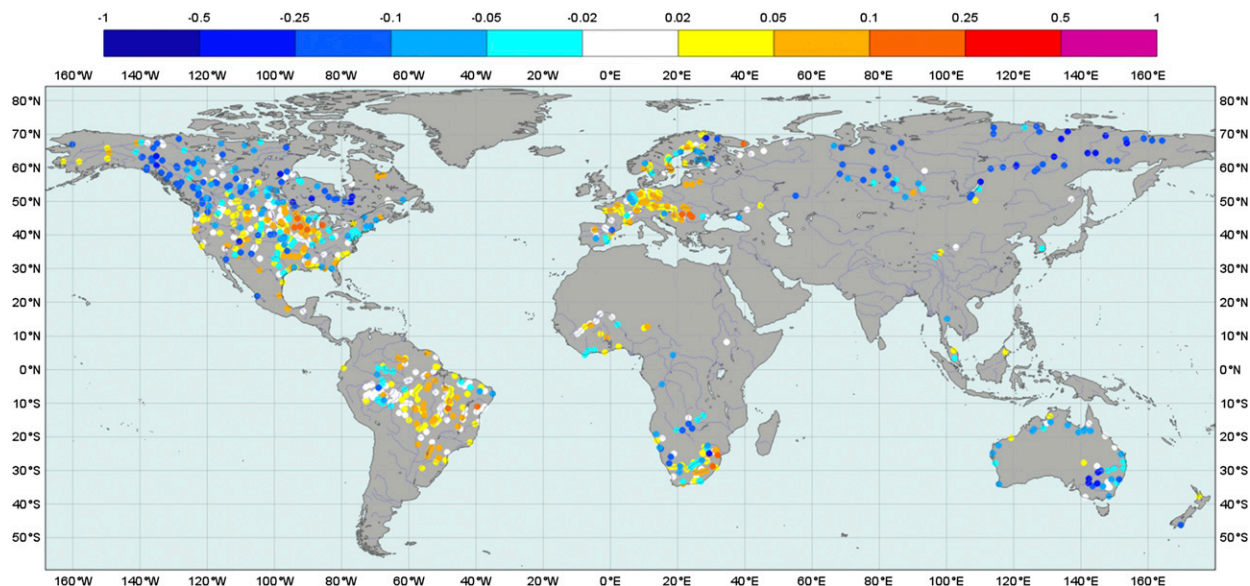
FIG. 6. Relative improvements in CORR by equal-weight average of ERAI-Land, ERAI, and MERRA-Land discharges. Values show the change in CORR compared with the best of ERAI-Land, ERAI, and MERRA-Land. Positive values show improvement while negative change means lower skill in the average than in the best of the three historical discharges.

Indonesia. The correlation between the simulated and observed time series shows a slightly more mixed picture, at least in some cases; especially Europe and Australia and Indonesia, the model is better than the observed climate. It is interesting to note that although the observed climate produces a high forecast time series correlation, in Asia the reanalysis discharge scores very low for all three sources. This could be related to the problematic handling of the snow in that area.

Figure 2 shows an example where MERRA-Land displayed a very strong negative bias. This example highlights the large variability among these data sources and is not an indication of the overall quality. Although MERRA-Land shows generally negative bias (not shown), the overall quality of the three reanalysis-driven historical discharge datasets is rather comparable. The highest skill and correlation is generally shown by ERAI-Land for most of the regions with the exception of Africa and Australia, where MERRA-Land is superior. ERAI, as the oldest dataset, appears to be the least skillful. Reichle et al. (2011) have found the same relationship between MERRA-Land and ERAI using 18 catchments in the United States. Although they computed correlation between seasonal anomaly time series (rather than the actual time series evaluated here), they could show that runoff estimates had higher correlation of the anomaly time series in MERRA-Land than in ERAI.

The multimodel average of the three simulations is clearly superior in the global and also in the continental

averages, with very few exceptions that have marginally lower MMA scores compared with the best individual reanalysis. The MMA is able to improve on the best of the three individual datasets at about half of the stations globally, both in the MAESS and CORR. Figure 6 shows the improvements in correlation. The points where the combination of the three reanalyses helps to improve on the best model cluster are mainly over Europe, Amazonia, and the eastern United States. On the other hand, the Northern Hemisphere winter areas seem to show mainly deterioration. This again is most likely related to the difficulty in the snow-related processes, which can hinder the success of the combination if, for example, one model is significantly worse with larger biases than the other two. Further analysis could help identify these more detailed error characteristics, providing a basis for further potential improvements.

c. Improving the forecast distribution

Figure 7 displays example hydrographs of some analyzed forecast products for a single forecast run to provide a practical impression of our experiments. The forecasts from 18 April 2009 are plotted for the GRDC station of Lobith in the Netherlands. The thin solid colored lines are the four TIGGE models (ECMWF, UKMO, NCEP, and CMA) plotted together (MM) with ERAI-Land (red), ERAI (green), and MERRA-Land (blue) initializations. They start from very different levels that are quite far from the observation (thick black line), but then seem to converge to roughly the
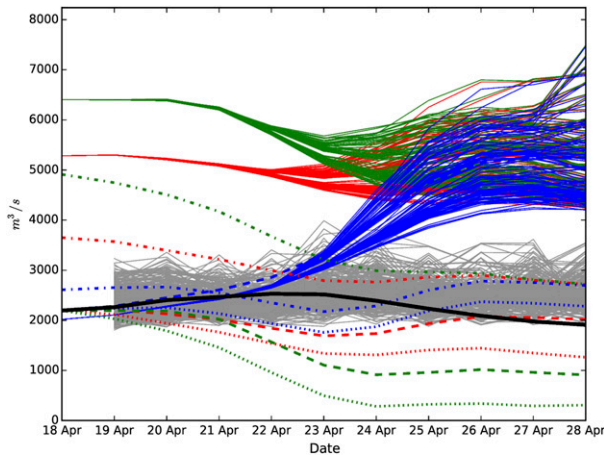
FIG. 7. Example of different discharge forecast products for the GRDC station of Lobith on the Rhine River in the Netherlands. All forecasts are from the run at 0000 UTC 18 Apr 2009 up to $T + 240$ h. The following products are plotted: multimodel combinations of four TIGGE models (ECMWF, UKMO, NCEP, and CMA) with ERAI-Land (solid red lines), ERAI (solid green lines), and MERRA-Land (solid blue lines) initializations; 30-day-corrected (dashed–dotted lines), initial-time-corrected (dashed lines), and 30-day- and initial-time-corrected (dotted lines) versions of the three multimodel combinations, each with all three initializations (with the respective colors); and finally, the BMA versions of the three multimodel combinations (all with gray lines, only from $T + 24$ h). The verifying observations are displayed by the black line.

FIG. 8. Discharge forecast performance for forecast ranges from $T + 0$ h to $T + 240$ h from August 2008 to May 2010 as global averages of CRPSS (computed at each station over the whole period) with the following forecast products. Gray lines indicate the four TIGGE models (ECMWF, UKMO, NCEP, and CMA) with ERAI-Land initialization, and a multimodel combination of these four models with ERAI-Land (red line), ERAI (green line), and MERRA-Land (blue line) initialization is also shown. The orange line represents a grand combination of these three multimodels, and grand combinations for six postprocessed products are shown: the multimodel of the 30-day correction (burgundy dashed line), the initial error correction (purple dashed line with markers), the 30-day and initial error correction combination (solid burgundy line with markers), two BMA versions of the multimodel—one with the uncorrected forecasts (black line without markers) and one with the uncorrected forecasts extended by the persistence as predictor (black line with circles)—and the persistence forecast (light blue dashed line with circles). The CRPSS is positively oriented and has a perfect value of 1. The 0 value line represents the quality of the reference system, the daily observed discharge climate.

same range in this example. The ensemble mean of the initial error-corrected MMs (from the three initializations with dashed lines), which by definition start from the observed discharge at $T + 0$ h, then follow faithfully the pattern of the mean of the respective MMs. The 30-day-corrected forecasts (dashed–dotted lines) follow a pattern relative to the MM ensemble means set by the performance of the last 30 days. The combination of the two bias-correction methods (dotted lines) blends the characteristics of the two; all three versions start from the observation (as first the initial error is removed) and then follow the pattern set by the past 30-day performance of this initial time-corrected forecast. Finally, the BMA-transformed (uncorrected) MMs (thin gray lines) happen to be closest to the observations in this example, showing a rather uniform spread throughout the processed range from $T + 24$ h to $T + 240$ h.

The quality of the TIGGE discharge forecasts based on the verified period from August 2008 to May 2010 is strongly dependent on the historical discharge that is used to initialize them. Figure 5 highlighted that the daily observed discharge climate is a better predictor than any of the three historical reanalysis-driven discharges (MAESS < 0). It is therefore not surprising that the uncorrected TIGGE forecasts show similarly low relative skill based on the CRPS (Fig. 8). Figure 8 also
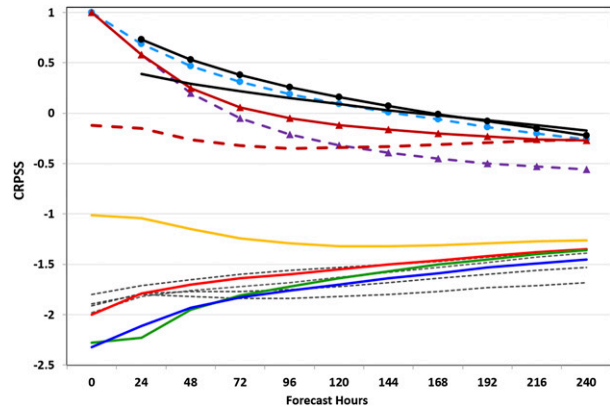
shows the performance of the four models (gray dashed lines). In this study, we concentrate on the added value of the multimodel combination and do not distinguish between the four raw models. The scores change very little over the 10-day forecast period, showing a marginal increase in CRPSS as lead time increases. This is indicative of the incorrect initialization, with the forecast outputs becoming less dependent on initialization further into the medium range, and slowly converging toward climatology.

The first stage of the multimodel combination is the red line in Fig. 8, the combination of the uncorrected four models with the same ERAI-Land initialization. On the basis of this verification period and global station list, the simple equal-weight combination of the ensembles does not really seem to be able to improve on the best model. However, we have to acknowledge that the performance in general is very low.

The other area where we expect improvements through the multimodel approach is the initialization. Figure 8 highlights a significant improvement when using

three historical discharge initializations instead of only one. The quality of the ERAI-Land (red), ERAI (green), and MERRA-Land (blue) initialized forecasts (showed here only the multimodel combination versions) are comparable, with the ERAI-Land slightly ahead, which is in agreement with the results of the direct historical discharge comparisons presented in section 4b. However, the grand combination of the three is able to improve significantly (orange line) on all of them. The improvement is much larger at shorter lead times as the TIGGE meteorological inputs provide lower spread, and therefore the spread introduced by the different initializations is able to have a bigger impact.

The quality of the discharge forecasts could be improved noticeably by introducing different initial conditions. However, the CRPSS is still significantly below 0, pointing to the need for postprocessing. In this study, we have experimented with a few methods that were proven to be beneficial.

The 30-day correction removed the mean bias of the most recent 30 runs from the forecasts. Figure 8 shows the grand combination of the 30-day bias-corrected multimodels (with all three initializations), which brings the CRPSS to almost 0 throughout the 10-day forecast range (burgundy dashed line in Fig. 8). This confirms that the forecasts are severely biased. In addition, the shape of the curve remains fairly horizontal, suggesting this correction is not making the best use of the temporal patterns in the bias.

Further significant improvements in CRPSS are gained at shorter forecast ranges by using the initial time correction (purple dashed line with markers in Fig. 8), which does make use of temporal patterns in the bias. The shape of this error curve shows a typical pattern with the CRPSS decreasing with forecast range, reflecting the decreasing impact of the initial time correction and increased uncertainty in the forecast. The impact of the initial time errors gradually decreases until it finally disappears by around day 5 or 6, when the 30-day correction becomes superior.

The combination of the two methods, by applying the 30-day bias correction to forecasts already adjusted by the initial time correction, blends the advantages of both corrections. The CRPSS is further improved mainly in the middle of the 10-day forecast period with disappearing gain by $T + 240$ h (solid burgundy line with markers in Fig. 8).

The fact that the performance of the 30-day correction is worse in the short range than the initial time correction highlights that the impact of the errors at initial time has a structural component that cannot be explained by the temporally averaged bias. Similarly, the initial time correction cannot account exclusively

for the large biases in the forecasts as its impact trails off relatively quickly.

The persistence forecast shows a distinct advantage over these postprocessed forecasts (light blue dashed line with circles in Fig. 8). There is positive skill up to $T + 144$ h and the advantage of the persisted observation as a forecast diminishes, so that by $T + 240$ h its skill is similar to that of the combined corrected forecasts. This further highlights that the utilization of the discharge observations in the forecast production promises to provide a really significant improvement.

It is suggested that the structure of the initial errors has two main components: (i) biases in the reanalysis initializations due to biases in the forcing (e.g., precipitation) and in the simulations (e.g., evapotranspiration) and (ii) biases introduced by timing errors in the routing model due, in part, to the lack of optimized model parameters. A further evaluation of the weight of each of these error sources is beyond the scope of this study.

The final of our trialed postprocessing methods is the BMA. In Fig. 8, similarly to the other postprocessed products, only the grand combination is displayed of the three BMA-transformed MMs with the different initializations. The BMA of the uncorrected forecasts was able to increase further the CRPSS markedly across all forecast ranges except $T + 24$ h (black line without markers). The results for $T + 24$ h suggest that at this lead time the perfect initial error correction from $T + 0$ h still holds superior.

The other two BMA versions, one with the uncorrected forecasts extended by the persistence as predictor (black line with circles) and one with the initial-time-corrected forecasts (not shown), both provide further skill improvements. The one with the persistence performs overall better, especially in the first few days. The BMA incorporating the persistence forecast remains skillful up to $T + 168$ h, the longest lead time of any of the forecast methods tested. At longer lead times (days 8–10) the BMA of the uncorrected model forecast appears to provide the highest skill of all the postprocessed products. This is evidence that the training of the BMA is not optimal. This is in part due to the estimation methodology used. More significantly experiments (not reported) show that the optimal training window for the BMA varies across sites, showing a different picture for the BMA with or without persistence, and also delivering potentially higher global average skill using a longer window.

Although Fig. 8 shows only the impact of the four postprocessing methods on the grand combination of the MM forecasts, the individual MMs with the three initializations show the same behavior. The GMMs
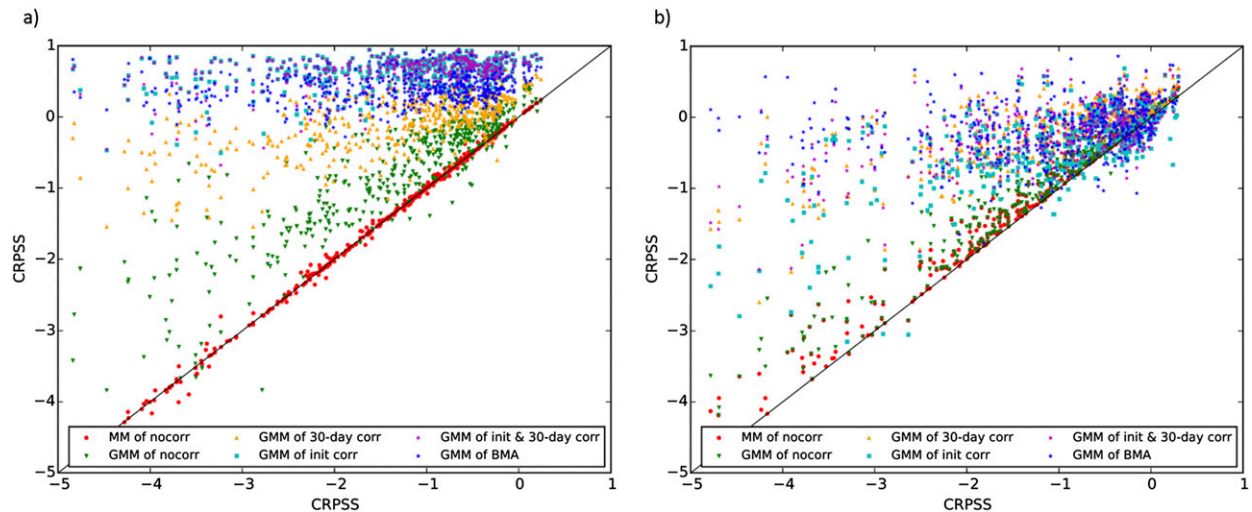
FIG. 9. Distribution of the skill increments over all stations provided by six combination and postprocessing products for two time ranges: (a) $T + 24$ h and (b) $T + 240$ h. The x axis shows the reference skill, the average CRPSS of the four TIGGE models with the ERA-Land initialization, while the y axis displays the CRPSS of the postprocessed forecasts at the stations. The six products are the MM combination of the four models with ERAI-Land (red circles), the GMM combination of the three uncorrected MMs (green triangles), and the GMM combination of four postprocessed products: the 30-day-corrected MMs (orange triangles), the initial-time-corrected MMs (cyan squares), the combined 30-day- and initial-time-corrected MMs (purple stars), and finally the BMA-transformed MMs of the uncorrected forecasts (blue stars), where all the MMs are the three MM with the different initializations. The diagonal line represents no skill improvement; above this line the six products are better, while below it they are worse than the reference. The CRPSS values are computed based on the period from August 2008 to May 2010. Some of the stations that have reference CRPSS below −5 are not plotted.

always outperform the three MMs for all the post-processing products; for example, for the most skillful method, the BMA, the grand combination extends the positive skill by ~1 day (from around 5 days to 6 days, not shown).

The distribution of the skill increments over all stations provided by different combination and post-processing products is summarized in Fig. 9 at $T + 24$ h (Fig. 9a) and $T + 240$ h (Fig. 9b). The reference skill is the average CRPSS of the four TIGGE models with the ERA-Land initialization (these values are represented by the gray dashed lines in Fig. 8). Figure 9 highlights the structure of the improvements in different ranges of the CRPSS for the different methods over all verified stations in the period from August 2008 to May 2010. The picture is characteristically different at different lead times, as suggested by the $T + 24$ h and $T + 240$ h plots. At short range, the improvements of the different products scale nicely into separate bands. The relatively simple MM combination of the four models with ERAI-Land (red circles) does not improve on the forecast; the increments are small and with mixed sign. The GMM combination of the three uncorrected MMs (green triangles) shows a marked improvement, and the 30-day correction version (orange triangles) improves further while the initial time correction products (cyan squares and purple stars) show the largest improvement over most of the stations. At this short $T + 24$ h range, the

BMA (blue stars) of the uncorrected forecasts is slightly behind, which is a general feature across the displayed CRPSS range from −5 to 1.

In contrast to the short range, $T + 240$ h provides a significantly different picture. The relatively clear ranking of the products is gone by this lead time. The MM and GMM combinations are able to improve slightly for most of the stations, but at this range the contribution seems to be generally always positive. The postprocessing methods at this medium range, however, deteriorate the forecasts sometimes, especially in the range from −1 to 0.5 (the 30-day correction seems to behave noticeably better in this respect). The general improvements are clear though for most of the stations, and also the overall ranking of the methods seen in Fig. 8 is reflected, although much less clearly than at $T + 24$ h, with the BMA topping the list at $T + 240$ h.

Finally, Fig. 10 presents the discharge performance we could achieve in this study for all the stations that could be processed in the period from August 2008 to May 2010 at $T + 240$ h. It displays the CRPSS of the best overall product, the GMM with the BMA of the uncorrected forecasts (combination of the three BMA-transformed MMs with the three initializations without initial time or 30-day bias correction). The variability of the scores is very large geographically, but there are emerging patterns. Higher performance is observed in the Amazon and in central and western parts of the
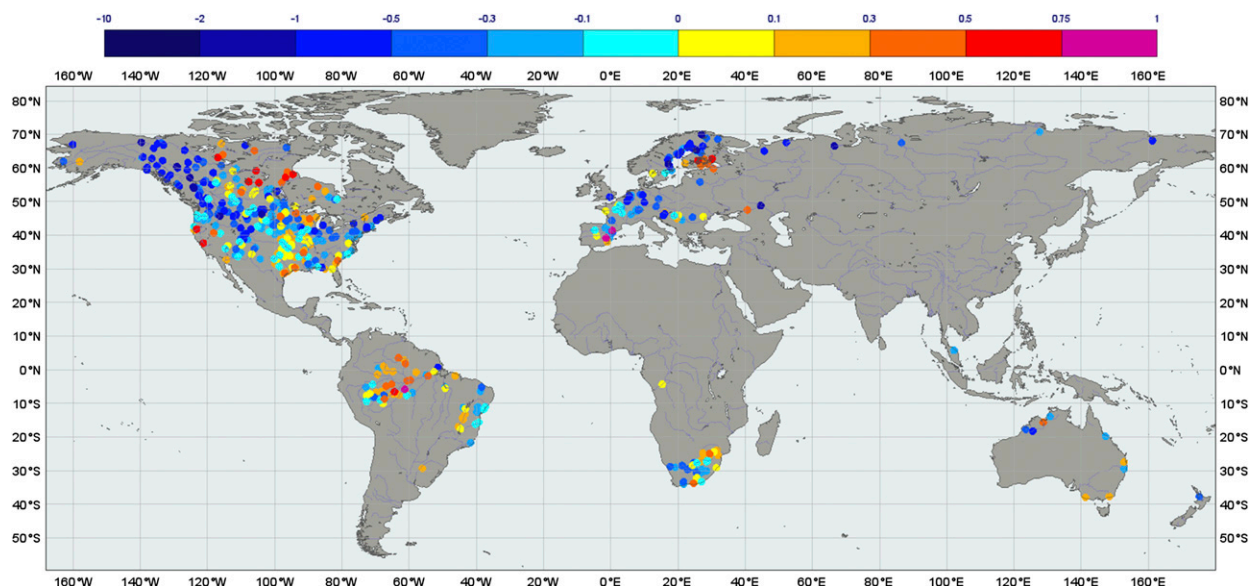
FIG. 10. Global CRPSS distribution of the highest quality postprocessed product at $T + 240$ h, the grand multimodel combination of the BMA-transformed uncorrected forecasts, based on the period from August 2008 to May 2010. The CRPSS is positively oriented and has a perfect value of 1. The 0 value represents the quality of the reference system, the daily observed discharge climate.

United States, while lower CRPSSs are seen over the Rocky Mountains in North America and in northerly points in Europe and Russia. Unfortunately, the geographical coverage of the stations is not good enough to draw more detailed conclusions.

5. Conclusions

This study has shown aspects of building a global multimodel hydrometeorological forecasting system using the TIGGE archive and analyzed the impact of the postprocessing required to run a multimodel system on the forecasts.

The atmospheric input was taken from four operational global meteorological ensemble systems, using data available from TIGGE. The hydrological component of this study was the HTESSEL land surface model while the CaMa-Flood global river-routing model was used to integrate runoff over the river network. Observations from the GRDC discharge archive were used for evaluation and postprocessing.

We have shown that the TIGGE archive is a valuable resource for river discharge forecasting, and three main objectives were successfully addressed: (i) the sensitivity of the forecasting system to the meteorological input variables, (ii) the potential improvements to the historical discharge dataset (which provides initial river conditions to the forecast routing), and (iii) improving the predictive distribution of the forecasts. The main outcomes can be grouped as follows:

(i) The impact of replacing or altering the input meteorological variables to fit the system requirements is small and allows the use of variables from the TIGGE archive for this hydrological study.

(ii) The multimodel average historical discharge dataset provides a very valuable source of uncertainty and a general gain in skill.

(iii) Significant improvements in the forecast distribution can be produced through the use of initial time and 30-day bias corrections on the TIGGE model discharge, or on the combination of the forecast models; however, the combination of techniques used has a big impact on the improvement observed, with the best BMA products providing positive skill up to 6 days.

The quality of the raw TIGGE-based discharge forecasts has been shown to be low, mainly determined by the limited performance of the reanalysis-driven historical river conditions analyzed in section 4b. The lower skill is in agreement with results found in other studies. For example, Alfieri et al. (2013) showed that in the context of GloFAS, the LISFLOOD hydrological model (Van Der Knijff et al. 2010), forced by ERAI-Land runoff, shows variable performance based on the 1990–2010 historical period. From the analyzed 620 global observing stations the Pearson's correlation coefficient reaches as low as -0.2, and only 71% of them provide correlation values above 0.5. Donnelly et al. (2015) highlighted similar behavior with the E-HYPE system based on 181 river gauges in Europe for 1981–2000. The

correlation component of the Kling–Gupta efficiency started around 0, and geographical distribution of values in Europe was very similar to our result (not shown). The lowest correlation was found mainly in Spain and in Scandinavia, with a comparable average value to our European mean of 0.6–0.7 (see Fig. 5).

The combination and postprocessing methods we applied to the discharge forecasts provided significant improvement of the skill. Although the simple multimodel combinations and the 30-day bias correction (removing the mean error of the most recent 30 days) both provide significant improvements, they are not capable of achieving positive global skill (i.e., outperform the daily observed discharge climate). The initial time correction, by adjusting to the observations at initial time and applying this error correction into the forecast, is able to provide skill in the short range (only up to 2–3 days), especially when combined with the 30-day correction. However, the impact quickly wears off and for longer lead times (up to about 6 days) only the BMA postprocessing method is able to provide positive average global skill (closely followed by the persistence).

Although other studies could show significant improvement by using multiple meteorological inputs (e.g., Pappenberger et al. 2008), in this study the impact of combining different TIGGE models is rather small. This is most likely a consequence of the overwhelming influence of the historical river conditions on the river initialization. The grand combinations, when we combine the forecasts produced with different reanalysis-driven historical river conditions, however, always outperform the individual MMs (single initialization) for all the postprocessing products. They provide a noticeable overall skill improvement, which in our study translated into an extension of the lead time, when the CRPSS drops below 0, by about one day as a global average for the most skillful BMA forecasts.

In the future we plan to extend this study to address other aspects of building a skillful multimodel hydrometeorological system. The following areas are considered:

(i) Include other datasets that provide global coverage of runoff data on high enough horizontal resolution, such as the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015) or the NCEP Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) to provide further improvements in the initial river condition estimates.

(ii) Introduce the multihydrology aspect by adding an additional land surface model such as the Joint UK Land Environment Simulator (JULES; Best et al. 2011).

(iii) The presented scores in this study are relatively low even with the postprocessing methods applied. To achieve significantly higher overall scores, the information on the discharge observations should be utilized in the modeling.

(iv) Similarly, the discharge quality could be significantly improved by better calibration of many of the watersheds in the CaMa-Flood routing.

(v) Alternatively, the application of different river-routing schemes such as LISFLOOD, which is currently used in the GloFAS, would also provide potential increase in the skill through the multimodel use.

(vi) Further analysis of the errors and the trialing of other postprocessing methods could also lead to potential improvements. In particular, better allowance should be made for temporal correlation in the forecast errors. The use of the extreme forecast index (Zsótér 2006) as a tool to compare the forecasts to the model climate could potentially bring added skill into the flood predictions.

REFERENCES

Agustí-Panareda, A., G. Balsamo, and A. Beljaars, 2010: Impact of improved soil moisture on the ECMWF precipitation forecast in West Africa. *Geophys. Res. Lett.*, **37**, L20808, doi:10.1029/2010GL044748.

Ajami, N. K., Q. Duan, and S. Sorooshian, 2007: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.*, **43**, W01403, doi:10.1029/2005WR004745.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, doi:10.5194/hess-17-1161-2013.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeor.*, **10**, 623–643, doi:10.1175/2008JHM1068.1.

——, S. Boussetta, P. Lopez, and L. Ferranti, 2010: Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation

over the U.S.A. ERA Rep. 01/2010, ECMWF, 10 pp. [Available online at http://www.ecmwf.int/sites/default/files/elibrary/2010/7926-evaluation-era-interim-and-era-interim-gpcp-rescaled-precipitation-over-usa.pdf.]

——, F. Pappenberger, E. Dutra, P. Viterbo, and B. van den Hurk, 2011: A revised land hydrology in the ECMWF model: A step towards daily water flux prediction in a fully-closed water cycle. *Hydrol. Processes*, **25**, 1046–1054, doi:10.1002/hyp.7808.

——, and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, doi:10.5194/hess-19-389-2015.

Bao, H. H., and O. Zhao, 2012: Development and application of an atmospheric–hydrologic–hydraulic flood forecasting model driven by TIGGE ensemble forecasts. *Acta Meteor. Sin.*, **26**, 93–102, doi:10.1007/s13351-012-0109-0.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, doi:10.5194/gmd-4-677-2011.

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, doi:10.1175/2010BAMS2853.1.

Candille, G., and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150, doi:10.1256/qj.04.71.

Cane, D., S. Ghigo, D. Rabuffetti, and M. Milelli, 2013: Real-time flood forecasting coupling different postprocessing techniques of precipitation forecast ensembles with a distributed hydrological model. The case study of May 2008 flood in western Piemonte, Italy. *Nat. Hazards Earth Syst. Sci.*, **13**, 211–220, doi:10.5194/nhess-13-211-2013.

Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, doi:10.1016/j.jhydrol.2009.06.005.

Coccia, G., and E. Todini, 2011: Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrol. Earth Syst. Sci.*, **15**, 3253–3274, doi:10.5194/hess-15-3253-2011.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Demargne, J., and Coauthors, 2014: The science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bull. Amer. Meteor. Soc.*, **95**, 79–98, doi:10.1175/BAMS-D-12-00081.1.

Dong, L., L. Xiong, and K. Yu, 2013: Uncertainty analysis of multiple hydrologic models using the Bayesian model averaging method. *J. Appl. Math.*, **2013**, 346045, doi:10.1155/2013/346045.

Donnelly, C., J. C. M. Andersson, and B. Arheimer, 2015: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol. Sci. J.*, **61**, 255–273, doi:10.1080/02626667.2015.1027710.

Dutra, E., C. Schär, P. Viterbo, and P. M. A. Miranda, 2011: Land–atmosphere coupling associated with snow cover. *Geophys. Res. Lett.*, **38**, L15707, doi:10.1029/2011GL048435.

Emerton, R. E., and Coauthors, 2016: Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev.: Water*, **3**, 391–418, doi:10.1002/wat2.1137.

Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, **138**, 190–202, doi:10.1175/2009MWR3046.1.

Gneiting, T., and M. Katzfuss, 2014: Probabilistic forecasting. *Annu. Rev. Stat. Appl.*, **1**, 125–151, doi:10.1146/annurev-statistics-062713-085831.

Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeor.*, **12**, 869–884, doi:10.1175/2011JHM1324.1.

Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1814–1827, doi:10.1002/qj.1895.

He, Y., F. Wetterhall, H. L. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor, 2009: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteor. Appl.*, **16**, 91–101, doi:10.1002/met.132.

——, and Coauthors, 2010: Ensemble forecasting using TIGGE for the July–September 2008 floods in the Upper Huai catchment: A case study. *Atmos. Sci. Lett.*, **11**, 132–138, doi:10.1002/asl.270.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:10.2151/jmsj.2015-001.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550, doi:10.1126/science.285.5433.1548.

Krzysztofowicz, R., 1997: Transformation and normalization of variates with specified distributions. *J. Hydrol.*, **197**, 286–292, doi:10.1016/S0022-1694(96)03276-3.

Liang, Z., D. Wang, Y. Guo, Y. Zhang, and R. Dai, 2013: Application of Bayesian model averaging approach to multimodel ensemble hydrologic forecasting. *J. Hydrol. Eng.*, **18**, 1426–1436, doi:10.1061/(ASCE)HE.1943-5584.0000493.

Liu, Y., Q. Duan, L. Zhao, A. Ye, Y. Tao, C. Miao, X. Mu, and J. C. Schaake, 2013: Evaluating the predictive skill of postprocessed NCEP GFS ensemble precipitation forecasts in China's Huai River basin. *Hydrol. Processes*, **27**, 57–74, doi:10.1002/hyp.9496.

Olsson, J., and G. Lindström, 2008: Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *J. Hydrol.*, **350**, 14–24, doi:10.1016/j.jhydrol.2007.11.010.

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo, 2008: New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophys. Res. Lett.*, **35**, L10404, doi:10.1029/2008GL033837.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, doi:10.1175/MWR2906.1.

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Schefzik, R., L. T. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, doi:10.1214/13-STS443.

Thielen, J., J. Bartholmes, M. H. Ramos, and A. de Roo, 2009: The European Flood Alert System—Part 1: Concept and development. *Hydrol. Earth Syst. Sci.*, **13**, 125–140, doi:10.5194/hess-13-125-2009.

Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manage.*, **6**, 123–137, doi:10.1080/15715124.2008.9635342.

Van Der Knijff, J. M., J. M. J. Younis, and A. P. J. De Roo, 2010: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.*, **24**, 189–212, doi:10.1080/13658810802549154.

Velázquez, J. A., F. Anctil, M. H. Ramos, and C. Perrin, 2011: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Adv. Geosci.*, **29**, 33–42, doi:10.5194/adgeo-29-33-2011.

Vrugt, J. A., and B. A. Robinson, 2007: Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.*, **43**, W01411, doi:10.1029/2005WR004838.

Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, **47**, W04501, doi:10.1029/2010WR009726.

Zsótér, E., 2006: Recent developments in extreme weather forecasting. *ECMWF Newsletter*, ECMWF, Reading, United Kingdom, 8–17. [Available online at http://www.ecmwf.int/sites/default/files/elibrary/2006/14618-newsletter-no107-spring-2006.pdf.]

A5: Can seasonal hydrological forecasts inform local decisions and actions? A decision-making activity

This paper presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

Neumann, J. L., L. Arnal, **R. E. Emerton,** H. Griffith, S. Hyslop, S. Theofanidi and H. L. Cloke, 2018: Can seasonal hydrological forecasts inform local decisions and actions? A decision-making activity, *Geoscience Communication*, **1**, 35-57, doi:10.5194/gc-1-35-2018[*]

R.E. helped to run the decision-making activity during the focus group, recorded discussions and observations throughout the activity, and commented on the manuscript before and during publication.

Can seasonal hydrological forecasts inform local decisions and actions? A decision-making activity

Jessica L. Neumann[1], **Louise Arnal**[1,2], **Rebecca E. Emerton**[1,2], **Helen Griffith**[1], **Stuart Hyslop**[3], **Sofia Theofanidi**[1], and **Hannah L. Cloke**[1,4,5]

[1]Department of Geography and Environmental Science, University of Reading, Reading, UK
[2]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
[3]Environment Agency, Kings Meadow House, Reading, UK
[4]Department of Meteorology, University of Reading, Reading, UK
[5]Department of Earth Sciences, Uppsala, Sweden

Correspondence: Jessica L. Neumann (j.l.neumann@reading.ac.uk)

Abstract. While this paper has a hydrological focus (a glossary of terms highlighted by asterisks in the text is included in Appendix A), the concept of our decision-making activity will be of wider interest and applicable to those involved in all aspects of geoscience communication.

Seasonal hydrological forecasts (SHF) provide insight into the river and groundwater levels that might be expected over the coming months. This is valuable for informing future flood or drought risk and water availability, yet studies investigating how SHF are used for decision-making are limited. Our activity was designed to capture how different water sector users, broadly flood and drought forecasters, water resource managers, and groundwater hydrologists, interpret and act on SHF to inform decisions in the West Thames, UK. Using a combination of operational and hypothetical forecasts, participants were provided with three sets of progressively confident and locally tailored SHF for a flood event in 3 months' time. Participants played with their "day-job" hat on and were not informed whether the SHF represented a flood, drought, or business-as-usual scenario. Participants increased their decision/action choice in response to more confident and locally tailored forecasts. Forecasters and groundwater hydrologists were most likely to request further information about the situation, inform other organizations, and implement actions for preparedness. Water resource managers more consistently adopted a "watch and wait" approach. Local knowledge, risk appetite, and experience of previous flood events were important for informing decisions. Discussions highlighted that forecast uncertainty does not necessarily pose a barrier to use, but SHF need to be presented at a finer spatial resolution to aid local decision-making. SHF information that is visualized using combinations of maps, text, hydrographs, and tables is beneficial for interpretation, and better communication of SHF that are tailored to different user groups is needed. Decision-making activities are a great way of creating realistic scenarios that participants can identify with whilst allowing the activity creators to observe different thought processes. In this case, participants stated that the activity complemented their everyday work, introduced them to ongoing scientific developments, and enhanced their understanding of how different organizations are engaging with and using SHF to aid decision-making across the West Thames.

1 Introduction

There has been a recent shift away from the conventional linear model of science, where research is carried out within the scientific community with the expectation that users will be able to access and apply the information, towards co-production and stakeholder-led initiatives that bring together scientists and decision-makers to frame and deliver "actionable research" (Asrar et al., 2012; Lemos et al., 2012; Meadow et al., 2015). Regular and clear communication between scientists and policy-makers and practitioners in

workshops, focus groups, consultations, and interviews, and through the development of games, activities, and interactive media, is imperative for ensuring that projects deliver impact outside of the academic environment. Here, we share findings from an activity that explored the use of seasonal hydrological forecasts* for local decision-making. This was conducted as part of an IMPREX (IMproving PRedictions and management of hydrological Extremes) stakeholder focus group for the West Thames, UK (van den Hurk et al., 2016; IMPREX, 2018a), co-organized by the University of Reading (UoR), UK, Environment Agency (EA) and supported by the European Centre for Medium-Range Weather Forecasts (ECMWF).

Seasonal hydrological forecasts (SHF) have the ability to predict principal changes in the hydrological environment such as river flows and groundwater levels weeks or months in advance. This has the potential to benefit humanitarian action and economic decision-making, e.g. to provide early warning of potential flood and drought events, assist with water quality monitoring, and ensure optimal management and use of water resources for public water supply, agriculture, and industry (Chiew et al., 2003; Arnal et al., 2017; Li et al., 2017; Meißner et al., 2017; Turner et al., 2017). SHF systems covering a range of spatial scales have been developed – Hydrological Outlook UK forecasts at a national level (Prudhomme et al., 2017; CEH, 2018) – while the Copernicus European and Global Flood Awareness Systems (EFAS and GloFAS) provide operational forecasts over larger scales (JRC, 2018a, b). Recent research has demonstrated improvements in SHF quality*, including increased accuracy out to 4 months for high-flow events during the winter in Europe (Arnal et al., 2018; Emerton et al., 2018).

There is growing interest in SHF amongst policy-makers and practitioners; however, in many cases, there is limited information about whether SHF products are *actually* being used. Research output has focused largely on technical system development and improvements to forecast skill* (see the review by Yuan et al., 2015), with relatively fewer studies exploring how users engage with and apply SHF to inform decisions (see Crochemore et al., 2015; Viel et al., 2016). Many seasonal forecasting studies, including those investigating the application of seasonal meteorological forecasts* (which provide information about future weather variables, rather than hydrology more specifically), have identified forecast uncertainty*, whereby forecast skill and sharpness* decrease with increasing lead time* (Wood and Lettenmaier, 2008; Soares and Dessai, 2015), and how this uncertainty can be communicated effectively as key barriers to use (Arnal et al., 2016; Vaughan et al., 2016). Non-technical factors, including the level of knowledge and training required to interpret and apply SHF information effectively (Bolson et al., 2013; Soares and Dessai, 2016), the visualization, format, and compatibility of the information provided (Fry et al., 2017; Soares et al., 2018), and the level of communication between different users in the water sector and between

research developers and practitioners (Golding et al., 2017), have all been found to act as both barriers and enablers, depending on the user group in question.

The potential for SHF to meet the needs of the water sector is recognized by a host of UK environmental organizations, including the EA, the Met Office, and research centres (see Prudhomme et al., 2017). The West Thames specifically is underlain by a slowly responding, largely groundwater-driven hydrogeological system (Mackay et al., 2015), meaning that there is potential for extreme hydrological events such as the drought of 2010–2012 (Bell et al., 2013) and winter floods of 2013–2014 (Neumann et al., 2018) to be detected weeks or months in advance. It also has a dense population and high demands for water which require effective long-term management of resources for public drinking supply, industry, agriculture, and wastewater treatment (further details about the West Thames can be found in Sect. 2.2). The value of using SHF in the West Thames is of particular interest to the EA; however, information on the level of understanding, uptake, and application is currently unknown. We therefore aimed to develop a clearer understanding about how different professional water sector users – broadly forecasters, groundwater hydrologists, and water resource managers – are currently engaging with SHF in the West Thames using a decision-making activity.

In the context of flood science communication with experts, real-time activities such as simulation exercises (that imitate real-world processes and behaviours) or roleplay (where participants engage with real-world scenarios but take on personas and positionalities that differ from their own) are known to be effective when engaging with stakeholders who bring a range of scientific ideas and perspectives to the table (McEwen et al., 2014). Such activities encourage participants to apply their knowledge to realistic situations and to reflect on issues and the perspectives of other stakeholders (Pavey and Donoghue, 2003, p. 7). They are also valuable for understanding decision-making processes, e.g. for environmental hazards and conflicting community views (Harrison, 2002), for capacity building in response to new water legislation (Farolfi et al., 2004), and for understanding climate forecasts and decision-making (Ishikawa et al., 2011). Our decision-making activity provided an interactive and entertaining platform that encouraged participants to engage with real-world scenarios whilst fostering discussions about the barriers and enablers to use of SHF. Using three activity stages, participants were provided with sets of progressively confident and locally tailored SHF for the next 3 to 4 months. The SHF were produced using output from operational systems including Hydrological Outlook UK and the European Flood Awareness System (EFAS), and hypothetical forecasts generated through scientific research (see Neumann et al., 2018). Participants were asked to play in real time, i.e. as if receiving the forecasts on the day for the next 3 to 4 months. They did not know in advance whether the SHF represented a flood, drought, or business-as-usual scenario

and had to use their knowledge and experiences to make informed decisions based on the maps, hydrographs*, tables, and text provided. In reality, all three sets of SHF represented the same time period: winter 2013–2014 (a period of extensive flooding nationwide that occurred at the end of 2 years of drought conditions in the UK). Between December 2013 and February 2014 the West Thames experienced extreme flooding from fluvial and groundwater sources which had knock-on impacts for local water quality, sewage treatment, and water resource management – opening up discussions for all participants. Given that issues relating to flood and drought risk, water quality, and water resource management in the West Thames are generally managed by local and regional-area authorities (Thames Water, 2010), the activity focused on whether SHF can be used to support decision-making at the local level. To the best of our knowledge, this scale of practical application has yet to be explored, we suspect mainly due to the lower skill of seasonal meteorological forecasts in Europe, particularly with respect to precipitation, which is a key variable of interest for hydrology (Arribas et al., 2010; Doblas-Reyes et al., 2013). A brief overview of the focus group is provided in Sect. 2, the full activity set-up is detailed in Sect. 3, and the findings and the discussion are presented in Sects. 4 and 5.

2 Overview of the focus group

2.1 Aims of the focus group

The focus group was developed in collaboration with the EA and in line with the objectives of the IMPREX project. The aims were the following.

- Introduce and discuss current SHF projects, products, and initiatives for the UK and Europe.

- Engage with participants' experiences and knowledge of using SHF.

- Learn how SHF are being applied in the West Thames and recognize how different users in the water sector approach and apply SHF information for decision-making.

- Identify limitations and barriers to use.

- Identify future opportunities for SHF application and research.

These aims were delivered through a series of four interactive sessions designed to actively engage participants to share their knowledge and experiences of SHF, and short presentations that introduced the main topics surrounding SHF and informed participants about current SHF projects and developments in the scientific research. While this paper focuses on the decision-making activity (interactive session 2), discussions from the other sessions are also presented where

relevant. An outline of the focus group programme is provided in Supplement 1 and a full report of the activities is available; see Neumann et al. (2017).

2.2 The West Thames in southern England

2.2.1 Physical geography

The West Thames refers to the non-tidal portion of the Thames River Basin*, from its source in the Cotswolds in the west of England to 230 km downstream at Teddington Lock in western London (Fig. 1). It covers an area of 9857 km^2 (the Thames basin is 16 980 km^2) and comprises 10 river catchments* that are the tributaries* that feed directly into the River Thames (Fig. 1). The western catchments are predominantly rural; land use is a mix of agriculture and woodland with rolling hills and wide, flat floodplains (elevation up to 350 m a.s.l.). Towards the centre and east, the region becomes increasingly urbanized, encompassing the towns of Reading and Slough and outskirts of Greater London (elevation 4 m a.s.l. at Teddington Lock). Lithology* varies markedly across the West Thames. Catchments overlaying the Cotswolds (upstream) and the Chilterns (middle sections) are dominated by chalk and limestone aquifers* with high baseflow*, while a band of less-permeable clays and mudstones separates these two areas. Sandstones, mudstones, and clays are also prevalent towards London (downstream) – these catchments have higher levels of surface runoff* and can exhibit a flashier* response to storm events (Bloomfield et al., 2011; EA, 2009).

2.2.2 Water demands, risk, and management – why the West Thames is of interest

The West Thames is a highly pressured environment – 15 million people and a substantial part of the UK's economy rely directly on its water supply (EA, 2015). There are more than 2000 licensed abstraction points in the chalk aquifers and superficial alluvium and river terrace gravel deposits; 90 % of abstractions are for public water supply, the rest providing water for agriculture, aquaculture, and industry (Thames Water, 2010). There are 12 000 registered wastewater discharge points; pollution from sewage treatment works, transport, and urban areas affects more than 45 % of rivers, water bodies, and aquifers, largely towards London. Diffuse pollution and sedimentation from agricultural and forestry practice are the main contributors to poor water quality in the upper catchments, especially during times of high rainfall (EA, 2015).

Urbanization and land-use change in combination with more varied rainfall patterns have seen the region affected by a number of extreme drought and flood events in recent years (EA, 2009; Parry et al., 2015; Muchan et al., 2015). Across the Thames Basin, 200 000 properties are at risk from a 1 : 100*-year fluvial flood, with 10 000 at risk from a 1 : 5*-year event (EA, 2009). Low and high river flows also pose

Figure 1. Location and lithology of the West Thames and its 10 main river catchments.

risks to navigation and management of the canal network which is highly important for recreation, local living, and the economy (Wells and Davis, 2016).

2.3 Participants

2.3.1 Who took part?

SHF have the potential for wide-ranging application and it was important to capture the different perspectives of the West Thames water sector. The organizers agreed that the focus group would work well with a relatively small number of participants (up to 12) so that all perspectives could be heard. Based on discussions held between the organizers, individuals from local organizations working in established (i.e. long-term/permanent/leadership) roles relevant to SHF in the West Thames were invited; many but not all participants had previously collaborated with the University of Reading and/or EA. In some cases, an invitee was unable to attend due to prior commitments or because they had a colleague who they felt would be a better fit for the focus group. A total of 17 participants were invited from six organizations – 12 accepted and 11 took part on the day. They were responsible for flood and drought forecasting (F × 3), groundwater modelling and hydrogeology (GH × 2), navigation (N × 1), water resource and reservoir management (WR × 2), public water supply (WS × 2), and wastewater modelling and operations (WW × 1). They represented five organizations: two non-departmental public bodies (sponsored by government agencies), two science and research centres, one water service company, and one non-for-profit organization (Table 1).

2.3.2 Current engagement with SHF

By inviting local stakeholders we ensured that participants represented a range of different water sector personas and were familiar with the West Thames environment. We did not assume that participants had any prior knowledge of SHF and invitees were encouraged to attend even if they were unfamiliar with the concept as this would be an important indicator of the state of play in the West Thames (invite poster; see Supplement 1).

All 11 focus group participants were familiar with the concept of seasonal hydrological forecasting and 10 regularly used SHF in their everyday job (according to results from interactive session 1 – "What are seasonal hydrological forecasts?"). Using post-its, participants noted that Hydrological Outlook UK (CEH, 2018) and the associated raw forecasts from the analogue, hydrological, and meteorological models (produced by the UK Met Office, Centre for Ecology and Hydrology, British Geological Survey, EA, Natural Resources Wales, Scottish Environment Protection Agency, and Rivers Agency Northern Ireland) were the main sources of SHF information currently being used, primarily for flood and drought outlook, groundwater monitoring, and river flow projection purposes. Scientific research, operational planning, and sharing of information with other organizations in the water sector were also listed as reasons for engaging with SHF. It is important to note that no prior definitions or information were provided and no restrictions or guidance were placed on what participants should write down. This suggests that many in the water sector are using SHF to obtain an insight into whether the upcoming season will be drier or wetter than normal, but that they also believe SHF *potentially*

Table 1. Breakdown of participants who took part in the activity.

Job title	Organization type	Role in the activity
Modelling and Forecasting Team Leader	Public body/government agency (1)	Flood and drought forecaster
Chief Hydrometeorologist	Public body/government agency (2)	Flood and drought forecaster
Climate Scientist (Professor)	Science and research centre (1)	Flood and drought forecaster
Thames Water Resources Technical Specialist	Public body/government agency (1)	Groundwater modelling and hydrogeology
Groundwater Research Directorate	Science and research centre (2)	Groundwater modelling and hydrogeology
Principal Hydrologist for Water Management	Not-for-profit (charitable trust)	Navigation
Water Resources, Environment and Business Directorate	Public body/government agency (1)	Water resource and reservoir management
Abstraction and Transfers Analyst	Water service company	Water resource and reservoir management
Water Strategy and Resources Modeller	Water service company	Public water supply
Thames Region Hydrologist	Public body/government agency (1)	Public water supply
Wastewater Modelling Specialist	Water service company	Wastewater modelling and operations

have the capability to forecast possible flood and drought risk, which could be used to support decision-making and provide better preparedness. This is an encouraging starting point, although many participants noted that this potential is not currently being realized due to the uncertainty and coarse spatio-temporal resolution of SHF; e.g. Hydrological Outlook UK forecasts are only published monthly for the main UK river basins.

3 Set-up of the decision-making activity

3.1 Background

Our activity was inspired by the success of previous decision-making activities and games run by the HEPEX (Hydrological Ensemble Prediction EXperiment) community (e.g. Ramos et al., 2013; Crochemore et al., 2015; Arnal et al., 2016). The aim was to better understand how different water sector users in the West Thames interpret and act on SHF by providing them with hydrological context, maps, and forecasts for the region. The activity was designed for the West Thames so that we could capture the relationship between local stakeholders and the environment in which they work.

3.2 Activity design

3.2.1 Overview of the set-up

The set-up of the activity (illustrated in Fig. 2) had the following structure: Choose groups > Define the Objectives > Background Context > Stage 1 > Stage 2 > Stage 3.

Participants divided themselves into three groups based on their area of expertise and where they felt they could best contribute to the discussions. There were three flood and drought "forecasters" and two "groundwater hydrologists". The remaining participants (navigation, water resource and reservoir management, public water supply and wastewater operations) grouped themselves as "water resource managers". While the results and discussions focus on these three broad groups, individual perspectives are also included to

capture the variety of water sector personas present. There were also three research facilitators and three note-takers whose role it was to capture and record the key discussion points.

Groups were first provided with background context to the West Thames to set the scene, followed by three sets of progressively confident SHF for the next 3 to 4 months (Stages 1–3). Stage 1 forecasts were from Hydrological Outlook UK, Stage 2 were from EFAS-Seasonal (European Flood Awareness System) and Stage 3 were "improved" output from EFAS-Seasonal (Fig. 2 and Sect. 3.4). Participants were asked to discuss the information presented in their groups and make informed decisions about each of the 10 West Thames catchments (Fig. 1 and Sect. 3.3.2). All groups were provided with exactly the same information and discussion was encouraged. The activity took around 2 h and timings were only loosely controlled.

SHF at all three stages of the activity represented the same time period – dating from 1 November 2013 to 28 February 2014 (or 31 January 2014 for Hydrological Outlook UK, which only extends to 3 months; CEH, 2018). These dates captured a period of severe and widespread river and groundwater flooding in the West Thames (Huntingford et al., 2014; Kendon and McCarthy, 2015; Muchan et al., 2015). *Participants did not know the dates of the forecasts, nor were they informed whether the situation being forecasted was a high flow (flood), low flow (drought) or a business-as-usual scenario.* Dates were removed from all information, and streamflow- and groundwater-level units were removed from the Stage 2 and Stage 3 EFAS hydrographs, although exceedance thresholds were provided for context. The decision to remove units was advised by the EA. The concern was that participants familiar with average and high-flow values for specific catchments would deduce that the SHF must represent the 2013–2014 floods, which would bias their decision-making based on their previous experience and memories. No information on forecast skill or quality was given and participants were asked to treat all information as
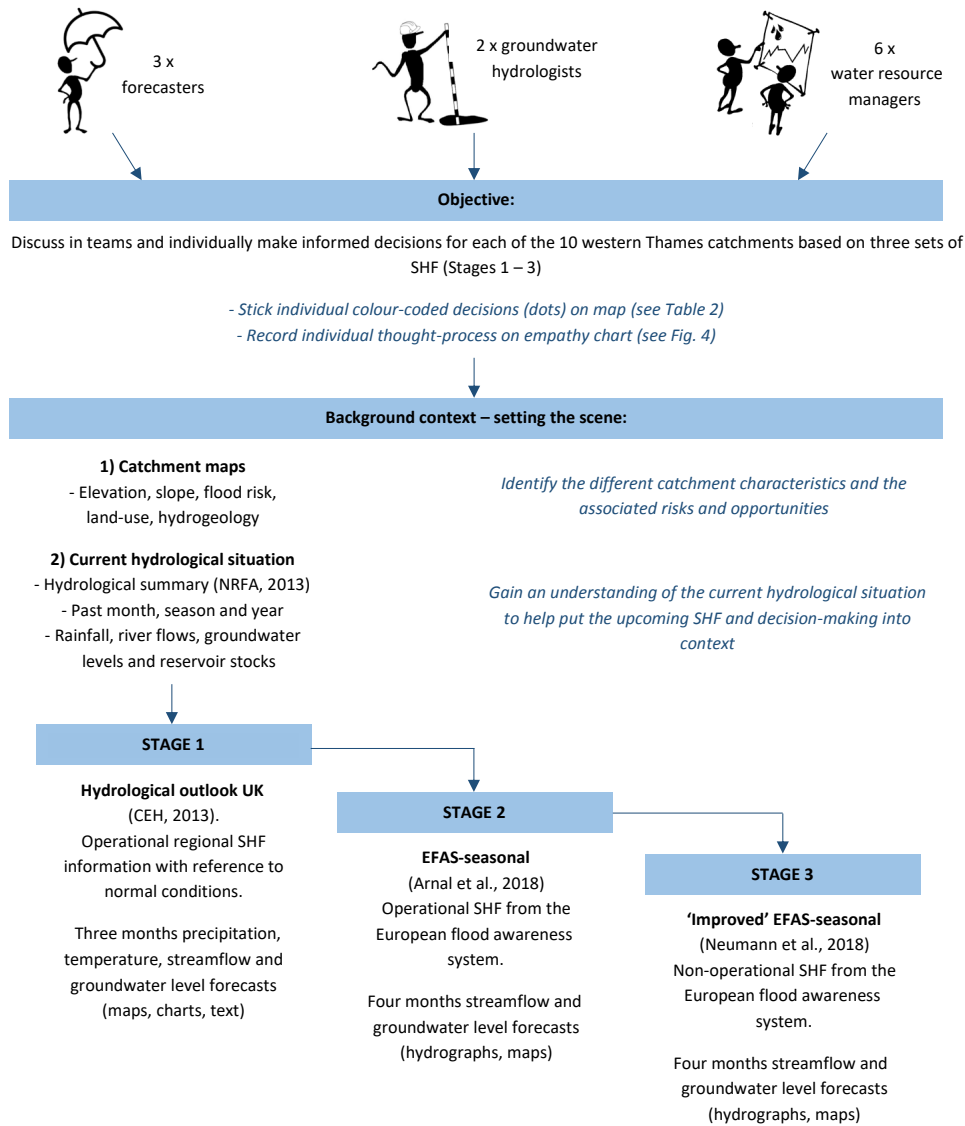
Figure 2. Set-up of the activity.

being "current", i.e. as if receiving the SHF today, for the next 3–4 months to create a realistic forecasting scenario.

3.2.2 Recording the decisions

In real life, a user's decision process can encompass a range of possible actions and associated consequences (Crochemore et al., 2015). Decisions can be controlled by providing participants with a set of options to choose from, e.g. to deploy temporary flood defences or not – the consequences of which usually determine the outcome of a game or activity. In this case, participants were asked to select from a broad range of colour-coded options (Table 2), but specific decisions were not defined as these had the potential to differ greatly between participants and might prompt unreal-

istic answers. At each stage, the colour-coded options were discussed by the three groups, simulating conversations that could happen in real life, but it was stressed that *the colour chosen was to be representative of what an individual participant, or their organization, would do with the SHF information in each catchment*. This was recorded on an A1 map using coloured sticky dots marked with the participant's initials ($n \sim 110$ dots per map (11 participants, 10 catchments)) (Fig. 3). In cases where participants were not familiar with all catchments, or did not feel able to make an informed decision, they did not place a dot. It was important to gather a written record explaining how and why the decisions were reached, and so participants were also asked to complete an A4 empathy map at each stage (Fig. 4). Originally designed as a collaborative tool to be used in business and marketing,
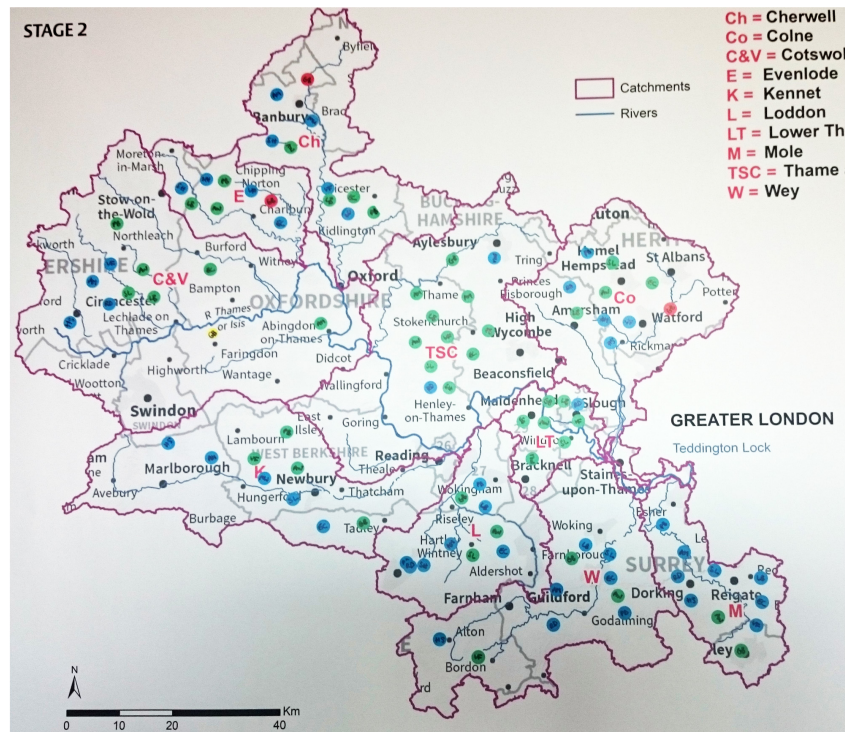
Figure 3. Participants' individual colour-coded decisions recorded on an A1 map.

Table 2. Colour codes and corresponding action or decision to be taken.

	Decision to be made or action to be taken
⚫	Ignore the SHF information: wait for the more skilful forecasts with shorter lead times (e.g. a 7–10-day forecast).
🔵	Look at the SHF information: decide there is no notable risk and do nothing at this point.
🟢	Look at the SHF information: discuss or pass the information on to relevant colleagues/departments in your organization and agree to keep an eye on the situation.
🟡	Look at the SHF information: discuss or pass the information on to relevant colleagues/departments in your organization *but also* external partners – actively request further information about the situation or seek advice on possible actions.
🔴	Look at the SHF information: decide to implement or set in motion action(s) in a catchment, e.g. to help with drought preparedness, early warning, repairs, or maintenance to flood defences.

empathy maps aim to gain a deeper understanding about an external user's experiences and decisions (Gray, 2017). Here, we adapted the traditional use by asking individuals to reflect on their own decisions based on their real-life experiences and discussions with other group members. This allowed us to capture individuals' thought processes, influences, discussions, and the potential risks and gains associated with their decision (Fig. 4). By combining the information recorded on

empathy maps for each group, we also gathered an overview of the shared understanding between forecasters, groundwater hydrologists, and water resource managers and how their SHF needs and expectations match and differ when it comes to decision-making.
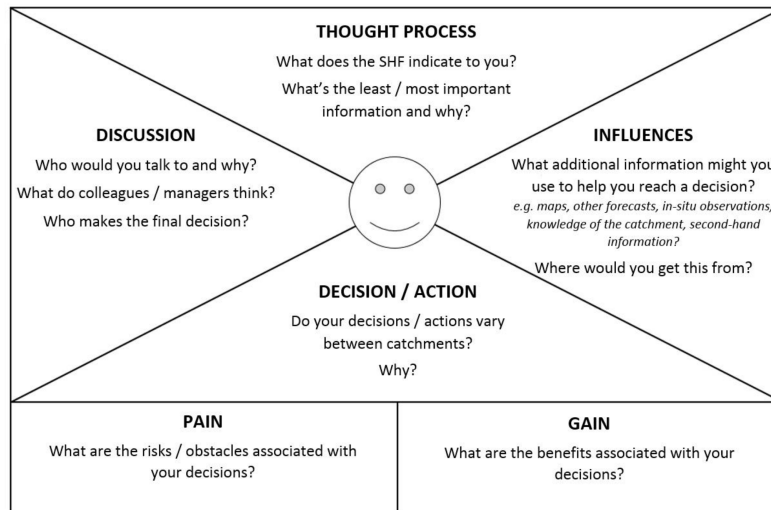
Figure 4. Empathy map completed by each participant during Stages 1–3.

3.3 Background context

Groups were given information about the West Thames catchment characteristics and "current" hydrological conditions (units and dates removed) to place the upcoming SHF into context and aid interpretation.

3.3.1 Catchment characteristics – driving factors, risks and opportunities

Five maps (Supplement 2) that provided a visual representation and a numerical breakdown of the characteristic differences between each catchment were given to participants.

- Hydrogeology* – dominant geological type (sandstone, chalk, clay)

- Elevation – minimum, maximum and mean elevation (m a.s.l.)

- Slope – minimum, maximum and standard deviation of slope angle (degrees)

- Land cover – dominant land use (urban, woodland, agricultural, semi-natural)

- Flood risk – flood warning and flood alert areas and an indication of "urban flood risk"

Participants were asked to discuss and identify the key differences between catchments and highlight the associated risks and opportunities. As some participants were more familiar with specific areas/catchments based on their day job, the maps provided a wider view of where catchment characteristics differ across the West Thames region.

3.3.2 Current hydrological situation

To help set the scene with respect to initial conditions, i.e. the "current" levels of water contained in the soil, groundwater, rivers, and reservoirs, groups were provided with information from the Hydrological Summary (NRFA, 2018) for the last month, past season, and past year (October 2013, June to September 2013, and November 2012 to October 2013 with dates removed). The Hydrological Summary (Supplement 3) focuses on rainfall, river flows, groundwater levels, and reservoir stocks and places the events of each month, and the conditions at the end of the month, into a historical context. In the real world, decision-makers are already prepared with this information; thus, providing evidence about whether hydrological conditions were wet, dry, or normal at the point of receiving the forecasts was an important piece of information for the participants to consider.

3.4 Activity Stages 1–3: the seasonal hydrological forecasts

3.4.1 Stage 1 – Hydrological Outlook UK

The first set of SHF information provided to participants was the Hydrological Outlook UK (from 1 November 2013 to 31 January 2014, with dates removed) (CEH, 2013). This provided regional information for the next 3 months with reference to normal conditions for precipitation, temperature, river flows and groundwater levels. Hydrological Outlook UK uses observations, ensemble models and expert judgement (CEH, 2018) to produce the seasonal forecasts. Information is publicly available and consists of text, graphs, tables and regional maps (examples are shown in Fig. 5 and the full set of forecasts provided to participants are in Supplement 4).
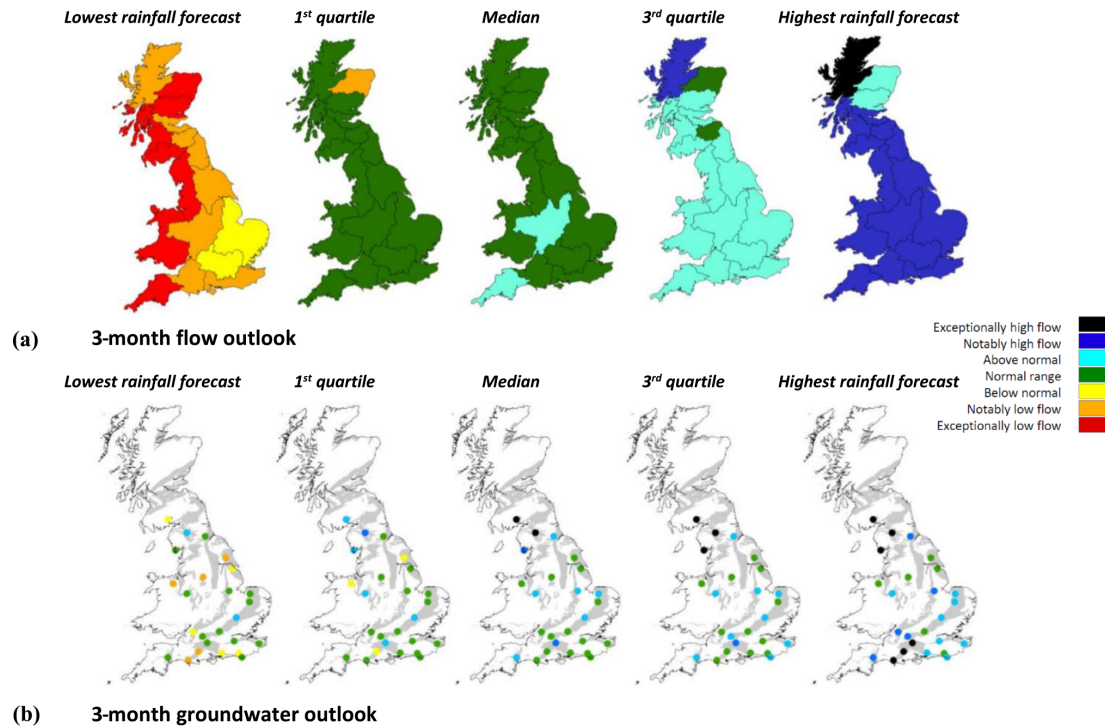
Figure 5. UK 3-month outlook maps from November 2013 (colours based on the percentile range of historical observed values). **(a)** Regional river flow forecasts created from climate forecasts. **(b)** Groundwater level forecasts at 25 UK boreholes created from climate forecasts (CEH, 2013).

3.4.2 Stage 2 – EFAS-Seasonal

EFAS-Seasonal (European Flood Awareness System) is an operational system that monitors and forecasts streamflow* across Europe, with the potential to predict higher than normal streamflow events up to 2 months ahead in an operational capacity, and up to 7 months in practice (JRC, 2018a; Arnal et al., 2018). It runs on a 5 km × 5 km grid and uses the LIS-FLOOD hydrological model (Van der Knijff et al., 2010; Alfieri et al., 2014). Seasonal ensemble* meteorological forecasts from the ECMWF's "System 4" operational meteorological forecasting system (Molteni et al., 2011) are used as input to LISFLOOD, from which seasonal ensemble hydrological forecasts are generated on the first day of each month (see Arnal et al., 2018, for details).

For the activity, SHF were produced from 1 November 2013 out to 4 months to focus on the period of extreme stormy weather and flooding experienced. As EFAS-Seasonal is designed to run at the scale of large river basins (i.e. the whole Thames basin), GIS shapefiles were used to extract forecast information for the 10 West Thames catchments using Python v3.5. This provided more locally tailored forecasts compared with Hydrological Outlook UK (Stage 1).

To ascertain whether participants had a preference for how SHF information is presented, the Stage 2 forecasts were presented as both hydrographs and choropleth* maps (Fig. 6).

Ensemble hydrographs for streamflow ($m^3 s^{-1}$) and groundwater levels (mm) indicated the predicted trajectory of the hydrological conditions for the next 4 months in each of the 10 catchments (n.b. the greater the spread, the more uncertain the forecast) (Fig. 6a). Units and dates were removed; however, exceedance thresholds*, based on daily observed streamflow and groundwater records between 1994 and 2014 for each of the catchments, were provided for context (EA, 2017; NRFA, 2017). Q50 (median) indicated average streamflow and groundwater conditions for the catchment. Q10 (90th percentile) indicated high streamflow/high groundwater level conditions – 90 % of all recorded observations over the previous 20-year period fell below this line.

The choropleth maps showed the maximum probability that the full forecast ensemble for a catchment exceeded the Q10 (90th percentile) threshold in a given month (Fig. 6b), thus providing a snapshot of the probability of potentially extreme conditions at catchment level. The full set of EFAS-Seasonal SHF provided to participants can be found in Supplement 5.

3.4.3 Stage 3 – "Improved" EFAS-Seasonal

Stage 3 followed the exact same set-up and provided the same style output (Fig. 7a, b) as Stage 2 – the only difference being that the seasonal meteorological forecasts used as input to LISFLOOD were taken from a set of atmospheric re-
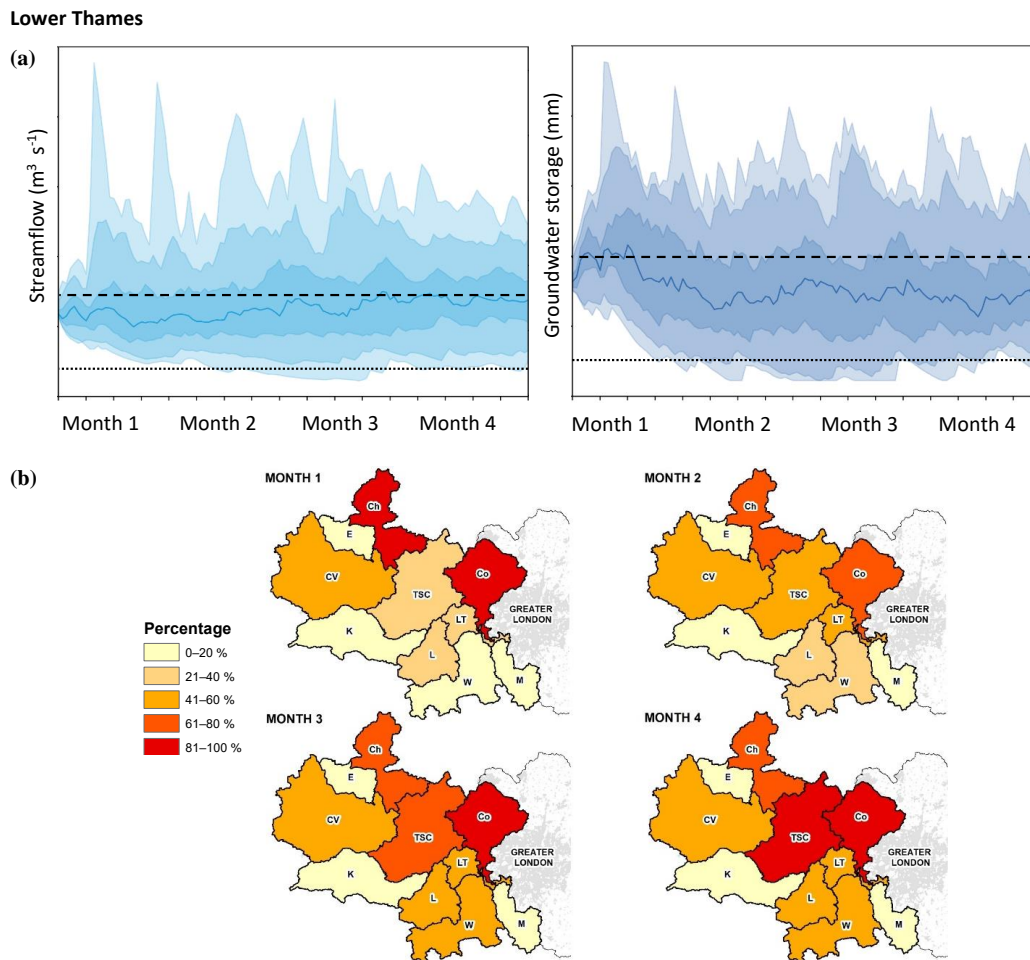
Figure 6. Four-month hydrological forecasts from EFAS-Seasonal (Stage 2). **(a)** Ensemble hydrographs for streamflow (light blue) and groundwater levels (dark blue) for the Lower Thames (LT) catchment. Exceedance thresholds (based on records from 1994 to 2014) are shown as Q10 (dashed line) and Q50 (dotted line). **(b)** Choropleth map shows the maximum probability that the full hydrograph ensemble for a catchment exceeds the Q10 streamflow threshold in a given month.

laxation experiments* conducted as part of a scientific study in the West Thames (see Neumann et al., 2018) rather than the operational seasonal meteorological forecasts from "System 4".

Atmospheric relaxation experiments were conducted by the ECMWF in late 2014 *after* the extreme weather and flooding (Rodwell et al., 2015). The aim was to recreate the atmospheric conditions that prevailed between November 2013 and February 2014, so that the ECWMF could better understand how weather anomalies across the globe contributed to the flooding experienced in the West Thames (Neumann et al., 2018). The SHF at Stage 3 represented near "perfect" forecasts as they were produced *once the floods had happened and the weather conditions were known*. The hydrographs are thus much sharper and more accurate than those presented to the participants at Stage 2 (Fig. 7, Supplement 6). It is important to note that this is not something that

can be achieved by operational systems currently, but does represent the theoretical upper level of forecast skill that may be available to water sector users in the future.

4 Results

4.1 Background context

4.1.1 Catchment differences – "hydrogeology is the driving factor of risks and opportunities"

All groups recognized spatial variability between the catchments and general consensus was that hydrogeology was the most important factor determining flood risk, drought risk, and water availability in the West Thames (Supplement 2). All groups were interested in the persistence, hydrological memory, and slower response of the groundwater-driven catchments upstream (e.g. the Evenlode, Thames, and
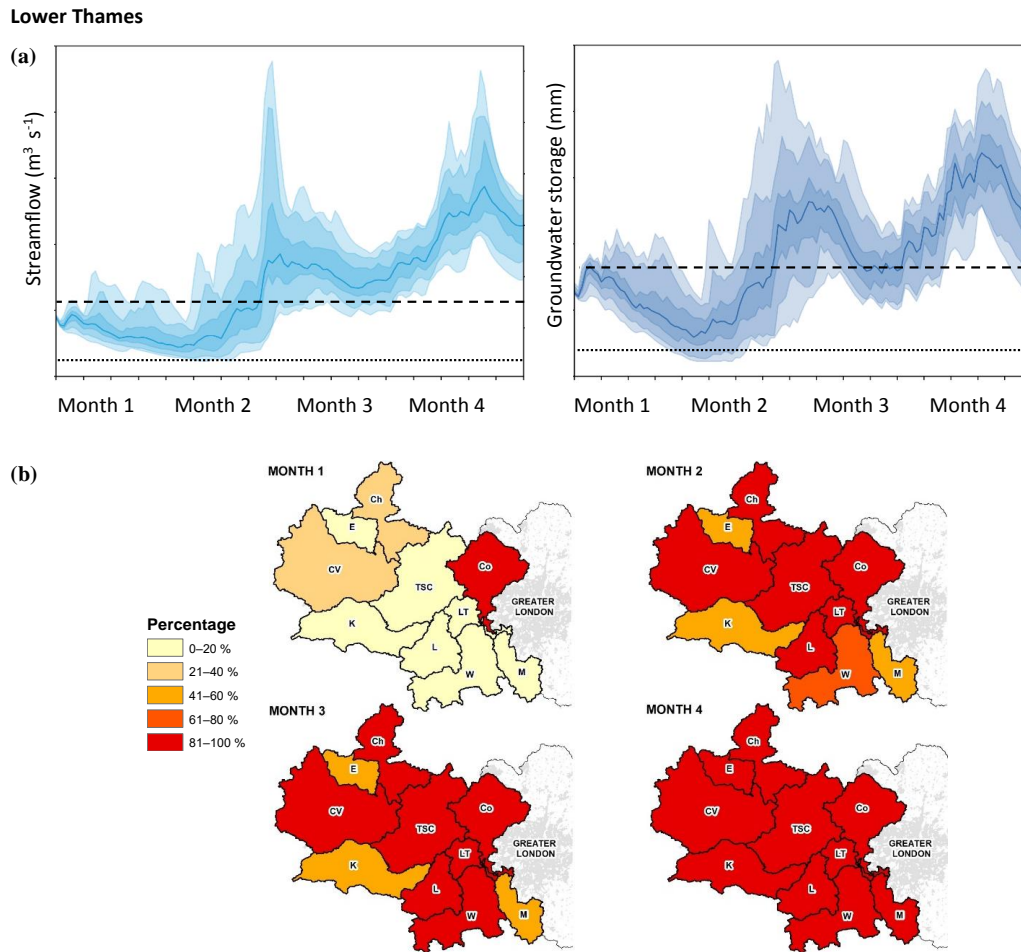
Figure 7. Four-month hydrological forecasts from the "Improved" EFAS-Seasonal (Stage 3). **(a)** Ensemble hydrographs for streamflow (light blue) and groundwater levels (dark blue) for the Lower Thames (LT) catchment. Exceedance thresholds (based on records from 1994 to 2014) are shown as Q10 (dashed line) and Q50 (dotted line). **(b)** Choropleth map shows the maximum probability that the full hydrograph ensemble for a catchment exceeds the Q10 streamflow threshold in a given month.

South Chilterns and Kennet) as these provided the greatest opportunity for water supply but also increased risk of local groundwater flooding and widespread fluvial flooding further downstream. Forecasters also highlighted the risks posed by impermeable catchments (e.g. the Cherwell and Lower Thames) that have a flashier response to rainfall. Water resource managers stated that upstream reservoirs were at increased risk of pollution (from agriculture), whilst dry weather (drought) was a greater issue towards London.

4.1.2 Current hydrological situation – "normal"

Hydrological Summary placed the "current" hydrological conditions for river flows, groundwater levels, and reservoir stocks within the "normal" range (Supplement 3). Maps indicated that rainfall was below average over the past season but above average the previous month. All groups were happy

with the current hydrological situation (no risks currently), although water resource managers stated that rainfall deficiency in the background should be kept in mind due to future drought potential.

4.2 Participant responses from Stages 1 to 3

The findings from each stage of the activity are presented below. At no point did participants ignore the SHF information (no black stickers were placed on the maps), which matched previous discussions about organizations' current use of SHF (Sect. 2.3.2). Colour-coded decisions made by all participants (calculated by counting the stickers on the A1 catchment maps) are represented as pie charts. An accompanying bar chart details the breakdown of choices made by each participant and their specific role in the water sector (Fig. 8a–c). Quotes and information in the text are taken from discussions

(a)

F = Flood and drought forecasters   WR = Water resource specialists
GH = Groundwater hydrologists   WS = Public water supply managers
N = Navigations officer   WW = Wastewater operations
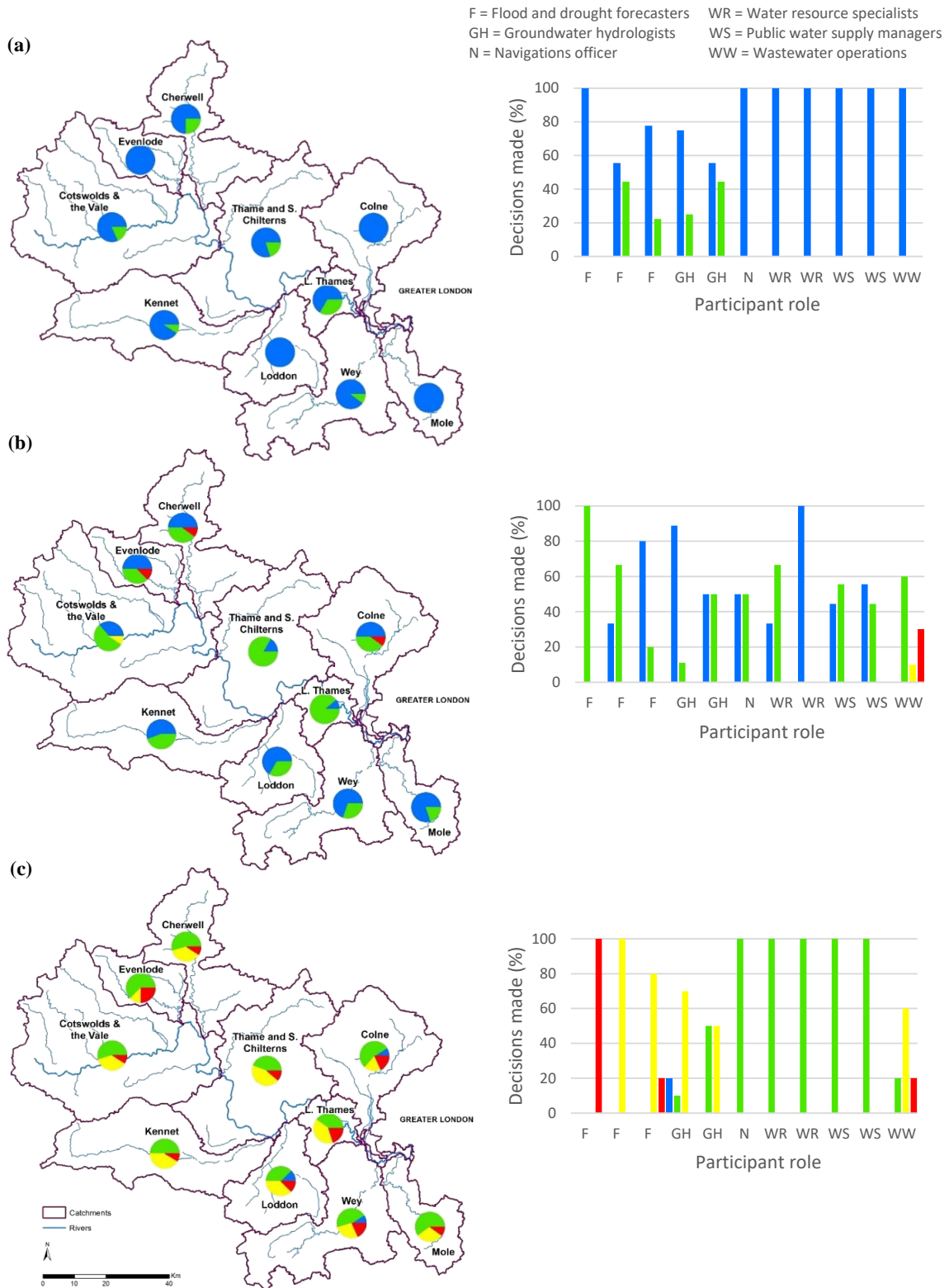

(b)

(c)

Figure 8. Summary of decisions and actions taken by different water sector personas based on **(a)** Hydrological Outlook UK; **(b)** EFAS-Seasonal; and **(c)** "Improved" EFAS-Seasonal. Blue – no notable risk; green – discuss internally; yellow – discuss externally and seek advice; red – implement action. Refer to Table 2 for full colour code descriptors.

recorded on the day and empathy maps – these are presented for the three groups (forecasters, groundwater hydrologists, and water resource managers).

4.2.1 Stage 1 – Hydrological Outlook UK

General consensus was for normal or above-normal conditions over the next 3 months; however, the information was "too vague to be actionable". Forecasters and groundwater hydrologists were more likely to discuss the situation with colleagues and keep an eye on the situation (green/blue), although there was some disagreement about the level of risk. Those involved in water resources, water supply, navigation, and wastewater operations (water resource managers) identified no risks requiring action (blue) (Fig. 8a).

Key statements:

"**Analogy with the summer 2007 floods**[*] suggests that **there's a risk that might be worth communicating internally**. Political influences e.g. known flooding hotspots might also be singled out for further engagement. However, there's not much evidence to divert from a normal pattern of preparedness."

[*]The UK suffered extensive flooding during June and July 2007 (the West Thames was flooded in late July). Thirteen people died and damages exceeded 3.2 billion GBP nationwide (Chatterton et al., 2010).

"**No major issues currently** but there is a **signal for rising groundwater levels**, potentially leading to flood risk – discuss with colleagues and keep an eye on borehole observations and new forecasts."

"Conditions are **favourable from a water resources perspective** – possibly heading more towards flood than drought conditions but currently **no notable risk and no concerns**. Discussions may arise during regular business briefings, but unlikely to be pursued unless changes are observed."

4.2.2 Stage 2 – EFAS-Seasonal

General consensus was for above-average streamflow and groundwater levels. Although the SHF provided more detail compared with Hydrological Outlook UK (Stage 1), clarity remained an issue. There was a general shift towards more internal communication (green), although actions were taken

by the wastewater operations manager in the water resource managers' group (yellow/red) (Fig. 8b).

Key statements:

"**Repeated rainfall events can lead to accumulated flood risk** in the Lower Thames and Thame and South Chilterns. Streamflow appears to convey more risk than groundwater levels. Would discuss in general terms with colleagues and internal decision-makers to avoid an over-reaction at senior level."

"A **moderate risk of groundwater flooding** (especially if the time period is for autumn – winter) but river flows do not appear to contribute much to groundwater risk at this stage and the forecasts are uncertain. Our **attention is focused on the chalk catchments and Thames gravels**; no direct actions are taken at the moment but we'd keep an eye on the situation and discuss at monthly meetings."

"**No significant concerns** from a water resources or navigation perspective however, there is **potential for localised flood risk which may impact on water supply and turbidity**. Not all catchments are affected so focus attention on Cotswolds and the Vale, Cherwell, Thame and South Chilterns and Colne where maps indicate high probability of Q10 exceedance. Discuss at internal briefings."

4.2.3 Stage 3 – "Improved" EFAS-Seasonal

General consensus was for confident forecasts that showed a high risk of streamflow and groundwater flooding in approximately 6 weeks' time. At this stage, forecasters and groundwater hydrologists were looking to verify the reliability and quality of the forecasts. Internal discussion and wider communication (green/yellow) were actively explored, although forecasters and groundwater hydrologists were still more likely to act on the information compared with water resource managers (Fig. 8c).

Key statements:

"Compared with our previous experiences of SHF these are very **sharp with a strong signal** and we would actively seek expert guidance as to the quality of the forecasts. If credible, our concern is that the signal is likely to **represent a nationwide flood risk** (not just the West Thames). **Low-consequence actions that deliver a measured message** should be implemented – e.g., identifying and locating resources and stocks, movement of temporary flood defences to high risk areas, completing projects, careful media release, strategic planning and staff briefing."

"There's **high probability of substantially exceeding the Q10 threshold**. Catchment characteristics are important to identify areas most at risk of groundwater flooding (chalk and gravels). **Drawing on previous experiences** we'd discuss the situation, obtain regular updates from partner organisations, use localised groundwater models to verify forecasts and consider communication via press release."
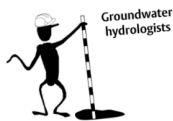
"These are **confident forecasts that give a good overview of magnitude and sequencing of possible flood events and subsequent knock-on effects to water quality**. Expect issues in 2–4 months so any actions taken would depend on how regularly forecasts are updated. We'd keep an eye on groundwater levels, hold internal briefings and discuss with groundwater team members to ensure they are kept informed and prepared. For navigation and wastewater operations where impacts can directly affect the public, we'd consider some open discussion with customers who will want to know how long an event might last."

5   Discussion

Our decision-making activity was designed to help understand how different water sector users engage with and act on SHF at a local level. The SHF for the three activity stages rep-

resented an extreme flood event between November 2013 and February 2014. There was clear evidence that more confident (sharper) and locally tailored forecasts led to increased levels of decision and action, although water sector users did not respond uniformly. Forecasters and groundwater hydrologists were most likely to inform other organizations, request further information about the situation, and implement action, while water resource managers more consistently adopted a "watch and wait" approach. In this section, the results are discussed in more detail and the findings are placed into the wider context of policy, practice, and next steps based on discussions captured during the focus group.

5.1   Operational SHF systems can support decision-making and uncertainty is expected

Throughout the focus group, participants expressed positively the potential for SHF to deliver better preparedness and early warning of flood and drought events, and the benefits associated with more consistent management of water resources, whilst recognizing that low skill and coarse resolution are current barriers to use (see also Soares and Dessai, 2015, 2016; Vaughan et al., 2016; Soares et al., 2018). These benefits and barriers were demonstrated during the activity as participants increased their level of decision-making in response to the more confident and locally tailored forecasts presented: Stage 1 Hydrological Outlook UK > Stage 2 EFAS-Seasonal > Stage 3 "Improved" EFAS-Seasonal.

Hydrological Outlook UK is the first operational SHF system for the UK and was the product that participants were most familiar with, likely due to its partnership set-up (Prudhomme et al., 2017). All groups indicated that the regional focus of the maps, i.e. the whole Thames basin, and lack of resolution and certainty as to the trajectory of the upcoming hydrological conditions, limited their ability to make informed decisions. No participants however ignored or dismissed the information despite there being no perceived risk. All agreed that on a day-to-day basis, Hydrological Outlook UK serves as a useful outlook tool when supplemented with additional sources of information including water situation reports (UK Gov, 2018) and other hydro-meteorological forecasts. As of 2017, exactly how the water sector uses Hydrological Outlook UK in practice had yet to be assessed (Bell et al., 2017), and here we provide a first step towards answering this question.

Stage 2 (EFAS-Seasonal) also represented an operational forecasting system designed to run at the scale of the whole Thames basin akin to Hydrological Outlook UK. The forecasts however were presented at a catchment level on a month-by-month basis to provide a more localized outlook. This finer spatio-temporal resolution allowed participants to supplement the SHF with their knowledge of local hydrogeology and other risk factors to identify those catchments where attention would likely be most needed. This led to increased levels of communication within organizations,

even though the overall hydrological outlook was very similar to that observed at Stage 1 (uncertain but with indication towards normal–high flows). The use of large-scale (regional or global) operational forecasting products that trigger worthwhile actions at the local level has been demonstrated at shorter lead times (e.g. Coughlan de Perez et al., 2016). While the development of higher-resolution seasonal meteorological forecasts and better representation of the coupled system and initial conditions are expected to lead to improvements in SHF (Lewis et al., 2015; Bell et al., 2017; Arnal et al., 2018), we pose the open question: do operational systems such as Hydrological Outlook UK *already* have the potential to support better communication and decision-making if they could be presented at a more local scale? This would require careful communication of the uncertainty, reliability, and skill of the forecast, and how to do this effectively is a topic of current interest in meteorological and hydrological forecasting (e.g. Ramos et al., 2013; Vaughan et al., 2016; Fry et al., 2017). Although communicating uncertainty was not a specific focus of our activity, one key message from the focus group was that "uncertainty is expected" with SHF and water sector users would engage with a local forecast, even if they chose not to act on it. As pointed out by Viel et al. (2016), "low skill" is not the same as "no skill", and SHF which may have minimal value from the perspective of a scientific researcher can sometimes elicit significant interest from the view of a water sector user who is familiar with the area. Importantly, it should also be noted that although no measures of forecast skill and quality were included in our activity, participants only expressed a need to verify the quality of the forecasts at Stage 3. In discussions as to why this was the case, the forecasters and groundwater hydrologists stated that holding internal briefings and increasing awareness of "at risk" catchments are suitable low-cost actions when dealing with SHF that indicate some degree of risk, even if the information is uncertain and unverified. At Stage 3, to obtain such confident SHF was well beyond current operational standards; thus, its reliability was questioned. Participants did agree however that even in the absence of information on forecast quality, a sharper, more confident forecast that indicated high potential flood risk would be more likely to provoke a response than a dispersive one, even if the maximum of the forecast ensemble indicated values of comparable magnitude in both cases.

5.2 Interactions with SHF are user-specific and should be tailored accordingly

The manner in which users approached and used SHF differed markedly depending on the perceived severity of the flood event; the responsibilities and risk appetite of an organization; and the local knowledge and experiences possessed by the individual (see also Kirchhoff et al., 2013; Golding et al., 2017). Forecasters and groundwater hydrologists displayed the lowest risk appetite, admitting that they

were likely to err on the side of caution to avoid negative media impacts, economic damages, and loss of trust by the public.

> "Analogy with the summer floods of 2007 . . . my previous experience makes me think that the risk is worth communicating. . . " – forecaster at Stage 1/2.

> "A much stronger and more coherent signal regarding river flows and groundwater levels, but the forecasts indicate that the potential impact isn't right now . . . we'll keep an eye on the situation" – water resource manager at Stage 3.

While a flood event is less of an immediate issue for water resource managers, secondary effects relating to closure of canals (navigation), turbidity, and sewer surcharge (wastewater operations) did invoke action where there was potential to impact on the public. Participants were notably proactive where they had had previous experience of extreme events, e.g. forecasters' analogies with the 2007 floods (Chatterton et al., 2010), or had been witness to poor management; e.g. the wastewater operations manager recognized high potential for groundwater flooding and sewer surcharge at 1 month's lead time in the Evenlode, Cherwell, and Colne (Fig. 7).

> "Based on previous operational issues, I'd advise pre-emptive actions such as the cleaning and maintenance of pumping stations for these catchments" – Wastewater operations manager at Stage 2/3.

This highlights the value of retaining institutional memory where possible (see also McEwen et al., 2012) and being aware of organizations' or individuals' pre-determined positions or perceived self-interests which may largely be founded on previous experiences (Ishikawa et al., 2011).

It is important to note that while this activity focused on a flood event, decisions made by the groups would almost certainly have differed if the SHF had indicated drought conditions. The impacts of drought have the potential to affect larger areas, for longer (Bloomfield and Marchant, 2013), notably with respect to agriculture (Li et al., 2017), reservoir management (Turner et al., 2017) and navigation (Meißner et al., 2017). The difference in response between water sector users supports the notion that tailoring SHF information to specific user groups will improve uptake and ability to inform decision-making (Jones et al., 2015; Lorenz et al., 2015; Vaughan et al., 2016; Soares et al., 2018), an area currently being explored by the IMPREX Risk Outlook (IMPREX, 2018b).

5.3 Communication is both a barrier and enabler to decision-making

Communication is one of the most frequently identified barriers when it comes to uptake and use of seasonal meteorological and hydrological forecasts (Soares and Dessai, 2015;

Vaughan et al., 2016; Golding et al., 2017; Soares et al., 2018). Discussions captured during the focus group and indicated on some empathy maps identified two key communication barriers in the West Thames: (1) between water sector users themselves and how they interpret and communicate SHF information and (2) a disconnect between scientists developing the forecasts and those involved in policy, practice and decision-making.

All groups said they felt better able to interpret and communicate the messages when presented with a range of complementary forms of SHF information including maps, hydrographs, and text, with maps being of particular value. This supports findings by Lorenz et al. (2015), who identified clear differences in users' comprehension of and preference for visualizations of climate information. Mapping information was also found to be important in the survey by Vaughan et al. (2016), while numerical representations were preferred over text and graphics in the study by Soares et al. (2018). Many participants said they would feel better prepared and able to discuss upcoming hydrological conditions if SHF information was visualized in a variety of ways and regular engagement was made a routine part of their job (see Sect. 5.4).

A number of participants also felt that scientific improvements and developments to SHF are not being adequately communicated to those involved in policy and practice. General consensus was that knowledge exchange events and information sharing services through projects such as IM-PREX are an excellent way of addressing this disconnect. Presentations during the focus group shared findings from other projects, including the European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales (EUPORIAS) (Met Office, 2018), the End-to-end Demonstrator for improved decision-making in the water sector in Europe (EDgE), Service for Water Indicators in Climate Change Adaptation (SWICCA) (Copernicus, 2017a, b), and Improving Predictions of Drought for User Decision Making (IMPETUS) (Prudhomme et al., 2015) – much of which was new knowledge to some participants. It was further expressed that stakeholder events yield maximum benefit for both the scientist and the user when they are co-produced with an organization that is involved in receiving, tailoring, and distributing SHF information (Rapley et al., 2014). Importantly, we do not want to be in the position whereby SHF skill has improved but the credibility and reliability of the information is questioned by decision-makers who have not been kept up to date with developments. The potential for this disconnect was demonstrated by both forecasters and groundwater hydrologists at Stage 3 ("Improved" EFAS-Seasonal) whereby decisions would only be made if the accuracy of the forecast could be verified.

> "Forecast signal is implausibly strong but, if valid, gives a clear signal for disturbed conditions"

> "Surprised at forecast and the strength of the signal... IF credible, then actions need to be taken"

> "Would definitely talk to the Environment Agency and search for other monitoring data to verify the forecast" – forecasters and groundwater hydrologists at Stage 3.

In this case, the SHF at Stage 3 were hypothetical and no information on forecast quality was given; however, the forecasts provided a good representation of what scientists hope to achieve with operational seasonal forecasting systems in the future (Neumann et al., 2018). This emphasizes the need to keep water sector users informed of scientific developments (see also Bolson et al., 2013), and to build awareness and knowledge around interpreting and using forecast quality information, as it is becoming more widely adopted in seasonal forecasting (see Copernicus, 2017a; Fry et al., 2017).

5.4 Implications for future policy and decision-making

The EA is the public body responsible for managing flood risk in the UK. They focus on maintaining a certain level of preparedness whilst recognizing that particular conditions and types of flooding/drought are more likely at different times of year. Currently, the EA use SHF predominantly as supporting information and rely on shorter-range forecasts for action. As co-developers of this focus group, the EA recognized the following points for future consideration.

1. To upskill and help staff interpret SHF information received.

2. To identify suitable low-consequence actions that could be taken based on SHF.

3. To move beyond the current position of using SHF for information only, to making conscious decisions as part of routine incident management strategies (relies on 1 and 2).

> "Regular review and discussion of extended outlooks (5–30 days) and the 1–3 months forecasts during weekly handover between the incoming and outgoing flood duty teams would improve familiarity of long range forecast products and dealing with the uncertainty that they present. This would be an excellent way of considering the possible conditions and the potential for disruption going forward." – EA activity co-developer.

In short, more engagement with SHF and improved clarity for easier interpretation by different users will ensure that SHF have a valuable role to play in future decision-making at the local scale.

5.5 Learning outcomes and future considerations

Encouragingly, we identified that SHF are being used, and participants agreed that the decision-making activity was an

entertaining platform for fostering discussions which complemented their everyday work and general understanding of SHF. From the participants' perspective, learning outcomes included knowing more about the ongoing scientific developments in SHF and a better understanding of how different organizations in the West Thames water sector are using SHF. Many also stated that the activity and focus group discussions enhanced their ability to think about possible decisions and actions that may be taken in the future. As the activity developers, we found that the group discussions stimulated participants' motivations and interests more so than would have been achieved by asking participants to engage on an individual basis. We also advocate the use of empathy maps or other forms of obtaining a written record of participant thought processes in addition to their decision choices.

Our activity was designed to provide a first insight into the current state of play regarding SHF in the West Thames. Although 11 participants was a small sample size, they represented an important and well-balanced mix of water sector decision-makers in the West Thames. The only exception was the agricultural sector, which could not attend, and thus it would be interesting to capture this perspective with ongoing research (e.g. Li et al., 2017). We also recognize the possibility that those who took part had a vested interest in SHF; however, we did encourage participants to attend even where they had no background knowledge or experience of SHF. Finally, we advocate that others conducting a similar activity may wish to consider whether participant interpretation can be subconsciously influenced by the information provided. For example, flood risk maps were provided as part of the background context, but may have inadvertently led participants to consider the upcoming forecasts with respect to high-flow events. Likewise, there is potential that the 3-month SHF (Stage 1) may have been interpreted differently to the 4-month forecasts (Stage 2 and Stage 3) and we do not know the degree to which individuals may have been swayed to place a particular colour on the map based on the conversations they had with their group members (and how big an influence such conversations play in real life). Discussions with the participants at the end of the activity with respect to these points would have been helpful.

6 Conclusions

Key findings were that engagement is user-specific and SHF have the potential to be more useful if they could be presented at a scale which matches that employed in decision-making. The ability to interpret messages is aided by complementary forms of SHF visualization that provide a wider overview of the upcoming hydrological outlook, with maps being of particular value. However, improved communication between scientists, providers, and users is required to ensure that users are kept up to date with developments. We conclude that the current level of understanding in the West Thames provides an excellent basis upon which to incorporate future developments of operational forecasts and for facilitating communication and decision-making between water sector partners.

Data availability. All data/graphs/information that were used by participants for the focus group activity are included in the Supplement. Individual participant results are not publicly available in order to protect anonymity. If readers require further information, this may be provided by contacting the corresponding author.

Appendix A: Glossary

Aquifer	underground layer of water-bearing permeable rock which can occur at various depths.
Atmospheric relaxation experiments	are used by meteorologists once an extreme weather event has happened. Put simply, when a seasonal forecast predicts the wrong weather, scientists "force" the conditions in the atmosphere so that they can try to recreate the extreme weather conditions and better understand what happened.
Baseflow	the portion of the river flow (streamflow) that is sustained between rainfall events and is fed into streams and rivers by delayed shallow subsurface flow. Not to be confused with "groundwater" which is water which has entered an aquifer, or "groundwater flow" where water enters a river having been in an aquifer.
Choropleth map	uses differences in shading, patterning or colouring in proportion to the value of a given variable in areas of interest.
Exceedance threshold	a user-defined threshold (e.g. 90 %) that is based on river flow or groundwater level observations (measurements) from the previous 20 years. E.g. if an exceedance threshold is set to the 90th percentile, this means that 90 % of all recorded observations over the past 20 years fell below this level.
Flashy	rivers and catchments that respond quickly to rainfall events.
Forecast ensemble	instead of running a single forecast (known as a deterministic forecast that has one outcome), computer models can run a forecast several times using slightly different starting conditions (to account for uncertainties in the forecasting process). The complete set of forecasts is referred to as the ensemble, and the individual forecasts are known as ensemble members. Each ensemble member represents a different possible scenario, and each scenario is equally likely to happen.
Forecast quality	the SHF is compared to, or verified against, a corresponding observation of what actually happened, or a good estimate of the true outcome. SHF quality describes the degree to which the forecast corresponds to what actually happened (see also "forecast skill").
Forecast sharpness	describes the spread or variability among the different ensemble members of a forecast (the different forecast values). The more concentrated (close together) the ensemble members are, the sharper the forecast is, and vice versa. Importantly, a forecast can be sharp even if it is wrong i.e. far from what actually happened. (See also "forecast ensemble".)
Forecast skill	the SHF quality can be compared to the quality of a benchmark or reference, usually another forecast. The relative quality of the SHF over this reference forecast is the SHF skill (see also "forecast quality").
Forecast uncertainty	the skill and accuracy of SHF tends to decrease with increasing lead time due to factors such as variations in weather conditions, how the hydrological model has been set-up to represent complex processes, and how well the hydrological model has captured the real-world hydrologic conditions at the time the forecast is started (e.g. how wet is the soil or how much water is currently in the river?). There is an element of uncertainty in all forecasts that can amplify with time. Ensemble forecasting is one way of representing forecast uncertainty. (See also "forecast ensemble".)
Hydrogeology	the area of geology that deals with the distribution and movement of below-ground water in the soil, rocks and aquifers.
Hydrograph	a graph showing how river and groundwater levels are expected to change over time at a specific location. Ensemble hydrographs show the full spread of the forecast ensemble.
Lead time	the length of time between when the SHF is started (initiated) and the occurrence of the phenomena (e.g. flood) being predicted. Can also be used to represent the point at which the SHF is started and the beginning of the forecast validity period (e.g. from 3 weeks).
Lithology	the general physical characteristics of rocks.
River basin	the largest and total area of land drained by a major river (in this case the River Thames) and all its tributaries. (See also "river catchment".)
River catchment	the area of land drained by a river. "Catchment" and "basin" are sometimes used interchangeably. Here catchments represent the drainage areas of the River Thames main tributaries, of which there are 10 in the West Thames.

Seasonal hydrological forecasts (SHF)	provide information about the hydrological conditions e.g. streamflow (river flows), groundwater levels and soil moisture levels, that might be expected over the next few months (e.g. from 3 weeks out to 7 months).
Seasonal meteorological forecasts	provide information about the weather conditions e.g. rainfall, air temperature, humidity, pressure, wind, that might be expected over the next few months (e.g. from 3 weeks out to 7 months).
Streamflow	the flow of water in a stream or river. Also known as river flow.
Surface runoff	the flow of water that occurs when water from excess rainfall, meltwater or drainage systems flows over the Earth's surface and not into the ground.
Tributary	a river or stream that flows into a larger stream, river or lake. Tributaries do not flow into the sea.
1 : 100-year flood event	a 100-year flood is a flood event that has a 1 % chance of occurring in any given year.
1 : 5-year flood event	a 1-in-5-year flood is a flood event that has a 20 % chance of occurring in any given year.

Information about the Supplement

- Supplement 1: Invitation flyer and programme for the focus group

- Supplement 2: West Thames catchment characteristic maps

- Supplement 3: Hydrological Summary: October 2013, June–September 2013 and November 2012–October 2013

- Supplement 4: Stage 1 Hydrological Outlook UK: November 2013–January 2014

- Supplement 5: Stage 2 EFAS-Seasonal: November 2013–February 2014

- Supplement 6: Stage 3 "Improved" EFAS-Seasonal: November 2013–February 2014

Author contributions. JLN and LA designed the decision-making activity. JLN, LA, SH, and HLC co-organized the set-up of the focus group. All the authors took part in delivering the focus group, including as note-takers, organizers, and presenters of their scientific research. JLN wrote the manuscript with input from all the authors.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. The information and findings in this paper are based on discussions and actions captured during the decision-making activity. They should not be taken as representing the views or practice of particular organizations or institutions.

References

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, J. Hydrol., 517, 913–922, https://doi.org/10.1016/j.jhydrol.2014.06.035, 2014.

Arnal, L., Ramos, M.-H., Coughlan de Perez, E., Cloke, H. L., Stephens, E., Wetterhall, F., van Andel, S. J., and Pappenberger, F.: Willingness-to-pay for a probabilistic flood forecast: a risk-based decision-making game, Hydrol. Earth Syst. Sci., 20, 3109–3128, https://doi.org/10.5194/hess-20-3109-2016, 2016.

Arnal, L., Wood, A. W., Stephens, E., Cloke, H., and Pappenberger, F.: An Efficient Approach for Estimating Streamflow Forecast Skill Elasticity, J. Hydrometeorol., 18, 1715–1729, https://doi.org/10.1175/JHM-D-16-0259.1, 2017.

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, Hydrol. Earth Syst. Sci., 22, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018, 2018.

Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A. A., Xavier, P., McLean, P., and Colman, A.: The GloSea4 Ensemble Prediction System for Seasonal Forecasting, Mon. Weather. Rev., 139, 1891–1910, https://doi.org/10.1175/2010MWR3615.1, 2010.

Asrar, G. R., Hurrell, J. W., and Busalacchi, A. J.: A need for "actionable" climate science and information: summary of WCRP open science conference, B. Am. Meteorol. Soc., 94, ES8–ES12, https://doi.org/10.1175/BAMS-D-12-00011.1, 2012.

Bell, V. A., Davies, H. N., Kay, A. L., Marsh, T. J., Brookshaw, A., and Jenkins, A.: Developing a large-scale water-balance approach to seasonal forecasting: application to the 2012 drought in Britain, Hydrol. Process., 27, https://doi.org/10.1002/hyp.9863, 2013.

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, Hydrol. Earth Syst. Sci., 21, 4681–4691, https://doi.org/10.5194/hess-21-4681-2017, 2017.

Bloomfield, J. P. and Marchant, B. P.: Analysis of groundwater drought building on the standardised precipitation index approach, Hydrol. Earth Syst. Sci., 17, 4769–4787, https://doi.org/10.5194/hess-17-4769-2013, 2013.

Bloomfield, J. P., Bricker, S. H., and Newell, A. J.: Some relationships between lithology, basin form and hydrology: A case study from the Thames basin, UK, Hydrol. Process., 25, 2518–2530, https://doi.org/10.1002/hyp.8024, 2011.

Bolson, J., Martinez, C., Breuer, N., Srivastava, P., and Knox, P.: Climate information use among southeast US water managers: beyond barriers and toward opportunities, Reg. Environ. Change, 13, 141–151, https://doi.org/10.1007/s10113-013-0463-1, 2013.

CEH: Hydrological Outlook – Further Information for November 2013, available at: http://www.hydoutuk.net/archive/2013/november-2013/further-information-november-2013/ (last access: 25 April 2018), 2013.

CEH: Hydrological Outlook UK, available at: http://www.hydoutuk.net/, last access: 9 April 2018.

Chatterton, J., Viavattene, C., Morris, J., Penning-Rowsell, E., and Tapsell, S.: The costs of the summer 2007 floods in England, Environment Agency Report SC070039, Rio House, Bristol, UK, 2010.

Chiew, F. H. S., Zhou, S. L., and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management, J. Hydrol., 270, 135–144, 2003.

Copernicus: EDgE, Climate Change Service, available at: http://edge.climate.copernicus.eu/ (last access: 31 May 2018), 2017a.

Copernicus: SWICCA: Service for Water Indicators in Climate Change Adaptation, SMHI, available at: http://swicca.climate.copernicus.eu/ (last access: 31 May 2018), 2017b.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., and Zsoter, E.: Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549–3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.

Crochemore, L., Ramos, M.-H., Pappenberger, F., van Andel, S. J., and Wood, A. W.: An experiment on risk-based decision-making in water management using monthly probabilistic forecasts, B. Am. Meteorol. Soc., 97, 541–551, 2015.

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. L: Seasonal climate predictability and forecasting: status and prospects, WIREs Clim. Change, 4, 245–268, https://doi.org/10.1002/wcc.217, 2013.

EA (Environment Agency): Thames Catchment Flood Management Plan – Managing Flood Risk, Summary Report December 2009, EA, Kings Meadow House, Reading, 2009.

EA (Environment Agency): The costs and impacts of the winter 2013 to 2014 floods, Technical Report SC140025, Defra/Environment Agency Joint R&D programme, 2015.

EA (Environment Agency): Groundwater Level Measurements (AfA075), Data contains Environment Agency information© Environment Agency and/or database right, All rights reserved, Data sourced under Environment Agency Conditional Licence, 2017.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev., 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Farolfi, S., Hassan, R., Perret, S., and MacKay, H.: A role-playing game to support multi-stakeholder negotiations related to water allocation in South Africa: First applications and potential developments, Midrand: Water Resources as Ecosystems: Scientists, Government and Society at the Crossroads, 2004.

Fry, M., Smith, K., Sheffield, J., Watts, G., Wood, E., Cooper, J., Prudhomme, C., and Rees, G.: Communication of uncertainty in hydrological predictions: a user-driven example web service for Europe, Geophys. Res. Abstr., EGU2017-16474, EGU General Assembly 2017, Vienna, Austria, 2017.

Golding, N., Hewitt, C., Zhang, P., Bett, P., Fang, X., Hu, H., and Nobert, S.: Improving user engagement and uptake of climate services in China, Climate Services, 5, 39–45, 2017.

Gray, D.: Gamestorming – Empathy Map, available at: http://gamestorming.com/empathy-mapping/ (last access: 1 May 2018), 2017.

Harrison, J.: Flood hazard management: Using an alternative community-based approach, Planet, 4, 5–6, 2002.

Huntingford, C., Marsh, T., Scaife, A. A., Kendon, E. J., Hannaford, J., Kay, A. L., Lockwood, M., Prudhomme, C., Reynar, N. S., Parry, S., Lowe, J. A., Screen, J. A., Ward, H. C., Roberts, M., Stott, P. A., Bell, V. A., Bailey, M., Jenkins, A., Legg, T., Otto, F. E. L., Massey, N., Schaller, N., Slingo, J., and Allen, M. A.: Potential influences on the United Kingdom's floods of winter 2013/14, Nat. Clim. Change, 4, 769–777, https://doi.org/10.1038/nclimate2314, 2014.

Ishikawa, T., Barnson, A. G., Kastens, K. A., and Louchouarn, P.: Understanding, evaluation, and use of climate forecast data by environmental policy students, in: Qualitative inquiry in geoscience education research, edited by: Feig, A. D. and Stokes, A., Geological Society of America Special Paper 474, 153–170, Geol. Soc. Am., Denver, CO, 2011.

IMPREX: Thames River Basin, available at: http://imprex.eu/thames-river-basin (last access: 8 April 2018), 2018a.

IMPREX: Risk Outlook Tool, available at: http://www.imprex.eu/innovation/risk-outlook (last access: 21 May 2018), 2018b.

Jones, L., Dougill, A., Jones, R. G., Steynor, A., Watkiss, P., Kane, C., Koelle, B., Moufouma-Okia, W., Padgham, J., Ranger, N., Roux, J.-P., Suarez, P., Tanner, T., and Vincent, K.: Ensuring climate information guides long-term development, Nat. Clim. Change, 5, 812–814, https://doi.org/10.1038/nclimate2701, 2015.

JRC: European Flood Awareness System, available at: https://www.efas.eu/ (last access: 9 April 2018), 2018a.

JRC: Global Flood Awareness System, available at: http://www.globalfloods.eu/user-information/seasonal_outlook/ (last access: 9 April 2018), 2018b.

Kendon, M. and McCarthy, M.: The UK's wet and stormy winter of 2013/2014, Weather, 70, 40–47, https://doi.org/10.1002/wea.2465, 2015.

Kirchhoff, C. J., Lemos, M. C., and Engle, N. L.: What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S., Environ. Sci. Policy, 26, 6–18, https://doi.org/10.1016/j.envsci.2012.07.001, 2013.

Lemos, M. C., Kirchhoff, C. J., and Ramprasad, V.: Narrowing the climate information usability gap, Nat. Clim. Change, 2, 789–794, 2012.

Lewis, H., Mittermaier, M., Mylne, K., Norman, K., Scaife, A., Neal, R., Pierce, C., Harrison, D., Jewell, S., Kendon, M., Saunders, R., Brunet, G., Golding, B., Kitchen, M., Davies, P., and Pilling, C.: From months to minutes – exploring the value of high-resolution rainfall observation and prediction during the UK winter storms of 2013/2014, Meteorol. Appl., 22, 90–104, 2015.

Li, Y., Giuliani, M., and Castelletti, A.: A coupled human–natural system to assess the operational value of weather and climate services for agriculture, Hydrol. Earth Syst. Sci., 21, 4693–4709, https://doi.org/10.5194/hess-21-4693-2017, 2017.

Lorenz, S., Dessai, S., Forster, P., and Paavola, J.: Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the

United Kingdom, Philos. T. Roy. Soc. A, 373, 20140457, https://doi.org/10.1098/rsta.2014.0457, 2015.

Mackay, J. D., Jackson, C. D., Brookshaw, A., Scaife, A. A., Cook, J., and Ward, R. S.: Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom, J. Hydrol., 530, 815–828, https://doi.org/10.1016/j.jhydrol.2015.10.018, 2015.

McEwen, L. J., Krause, F., Jones, O., and Garde Hansen, J.: Sustainable flood memories, informal knowledge and the development of community resilience to future flood risk, Transactions on Ecology and The Environment, 159, 253–263, 2012.

McEwen, L., Stokes, A., Crowley, K., and Roberts, C.: Using role-play for expert science communication with professional stakeholders in flood risk management, J. Geogr. Higher Educ., 38, 277–300, https://doi.org/10.1080/03098265.2014.911827, 2014.

Meadow, A., Ferguson, D., Guido, Z., Horangic, A., Owen, G., and Wall, T.: Moving toward the deliberate co-production of climate science knowledge, Weather, Clim. Soc., 7, 179–191, https://doi.org/10.1175/WCAS-D-14-00050.1, 2015.

Meißner, D., Klein, B., and Ionita, M.: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, Hydrol. Earth Syst. Sci., 21, 6401–6423, https://doi.org/10.5194/hess-21-6401-2017, 2017.

Met Office.: EUPORIAS Project, available at: https://www.metoffice.gov.uk/research/collaboration/euporias, last access: 16 June 2018.

Molteni, F., Stockdale, T., Alonso-Balmaseda, M., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T. N., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo., 656, 1–49, 2011.

Muchan, K., Lewis, M., Hannaford, J., and Parry, S.: The winter storms of 2013/2014 in the UK: hydrological responses and impacts, Weather, 70, 55–61, https://doi.org/10.1002/wea.2469, 2015.

Neumann, J. L., Arnal, L., Emerton, R., Griffith, H., Theofanidi, S., and Cloke, H.: Supporting the integration and application of seasonal hydrological forecasts in the West Thames, Technical Report for IMPREX, https://doi.org/10.13140/RG.2.2.19905.25447, 2017.

Neumann, J. L., Arnal, L. L. S., Magnusson, L., and Cloke, H. L.: The 2013/14 Thames basin floods: Do improved meteorological forecasts lead to more skilful hydrological forecasts at seasonal timescales?, J. Hydrometeorol., 19, 1059–1075, https://doi.org/10.1175/JHM-D-17-0182.1, 2018.

NRFA (National River Flow Archive): Search for Gauging Stations, available at: http://nrfa.ceh.ac.uk/data/search, last access: 10 July 2017.

NRFA (National River Flow Archive): Monthly Hydrological Summaries, available at: https://nrfa.ceh.ac.uk/monthly-hydrological-summary-uk?page=5, last access: 22 May 2018.

Parry, S., Prudhomme, C., Wilby, R., and Wood, P.: Chronology of drought termination for long records in the Thames catchment, in: Drought: Research and Science-Policy Interfacing, edited by: Andreu, J., Solera, A., Paredes-Arquiola, J., Haro-Monteagudo, D., and van Lanen, H., London, Taylor & Francis (CRC Press), 165–170, 2015.

Pavey, J. and Donoghue, D.: The use of role play and VLEs in teaching environmental management, Planet, 10, 7–10, 2003.

Prudhomme, C., Shaffrey, L. C., Woolings, T., Jackson, C. R., Fowler, H. J., and Anderson, B.: IMPETUS: Improving predictions of drought for user decision-making, in: Drought: Research and Science-Policy Interfacing, edited by: Andreu, J., Solera, A., Paredes-Arquiola, J., Haro-Monteagudo, D., and van Lanen, H., CRC Press, https://doi.org/10.1201/b18077-47, 2015.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R., and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, Hydrolog. Sci. J., 62, 2753–2768, https://doi.org/10.1080/02626667.2017.1395032, 2017.

Ramos, M. H., van Andel, S. J., and Pappenberger, F.: Do probabilistic forecasts lead to better decisions?, Hydrol. Earth Syst. Sci., 17, 2219–2232, https://doi.org/10.5194/hess-17-2219-2013, 2013.

Rapley, C. G., de Meyer, K., Carney, J., Clarke, R., Howarth, C., Smith, N., Stilgoe, J., Youngs, S., Brierley, C., Haugvaldstad, A., Lotto, B., Michie, S., Shipworth, M., and Tuckett, D.: Time for Change? Climate Science Reconsidered, Report of the UCL Policy Commission on Communicating Climate Science, 2014.

Rodwell, M. J., Ferranti, L., Magnusson, L., Weisheimer, A., Rabier, F., and Richardson, D.: Diagnosis of northern hemispheric regime behaviour during winter 2013/14, ECMWF Tech. Memo., 769, 1–12, 2015.

Soares, M. B. and Dessai, S. J.: Exploring the use of seasonal climate forecasts in Europe through expert elicitation, Climate Risk Management, 10, 8–16, 2015.

Soares, M. B. and Dessai, S. J.: Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe, Climatic Change, 137, 89–103, https://doi.org/10.1007/s10584-016-1671-8, 2016.

Soares, M. B., Alexander, M., and Dessai, S. J.: Sectoral use of climate information in Europe: A synoptic overview, Climate Services, 9, 5–20, 2018.

Thames Water: Hydrological Context for Water Quality And Ecology Preliminary Impact Assessments, Technical Appendix B, Thames Water Utilities Ltd 2W0H Lower Thames Operating Agreement (Cascade Consulting), 2010.

Turner, S. W. D., Bennett, J. C., Robertson, D. E., and Galelli, S.: Complex relationship between seasonal streamflow forecast skill and value in reservoir operations, Hydrol. Earth Syst. Sci., 21, 4841–4859, https://doi.org/10.5194/hess-21-4841-2017, 2017.

UK Gov: Water Situation Reports, available at: https://www.gov.uk/government/collections/water-situation-reports-for-england, last access: 5 May 2018.

Van der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci. 24, 189–212, https://doi.org/10.1080/13658810802549154, 2010.

van den Hurk, B. J. J. M., Bouwer, L. M., Buontempo, C., Döscher, R., Ercin, E., Hananel, C., Hunink, J. E., Kjellström, E., Klein, B., Manez, M., Pappenberger, F., Pouget, L., Ramos, M.-H., Ward, P. J., Weerts, A. H., and Wijngaard, J. B.: Improving predictions and management of hydrological extremes through climate services, Climate Services, 1, 6–11, 2016.

Vaughan, C., Buja, L., Kruczkiewicz, A., and Goddard, L.: Identifying research priorities to advance climate services, Climate Services 4, 65–74, 2016.

Viel, C., Beaulant, A.-L., Soubeyroux, J.-M., and Céron, J.-P.: How seasonal forecast could help a decision maker: an example of climate service for water resource management, Adv. Sci. Res., 13, 51–55, https://doi.org/10.5194/asr-13-51-2016, 2016.

Wells, M. and Davis, H.: Water transfer for public water supply via the CRT canal network, presentation Black and Veatch, 2016.

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, L14401, https://doi.org/10.1029/2008GL034648, 2008.

Yuan, X., Wood, E. F., and Ma, Z.: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, WIREs Water, 2, 523–536, https://doi.org/10.1002/wat2.1088, 2015.

A6: Global flood forecasting for averting disasters worldwide

This book chapter presents a co-author contribution arising through collaboration during this PhD, and has the following reference:

Hirpa, F. A., F. Pappenberger, L. Arnal, C. A. Baugh, H. L. Cloke, E. Dutra, **R. Emerton,** B. Revilla-Romero, P. Salamon, P. J. Smith, E. Stephens, F. Wetterhall, E. Zsoter and J. Thielen-del Pozo, 2018: Global flood forecasting for averting disasters worldwide. *Global Flood Hazard: Applications in Modeling, Mapping and Forecasting*, G. Schumann, P. D. Bates, H. Apel and G. T. Aronica, Eds., AGU Geophysical Monograph 233, John Wiley & Sons, Hoboken USA, 205-228

As this book chapter is not an open access publication, it has been included in this thesis for the purpose of the examination only and will not be available in any post-examination copies.

R.E. wrote the subsections "Global Flood Forecasting Systems" and "Continental-scale Flood Forecasting Systems" (12.3.1 and 12.3.2, pages 206-211, with the exception of section 12.3.1.1.2.), within the section "The Current Status of Large-Scale Flood Forecasting".