An assessment of the ECMWF extreme forecast index for water vapor transport during boreal winter

An Assessment of the ECMWF Extreme Forecast Index for Water Vapor Transport during Boreal Winter

DAVID A. LAVERS, ERVIN ZSOTER, DAVID S. RICHARDSON, AND FLORIAN PAPPENBERGER

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom

(Manuscript received 6 June 2017, in final form 21 July 2017)

ABSTRACT

Early awareness of extreme precipitation can provide the time necessary to make adequate event preparations. At the European Centre for Medium-Range Weather Forecasts (ECMWF), one tool that condenses the forecast information from the Integrated Forecasting System ensemble (ENS) is the extreme forecast index (EFI), an index that highlights regions that are forecast to have potentially anomalous weather conditions compared to the local climate. This paper builds on previous findings by undertaking a global verification throughout the medium-range forecast horizon (out to 15 days) on the ability of the EFI for water vapor transport [integrated vapor transport (IVT)] and precipitation to capture extreme observed precipitation. Using the ECMWF ENS for winters 2015/16 and 2016/17 and daily surface precipitation observations, the relative operating characteristic is used to show that the IVT EFI is more skillful than the precipitation EFI in forecast week 2 over Europe and western North America. It is the large-scale nature of the IVT, its higher predictability, and its relationship with extreme precipitation that result in its potential usefulness in these regions, which, in turn, could provide earlier awareness of extreme precipitation. Conversely, at shorter lead times the precipitation EFI is more useful, although the IVT EFI can provide synoptic-scale understanding. For the whole globe, the extratropical Northern Hemisphere, the tropics, and North America, the precipitation EFI is more useful throughout the medium range, suggesting that precipitation processes not captured in the IVT are important (e.g., tropical convection). Following these results, the operational implementation of the IVT EFI is currently being planned.

1. Introduction

The capability to provide early awareness of upcoming extreme precipitation and floods can allow discussions to be had at an early stage about appropriate event preparations. In recent years, it has been possible to provide awareness of extremes at longer lead times as a result of the increasing skill in global numerical weather prediction models (Bauer et al. 2015). This is exemplified by the warnings afforded for Hurricane Sandy in 2012 (Magnusson et al. 2014). In the midlatitude regions—namely, western Europe—recent research has built on predictability assessments of variables most relevant for predicting extreme precipitation (Lavers et al. 2014) to show that vertically integrated horizontal water vapor transport [integrated vapor transport (IVT)], a key driver of extreme events (e.g., Ralph et al. 2006; Lavers et al. 2011), can provide earlier awareness of extreme precipitation occurrence (Lavers et al. 2016a). In the study, the European Centre for Medium-Range Weather Forecasts (ECMWF) extreme forecast index (EFI; Lalaurette 2003; Zsoter 2006; Zsoter et al. 2015) for IVT and precipitation were compared. Results showed that the IVT EFI was more useful than the precipitation EFI in capturing extreme precipitation occurrence across western Europe toward the end of the 10-day forecast range for forecasts initialized in a positive phase of the North Atlantic Oscillation (NAO). The reasoning behind the result is that a positive NAO relates to a stronger westerly circulation, a phase in which high IVT is more likely, which, in turn, is then more likely to be the cause of extreme precipitation.

Denotes content that is immediately available upon publication as open access.

Corresponding author: David Lavers, david.lavers@ecmwf.int

DOI: 10.1175/WAF-D-17-0073.1

Conversely, in the negative NAO phase, the precipitation EFI was shown to be more useful.

In this study, the potential to use IVT as a variable to monitor precipitation extremes is investigated further by 1) undertaking a global assessment, with emphasis on the Northern Hemisphere winter, of how well the EFI for IVT and precipitation captures observed extreme precipitation; and 2) evaluating the whole medium-range forecast horizon (out to 15 days). This assessment is undertaken using ECMWF ensemble forecasts for 2015/16 and 2016/17 and daily observed precipitation. The analysis will identify in which regions and at what lead times it may be possible to use the IVT EFI (rather than the precipitation EFI) to provide earlier awareness of extreme daily precipitation.

2. Data and methods

a. ECMWF ensemble forecasts and reforecasts

The control and 50 perturbed ensemble members (out to 15 days) from the ECMWF Integrated Forecasting System ensemble (ENS) were retrieved from the ECMWF Meteorological Archival and Retrieval System (MARS) for the 0000 UTC initialization for two winter seasons, 1 November 2015–29 February 2016 and 20 October 2016–28 February 2017 (253 forecasts). Daily total surface precipitation accumulated at 0000 UTC was retrieved, and the specific humidity and the zonal and meridional winds at 300, 400, 500, 700, 850, 925, and 1000 hPa were retrieved at 0000 and 1200 UTC. The daily-averaged (using 0000, 1200, and 0000 UTC of the next day) vertically integrated horizontal zonal and meridional water vapor transport components were calculated in an Eulerian framework (e.g., Neiman et al. 2008), and combined into the total water vapor transport (IVT; although IVT is a vector, only its magnitude is considered in this analysis). Note that there was a horizontal resolution change from 32 km in 2015/16 to 18 km in 2016/17 owing to a model cycle upgrade.

ECMWF also generates reforecasts to provide a model climate against which the ENS can be compared. The reforecasts consist of 11 ensemble members and are run every Monday and Thursday from 0000 UTC for the past 20 years (e.g., reforecasts run from Thursday, 5 January 2017 have dates of 5 January 1997–5 January 2016). Herein, for the ENS from Monday to Wednesday (Thursday–Sunday), the model climate was built using the closest Monday (Thursday) reforecast date and four dates either side, in turn providing a climate of 1980 members (9 dates × 20 years × 11 members). For all reforecasts, as with the ENS, the daily total surface precipitation accumulated at 0000 UTC was retrieved, and the total water vapor transport (IVT) was calculated. The reforecasts of each winter had the same horizontal resolution as those of the corresponding real-time ENS.

b. ECMWF EFI

The EFI (Dutra et al. 2013; Lalaurette 2003; Zsoter 2006; Zsoter et al. 2015) compares the probability distribution of the ENS and the corresponding model climate, therefore revealing the extremeness of a forecast. This type of calibration mitigates the differences between the model cycles.

The EFI is calculated as

$$\mathrm{EFI} = \frac{2}{\pi}\int_0^1 \frac{p - F(p)}{\sqrt{p(1-p)}}\,dp, \tag{1}$$

where $F(p)$ is the proportion of ensemble members that lie below the pth percentile of the model climate. EFI values lie between −1 and 1, with −1 highlighting extremely low and 1 highlighting extremely high values with respect to the model climate.

The EFI was calculated for IVT and precipitation on each forecast day (1–15) for all 253 forecasts. A skill assessment was undertaken on the IVT and precipitation EFI forecasts to see how well they captured extreme precipitation events.

c. Observed precipitation and forecast verification

Daily land-based gauged precipitation observations (at 0000 UTC) from the World Meteorological Organization (WMO) Global Telecommunication System (GTS) were used as the verification dataset. Extreme observed precipitation events in a particular month were determined through comparison with the ECMWF model climate precipitation on forecast day 7 initialized from the middle of the month. An extreme event was identified if the observed total exceeded the 99th percentile of the model climate at the closest grid point to the gauge. Note that an observed climatology for each gauge was not built because the requirement of a 20-yr record (to match the reforecasts) would lead to over half of the gauges not being used, thus seriously affecting the validity of the study that focuses on extremes. Although the model climate may not reproduce the exact observed distribution because of model biases in some areas, it does take into account geographical variations of extremes and provides a sufficiently large sample that would otherwise not be possible.

The ability of the EFI forecasts to discriminate extreme precipitation (>99th percentile) is assessed using the hit rate (probability of detection) and the false alarm rate (probability of false detection); they are defined as

the fraction of forecast hits with respect to the total observed events and the number of false alarms to total nonevents, respectively. These hit and false alarm rates were evaluated with relative operating characteristic (ROC) curves for EFI thresholds ranging from 0 to 1 (Wilks 2006). The ROC area or score, which ranges from 0 to 1, was also calculated, with areas > 0.5 representing a skillful forecast. In this evaluation, we consider six geographical regions: 1) global, 2) Northern Hemisphere (>20°N), 3) tropics (between 20°N and 20°S), 4) Europe (35°–75°N, 12.5°W–20°E), 5) North America (25°–70°N, 170°–50°W), and 6) western North America (25°–70°N, 170°–117°W).

The statistical significance of the differences in ROC areas between IVT and precipitation was investigated using a bootstrap procedure. The bootstrap process involved resampling the 253 forecasts (for a particular forecast day) with replacement and calculating the resultant ROC area and then repeating the process 1000 times.

3. Results and discussion

a. Forecast evaluation using the ROC

In Fig. 1, the ROC areas and their bootstrap distributions for the IVT and precipitation EFI on each forecast day are shown for the six geographical areas. For all regions, the plot shows that in forecast week 1 the precipitation EFI has a higher ROC score than the IVT EFI, suggesting that it is more useful in discriminating the extreme precipitation events. This is highlighted, for example, by a ROC score difference on forecast day 1 of ≥ 0.2 in five of the six regions (except western North America; Fig. 1f). The reasoning behind the poorer IVT EFI performance is partly because of the large-scale characteristic of IVT fields, which gives a higher false alarm rate around an extreme. Throughout the medium range (out to 15 days), the difference in ROC scores becomes smaller owing to a more rapid score degradation for precipitation than for IVT, a property likely to be due to lower precipitation predictability (Lavers et al. 2014, 2016b). In the tropics (Fig. 1c) the IVT is less useful compared with other regions and hence has the largest ROC score difference with precipitation. This is because in the tropics, extreme precipitation is closely associated with convection (i.e., strong vertical uplift) and is less related to IVT.

In forecast week 2, the ROC scores for precipitation and IVT generally converge, and in two regions—namely, Europe and western North America (Figs. 1d and 1f, respectively)—the IVT has higher ROC scores, thus showing more usefulness (the statistical significance is discussed below). In these areas, extreme winter precipitation is connected to strong IVT, often within atmospheric rivers (e.g., Ralph et al. 2006; Ramos et al. 2015). It is likely that the large-scale nature of the IVT, a drawback at shorter lead times, enables it to capture the approximate location of extreme events, whereas the precipitation being more local in nature and tied to land–atmosphere interaction is less able to discriminate extremes at longer lead times. This finding suggests that the monitoring of IVT in forecast week 2 may yield earlier awareness of extreme precipitation in Europe and western North America. Note that a test was undertaken in which the European domain was extended from 20° to 40°E. For the larger domain out to 40°E, the IVT EFI had no added advantage over the precipitation EFI primarily because farther inland the extreme precipitation is less related to IVT.

The ROC curves on forecast day 13 for IVT and precipitation are shown for Europe and western North America in Figs. 2a and 2b, respectively. These curves underscore that the IVT EFI is more useful in discriminating between extreme precipitation events and nonevents, as the IVT ROC curve is almost everywhere closer to the top-left corner, which results in a higher ROC area. In Figs. 2c and 2d, we display box plots of the bootstrapped ROC score differences (IVT − precipitation) for forecast day 13 for Europe and western North America, respectively. For Europe, there is evidence, although not statistically significant, that using the IVT EFI leads to an improved ROC score and more usefulness. Future work will consider European areas with the strongest IVT link with precipitation in an attempt to raise this skill. For western North America, the box plot suggests a statistically significant difference at above the 97.5% level. This finding provides strong evidence that the IVT EFI is more useful than the precipitation EFI at this lead time.

b. Examples of the EFI over western North America and Europe

Having shown that the IVT EFI has more usefulness in forecast week 2, we present examples of where this benefit may have been realized. In winter 2016/17, western North America experienced many atmospheric rivers and heavy precipitation. Figure 3 shows the EFI on forecast day 13 (left column) and forecast day 1 (right column) valid for an extreme precipitation event in the Pacific Northwest on 15 February 2017. Strikingly, the EFI patterns on forecast day 13 (initialization: 0000 UTC 3 February 2017) are very similar to the day 1 patterns, which signifies an unusually high level of atmospheric predictability. It is thought that the strong Madden–Julian oscillation that was over the Maritime Continent

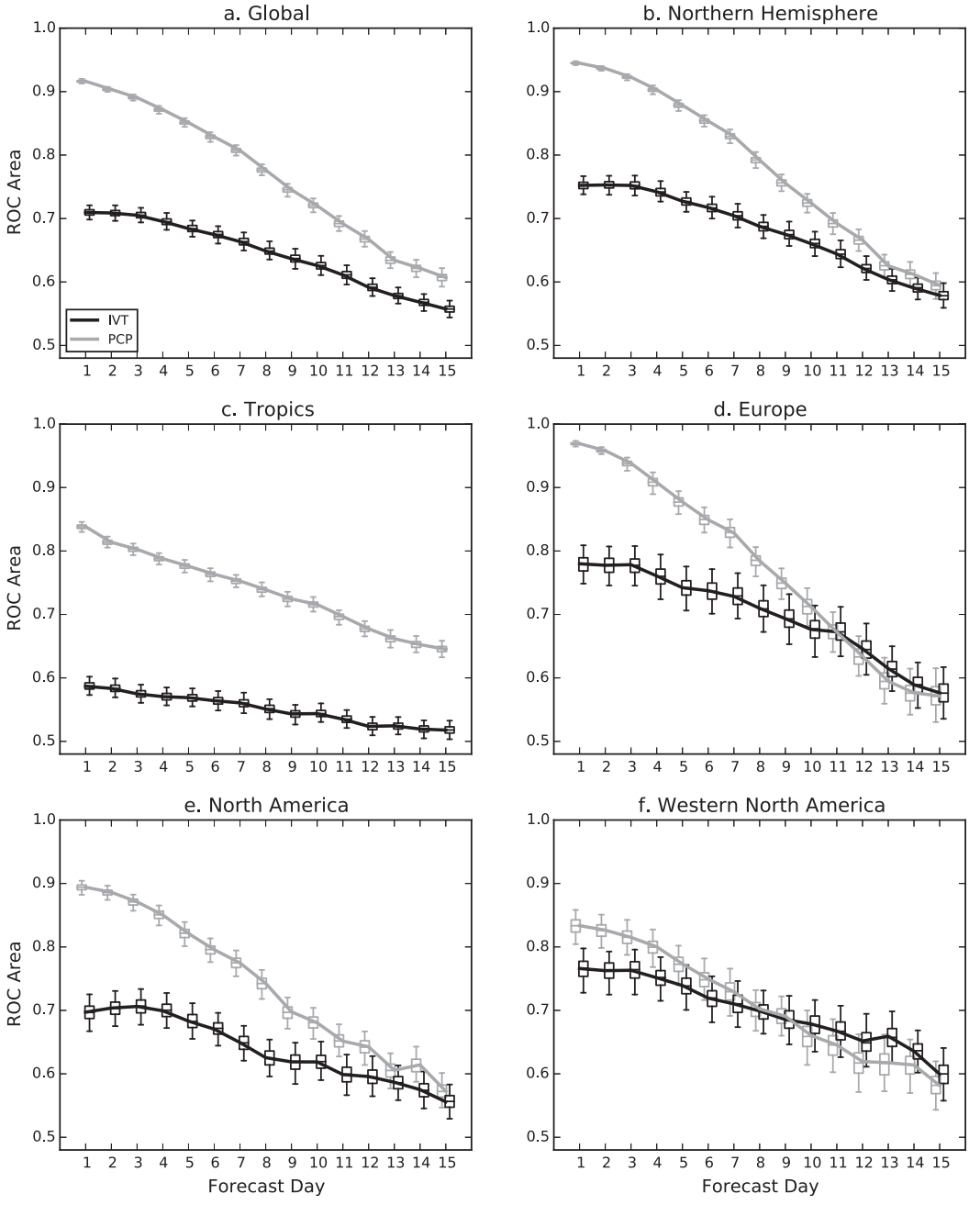

FIG. 1. The change in ROC areas or scores with forecast day using all 253 forecasts for (a) the whole globe, (b) Northern Hemisphere, (c) tropics, (d) Europe, (e) North America, and (f) western North America (see section 2c for the domain definitions), for IVT (black lines) and precipitation (gray lines), and the box plots show the distribution of the 1000 bootstrapped samples. The bottom and top of the boxes correspond to the 25th and 75th percentiles, respectively; the line in the box is the median; and the whiskers represent the 2.5nd and 97.5th percentiles, respectively.

and western Pacific Ocean in early February 2017 was behind this high predictability (e.g., Ferranti et al. 1990; Vitart and Molteni 2010). The IVT EFI clearly shows an atmospheric river affecting the region, and its ability to more accurately discriminate extreme precipitation events (black points) in Washington State compared to the precipitation EFI on forecast day 13 is found (cf. Figs. 3a and 3c). On forecast day 1, the issue raised in section 3a about the IVT EFI false alarm rate can be pictured because the high EFI values occur from northern

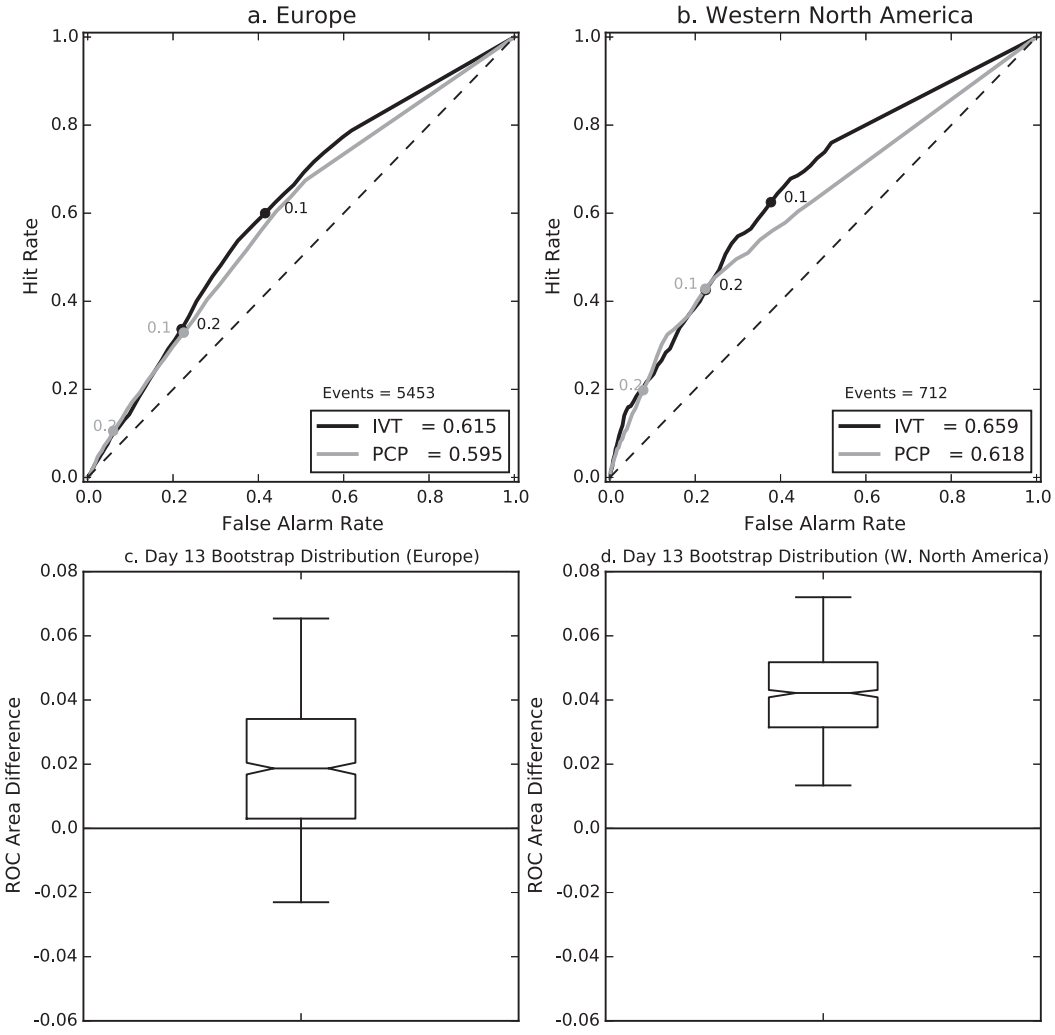

FIG. 2. The ROC curves for IVT (black lines) and precipitation (gray lines) for all 253 forecasts on forecast day 13 for (a) Europe and (b) western North America. The ROC areas are provided in the legends, and the number of extreme precipitation events is also given. Box plots of the ROC area differences (IVT − precipitation) on forecast day 13 for (c) Europe and (d) western North America calculated from a bootstrap process that was repeated 1000 times using resampling of the 253 forecast days with replacement. The bottom and top of the boxes correspond to the 25th and 75th percentiles, respectively; the line in the box is the median; and the whiskers represent the 2.5nd and 97.5th percentiles, respectively. In the boxes, the notch shows the 95% confidence interval around the median from a 1000-bootstrapped sample.

California to central British Columbia, in turn capturing a much larger swath around the extreme event than the precipitation EFI. Although at these short lead times the IVT EFI is less useful, it does provide a synoptic context to the event, and can thus aid process understanding.

In winter 2015/16, northwestern Europe was affected by multiple extratropical cyclones. Lavers et al. (2016a) showed the IVT and precipitation EFI valid for storm Desmond on 5 December 2015, a storm whose atmospheric river resulted in record precipitation and flooding. In Fig. 4, we further show that a signal for this event was found in the IVT EFI on forecast day 13 (initialization: 0000 UTC 23 November 2015) and that the IVT EFI was more able to capture the extreme precipitation across Ireland, the United Kingdom, and Sweden. Also, note that at short lead times the EFI generally has stronger values, which shows that the distribution of the ensemble forecasts differs more compared to that of the model climate.

4. Conclusions

The aim of this study was to 1) evaluate the ability of the EFI for IVT and precipitation to detect extreme

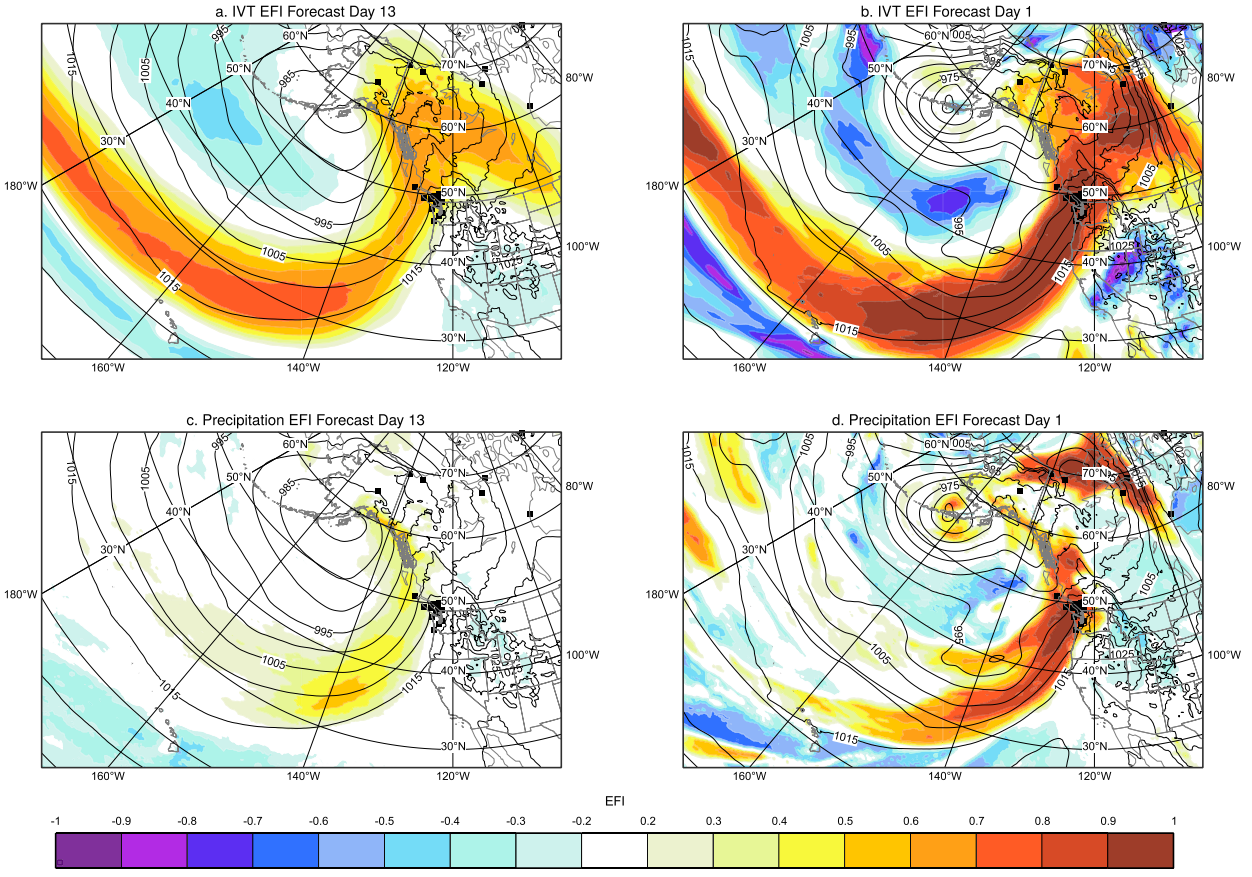

FIG. 3. The EFI and ensemble average mean sea level pressure (MSLP) fields valid for 15 Feb 2017. (a) IVT and (c) precipitation EFI from forecasts initialized at 0000 UTC 3 Feb 2017 (forecast day 13); MSLP from forecasts initialized at 0000 UTC 3 Feb 2017 valid at 1200 UTC 15 Feb 2017 ($T + 300$ h). (b) IVT and (d) precipitation EFI from forecasts initialized at 0000 UTC 15 Feb 2017 (forecast day 1); MSLP from forecasts initialized at 0000 UTC 15 Feb 2017 valid at 1200 UTC 15 Feb 2017 ($T + 12$ h). Precipitation gauges that exceeded the 99th percentile of the model climate at the closest grid point to the gauge are represented by black dots.

precipitation events across the globe and 2) assess the whole medium-range forecast horizon, thus extending the analysis in Lavers et al. (2016a). This research used ECMWF ensemble forecasts from two winters and WMO GTS precipitation observations. Results suggest that the IVT EFI is more useful than the precipitation EFI in capturing extreme events during forecast week 2 in Europe and western North America. These regions, being situated in the exit regions of the North Atlantic and North Pacific jet streams, climatologically have time-mean large-scale IVT convergence, with extreme winter precipitation associated with atmospheric rivers (e.g., Lavers et al. 2011; Ralph et al. 2006) that are often diagnosed by IVT. It is the large-scale IVT characteristics and higher predictability that are likely to allow the IVT to capture the location of precipitation extremes at longer lead times. These findings show the potential to use the IVT EFI for earlier awareness of upcoming precipitation extremes and flooding. Conversely, at shorter lead times across Europe and western North America, the precipitation EFI is more useful, partly because the IVT EFI covers a larger area and therefore has a high false alarm rate (see Figs. 3b and 4b). It is also possible that IVT is not the main driver of some precipitation events.

For the four other geographical regions assessed, the precipitation EFI is more skillful throughout the medium range. This is in part because factors other than the IVT are responsible for extreme precipitation, and this is exemplified for the tropics, where extreme precipitation is predominantly linked to convection and not IVT. We note that one limitation of the analysis undertaken is the study period length, which may affect the results, a point that will be addressed in future research by considering all seasons, with more focus specifically given to areas where IVT is a significant factor in precipitation. It is also

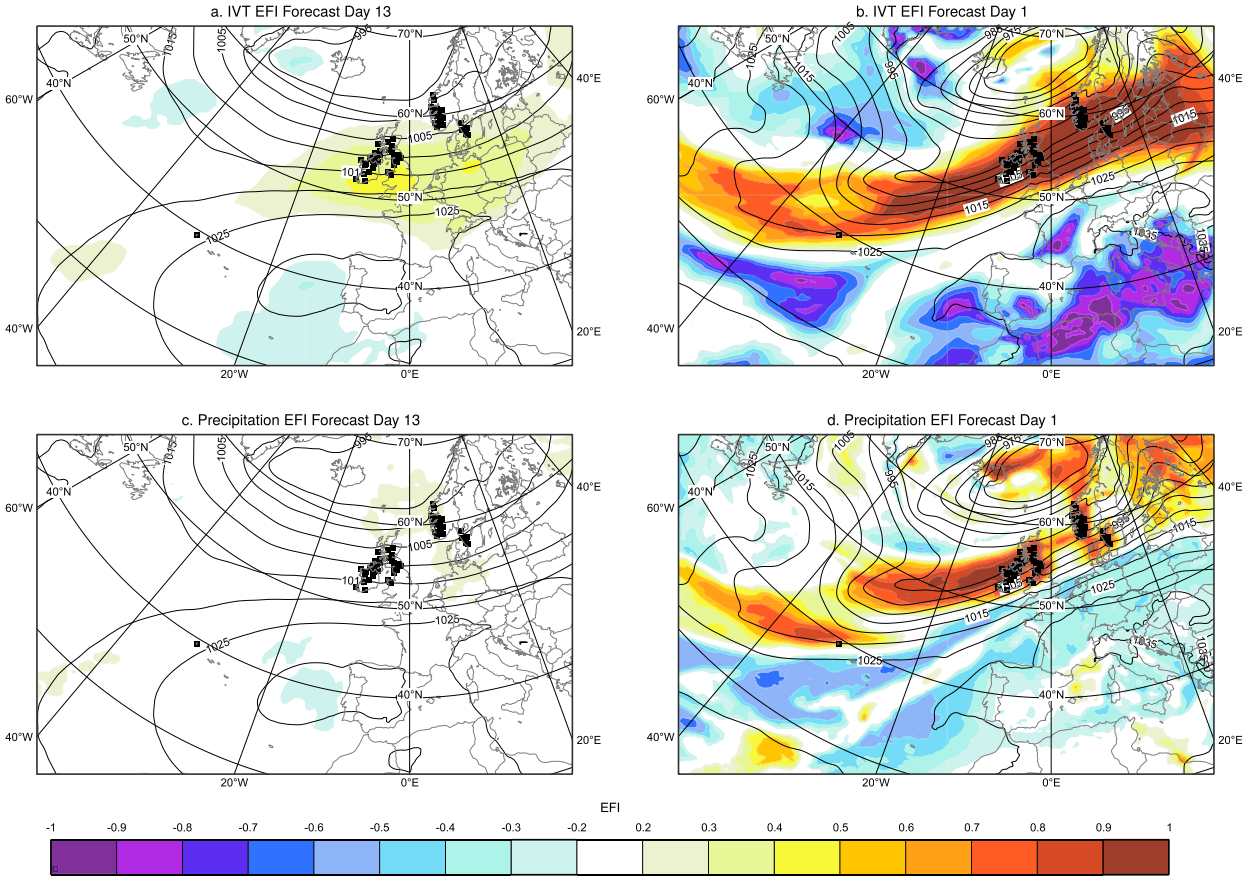

FIG. 4. The EFI and ensemble average MSLP fields valid for 5 Dec 2015. (a) IVT and (c) precipitation EFI from forecasts initialized at 0000 UTC 23 Nov 2015 (forecast day 13); MSLP from forecasts initialized at 0000 UTC 23 Nov 2015 valid at 1200 UTC 5 Dec 2015 (T + 300 h). (b) IVT and (d) precipitation EFI from forecasts initialized at 0000 UTC 5 Dec 2015 (forecast day 1); MSLP from forecasts initialized at 0000 UTC 5 Dec 2015 valid at 1200 UTC 5 Dec 2015 (T + 12 h). Precipitation gauges that exceeded the 99th percentile of the model climate at the closest grid point to the gauge are represented by black dots.

possible that analyses for the extratropical Southern Hemisphere (e.g., southwestern South America) winter will have similar results to those found herein for western North America and Europe.

In this research, conditional verification on large-scale atmospheric modes was not undertaken because of the too small sample size that would result. In Lavers et al. (2016a), by conditioning the verification on the NAO phase at initialization, the IVT EFI was shown to be more skillful over Europe than the precipitation EFI in a positive NAO phase at earlier lead times than shown herein. It is possible that similar results would occur in western North America with the Pacific North American pattern, and this will be considered in the future. Although at shorter lead times the IVT EFI is less useful, it can complement the precipitation EFI by providing the synoptic context for an event. Planning is currently underway for the operational implementation of the IVT EFI.

Acknowledgments. The authors gratefully acknowledge the financial support from the European Union Horizon 2020 IMPREX project (Grant Agreement 641811) and the ANYWHERE project (Grant Agreement 700099). We thank Ivan Tsonevsky and Paul Smith for discussions that helped this research, and Fernando Ii for help with Figs. 3 and 4. The data used are available on the ECMWF MARS server. The authors also thank the three anonymous reviewers for their comments, which helped to improve the paper.

REFERENCES

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, doi:10.1038/nature14956.

Dutra, E., M. Diamantakis, I. Tsonevsky, E. Zsoter, F. Wetterhall, T. Stockdale, D. Richardson, and F. Pappenberger, 2013: The extreme forecast index at the seasonal scale. *Atmos. Sci. Lett.*, **14**, 256–262, doi:10.1002/asl2.448.

Ferranti, L., T. N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical–extratropical interaction associated with the 30–60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177–2199, doi:10.1175/1520-0469(1990)047<2177:TEIAWT>2.0.CO;2.

Lalaurette, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteor. Soc.*, **129**, 3037–3057, doi:10.1256/qj.02.152.

Lavers, D. A., R. P. Allan, E. F. Wood, G. Villarini, D. J. Brayshaw, and A. J. Wade, 2011: Winter floods in Britain are connected to atmospheric rivers. *Geophys. Res. Lett.*, **38**, L23803, doi:10.1029/2011GL049783.

——, F. Pappenberger, and E. Zsoter, 2014: Extending medium-range predictability of extreme hydrological events in Europe. *Nat. Commun.*, **5**, 5382, doi:10.1038/ncomms6382.

——, ——, D. S. Richardson, and E. Zsoter, 2016a: ECMWF Extreme Forecast Index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophys. Res. Lett.*, **43**, 11 852–11 858, doi:10.1002/2016GL071320.

——, D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016b: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, **43**, 2275–2282, doi:10.1002/2016GL067765.

Magnusson, L., J.-R. Bidlot, S. T. K. Lang, A. Thorpe, N. Wedi, and M. Yamaguchi, 2014: Evaluation of medium-range forecasts of Hurricane Sandy. *Mon. Wea. Rev.*, **142**, 1962–1981, doi:10.1175/MWR-D-13-00228.1.

Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeor.*, **9**, 22–47, doi:10.1175/2007JHM855.1.

Ralph, F. M., P. J. Neiman, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, doi:10.1029/2006GL026689.

Ramos, A. M., R. M. Trigo, M. L. R. Liberato, and T. Ricardo, 2015: Daily precipitation extreme events in the Iberian Peninsula and its association with atmospheric rivers. *J. Hydrometeor.*, **16**, 579–597, doi:10.1175/JHM-D-14-0103.1.

Vitart, F., and F. Molteni, 2010: Simulation of the Madden–Julian Oscillation and its teleconnections in the ECMWF forecast system. *Quart. J. Roy. Meteor. Soc.*, **136**, 842–855, doi:10.1002/qj.623.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Zsoter, E., 2006: Recent developments in extreme weather forecasting. *ECMWF Newsletter*, No. 107, ECMWF, Reading, United Kingdom, 8–17.

——, F. Pappenberger, and D. Richardson, 2015: Sensitivity of model climate to sampling configurations and the impact on the Extreme Forecast Index. *Meteor. Appl.*, **22**, 236–247, doi:10.1002/met.1447.