The EUSTACE project: delivering global, daily information on surface air temperature

Article

Accepted Version

The EUSTACE project: delivering global, daily information on surface air temperature

Nick A. Rayner

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Renate Auchmann

Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

Janette Bessembinder

Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands

Stefan Brönnimann

Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

Yuri Brugnara

Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

Francesco Capponi,

Met Office Hadley Centre, Exeter, United Kingdom Now Data Scientist within the Decision Service Team at WorldPay.

Laura Carrea

Department of Meteorology

University of Reading, Reading, RG6 6AL, UK

Emma M. A. Dodd

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom

Darren Ghent

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom

Elizabeth Good

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Jacob L Høyer

Danish Meteorological Institute, Copenhagen, Denmark

John J Kennedy

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Elizabeth C. Kent

National Oceanography Centre, Southampton, UK

Rachel E Killick

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Paul van der Linden

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Finn Lindgren

University of Edinburgh, School of Mathematics, Edinburgh, United Kingdom

Kristine S. Madsen

Danish Meteorological Institute, Copenhagen, Denmark

Christopher J. Merchant

National Centre for Earth Observation and Department of Meteorology University of Reading, Reading, RG6 6AL, UK

Joel R Mitchelson

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Colin P Morice

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

Pia Nielsen-Englyst

Danish Meteorological Institute, Copenhagen, Denmark

Patricio F. Ortiz

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom now at Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, United Kingdom

John J Remedios

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom

Gerard van der Schrier

Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands

Antonello A. Squintu

Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands

Ag Stephens

Science and Technology Facilities Council, Didcot, United Kingdom

Peter W. Thorne

Department of Geography, National University of Ireland Maynooth, Maynooth, Ireland.

Rasmus T. Tonboe

Danish Meteorological Institute, Copenhagen, Denmark

Tim Trent

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom

Karen L Veal

National Centre for Earth Observation, University of Leicester, Leicester, United Kingdom

Alison M Waterfall

Science and Technology Facilities Council, Didcot, United Kingdom

Kate Winfield

Science and Technology Facilities Council, Didcot, United Kingdom

Jonathan Winn

Met Office Hadley Centre, Met Office, FitzRoy Road, Exeter, EX1 3PB

R. Iestyn Woolway

Department of Meteorology

University of Reading, Reading, RG6 6AL, UK Now at Centre for Freshwater and Environmental Studies Dundalk Institute of Technology, Dundalk, A91 K584, Ireland

Corresponding author: Nick A. Rayner, nick.rayner@metoffice.gov.uk

Capsule

The main goals and activities of the EUSTACE project are discussed along with some key results, including a global, multi-decadal daily air temperature record from satellite and *in situ* measurements.

Abstract

Day-to-day variations in surface air temperature affect society in many ways, but daily surface air temperature measurements are not available everywhere. Therefore, a global daily picture cannot be achieved with measurements made *in situ* alone and needs to incorporate estimates from satellite retrievals.

This article presents the science developed in the EU Horizon 2020-funded EUSTACE project (2015-2019, https://www.eustaceproject.org) to produce global and European, multi-decadal ensembles of daily analyses of surface air temperature complementary to those from dynamical reanalyses, integrating different ground-based and satellite-borne data types. Relationships between surface air temperature measurements and satellite-based estimates of surface skin temperature over all surfaces of Earth (land, ocean, ice and lakes) are quantified. Information contained in the satellite retrievals then helps to estimate air temperature and create global fields in the past, using statistical models of how surface air temperature varies in a connected way from place to place; this needs efficient statistical analysis methods to cope with the considerable data volumes. Daily fields are presented as ensembles to enable propagation of uncertainties through applications. Estimated temperatures and their uncertainties are evaluated against independent measurements and other surface temperature data sets.

Achievements in the EUSTACE project have also included fundamental preparatory work useful to others, for example: gathering user requirements; identifying inhomogeneities in daily surface air temperature measurement series from weather stations; carefully quantifying uncertainties in satellite skin and air temperature estimates; exploring the

interaction between air temperature and lakes; developing statistical models relevant to non-Gaussian variables; and methods for efficient computation.

Body text

EU Surface Temperature for All Corners of Earth (EUSTACE, https://www.eustaceproject.org) is a 4-yr research project funded by the European Union Horizon 2020 research and innovation programme (EU H2020; Grant Agreement 640171; see Appendix A for a list of the Consortium's institutions) that started on 1 January 2015. EUSTACE has used temperature estimates from satellites to boost the amount of information available beyond that provided by weather stations and ships to help to construct a prototype global, multi-decadal daily air temperature record presented on a 0.25^{o} latitude by 0.25^{o} longitude grid.

Near-surface air temperature (typically measured at a height of about 2 m above ground level at meteorological stations) is a fundamental quantity for many of the activities undertaken in climate science and in many of the societal concerns that climate services aim to support; it is something that we all experience directly in our day-to-day lives. Near-surface air temperature has been measured almost continuously in some places and across the global oceans by ships for well over a century. Designated as an Essential Climate Variable (ECV), these records allow for the construction of a useful climate data record (CDR) in those places for the period covered. Globally, however, there a number of locations where either access to the measurements is not possible, or no air temperature records exist. As well as long records of direct measurements of near-surface air temperature, we have information from satellite retrievals (i.e. remotely-sensed, indirect estimates) of temperature. However, satellite retrievals tend not to pertain to the air temperature that we experience directly, but either to an average temperature of a higher layer in the atmosphere or to the skin temperature of the surface of the Earth. These quantities are related to near-surface air temperature, more or less tightly depending on the type of surface and the surface-lower-atmosphere interactions. Therefore, it is possible to use satellite-derived temperatures together with near-surface air temperature measurements to create a more complete climate data record of air temperature. Thus, EUSTACE created a prototype global climate data record of near-surface air temperature for every day since January 1850 using both direct measurements of air temperature and estimates of it based on satellite skin temperature retrievals.

Near-surface air temperature products provide valuable information for a range of activities, from the monitoring of current conditions (e.g. Sánchez-Lugo et al. 2019) to the assessment of past variability (e.g. Osborn et al. 2017) to their use in seasonal-to-decadal forecasting (e.g. Kushnir et al. 2019), climate model evaluation (e.g. Walters et al. 2019), detection and attribution of climate change (e.g. Jones and Kennedy 2017), Intergovernmental Panel on Climate Change Assessments (e.g. Hartmann et al. 2013), agricultural modelling (e.g. Weedon et al. 2011), health modelling (e.g. Xu et al. 2019) and other downstream uses. Such a daily surface air temperature product could form part of the future operational monitoring system for surface air temperature over the polar regions, over Africa and South America. EUSTACE has already enabled monitoring of lake surface water temperature to be included in the annual State of the Global Climate reports (for the years 2015, 2016, 2017 and 2018; Woolway et al., 2016, 2017a and 2018; Carrea et al., 2019). EUSTACE products are complementary to products from dynamical reanalyses (e.g. Buizza et al. (2018)) with much of the work dedicated to the preparation of input surface temperature observations, for which EUSTACE has performed thorough uncertainty analyses, which were previously lacking.

Dynamical reanalyses combine historical and recent observations with numerical weather prediction models to produce dynamically-consistent reconstructions of past weather and climate. These reanalyses require observational data with well-characterised uncertainties. The new, validated, estimates of uncertainty in satellite surface skin temperature observations developed by EUSTACE are of benefit to them. EUSTACE products also provide an alternative source of near-surface air temperature data that is independent from numerical weather prediction models and extends further back in time than most dynamical reanalyses.

Results from scientific projects are often not produced in a format that can be used easily by others; in general, processing or translation is needed. Two-way interaction with potential users from the start of a project helps to increase the relevance and usability of products to various potential user groups. EUSTACE collected information on user requirements in several ways, via: user consultation workshops; questionnaires and interviews; a literature review on user requirements (Bessembinder et al. 2016; Bessembinder 2017, including the results from a large number of national and EU projects); testing of example mock-up datasets; and describing specific use cases with “trail blazer” users.

These activities resulted in greater insight into how climate data are used, data format preferences, and which variables are needed (i.e. not just daily mean temperature, but also minimum and maximum temperature), amongst other things. We used many of the user requirements collected to design the EUSTACE data file structure and the user guides; for example, a quick start guide is provided as part of the product user guide, together with example use cases.

While many of the ideas used within EUSTACE have been trialled elsewhere for individual regions (e.g. Cristóbal et al. (2008)), or for different time scales (e.g. Kilibarda et al. (2014)), EUSTACE has brought them together for the first time to create global, multi-decadal daily products. EUSTACE has performed an integrating function, bringing together products and expertise from a wide range of European, national and international initiatives. EUSTACE has also followed much of the road map of “recommended steps towards meeting societal needs for surface temperature understanding and information” set out previously in the EarthTemp Network Community Paper (Merchant et al. 2013). In particular, EUSTACE has made progress in seven out of the ten broad aims identified therein:

- develop more integrated, collaborative approaches to observing and understanding Earth’s various surface temperatures;
- build understanding of the relationships between different surface temperatures, where presently inadequate;
- make surface temperature datasets easier to obtain and exploit for a wider constituency of users;

- consistently provide realistic uncertainty information with surface temperature datasets;
- communicate differences and complementarities of different types of surface temperature datasets in readily understood terms;
- rescue, curate and make available valuable surface temperature data that are presently inaccessible; and
- build capacities to accelerate progress in the accuracy and usability of surface temperature datasets.

Computer code has been developed both to estimate air temperature from satellite data and to create daily maps of mean air temperature; this code has been publicly released (Rayner 2019). Information contained in the satellite retrievals helps to create more-complete fields in the past, via statistical models of how surface air temperature varies in a connected way from place to place. As the data volumes involved are considerable, the EUSTACE partnership included statisticians and computer scientists, enabling the development of efficient analysis methods. As a result, EUSTACE has been able to demonstrate that these methods can be built into a fully functional processing system, with research-level maturity (EUMETSAT, 2014) which exploits the features of modern high performance computing resources to deliver the more-complete datasets described below. This system could be used in future to update some of the EUSTACE data sets described here to enable their use in climate monitoring.

The datasets that are currently commonly used to monitor surface temperatures globally are constructed as a combination of air temperature observations over land and sea surface temperature observations over ocean. The current versions of the most widely used global near-surface temperature datasets, HadCRUT4 (Morice et al., 2012), NOAAGlobalTemp (Smith et al., 2008; Vose et al., 2012) and GISTEMP (Hansen, 2010), extend from the mid-19th century to present and are derived from *in situ* observations only; temperature retrievals from satellites are not used in their construction. These global temperature datasets are presented at monthly resolution because summaries of monthly average temperatures are more commonly available for individual meteorological stations and cover a greater region of the Earth than daily or sub-daily summaries in the 19th century and early 20th century. The density distribution of available *in situ* temperature observations limits the spatial resolution of these products. For example, HadCRUT4 is provided as monthly fields on an equi-angle latitude-longitude grid at 5^{o} resolution.

Surface air temperature datasets covering land regions, but not ocean or sea ice, are available at higher spatial and temporal resolutions. For example, Rhode (2013a; 2013b) use a larger number of meteorological stations than do HadCRUT4, NOAAGlobalTemp or GISTEMP, together with a statistical interpolation algorithm, to produce a monthly surface air temperature dataset at higher spatial resolution; an experimental daily analysis has also been produced. Other high-resolution datasets of air temperatures over land are available and are commonly used in climate modelling (Harris et al., 2013) and hydrological modelling (Weedon et al., 2011). Higher temporal resolution air temperatures derived from land meteorological station observations are also available, including the daily GHCN-D databank

(Menne et al., 2012), and the sub-daily HadISD databank (Dunn et al., 2016). Gridded temperature fields based on GHCN-D are available in the HadGHCN-D dataset (Caesar, et al., 2006) covering a time period from 1950 to present. HadISD is presented as time series for individual meteorological stations only. However, none of these latter datasets are based on homogenised data (see below).

The existing coarse-resolution global temperature datasets are widely used in global and regional climate assessments; however their utility is limited in some applications that require information at high temporal and/or spatial resolutions, such as the assessment of temperature extremes, national climate assessments, regional impact studies and validation of climate simulations from high-resolution climate models. These global temperature datasets are also often expressed in terms of temperature anomalies (temperatures relative to average conditions over some reference period), rather than in terms of absolute temperature information, which is commonly needed in these applications. EUSTACE provides products that can be used for the study of absolute temperatures, as well as providing information relevant to temperature anomalies.

Figure 1 provides an overview of the EUSTACE process and shows how different activities linked together to transform the source datasets (Appendix B) into the series of EUSTACE products (Appendix C). Source data sets were chosen to maximise our opportunity to quantify the components of uncertainty (in the case of satellite data) and the amount of historical daily information (in the case of weather station data). Wrapped around these scientific developments were interactions throughout the project with potential users.

Evaluation against independent reference measurements (Veal, 2019a) and comparison with other related products (Veal, 2019b) put EUSTACE work into context.

Through this development process, EUSTACE has contributed to advancing and enabling climate science in five main areas:

1) Detecting and correcting for non-climatic discontinuities in weather station series: to provide an accurate picture of variations in air temperature, measurements at weather stations have been checked for any jumps in the series and then corrected (Squintu et al., 2019a and b). Such discontinuities might have arisen from changes in the surroundings of the weather station, the instruments used, the location of the station, or the measurement procedure (Brugnara et al., 2019).
2) Estimating consistent skin temperature uncertainties: EUSTACE used satellite data on the surface skin temperature of the land, ocean and ice, obtained from European reprocessing projects with diverse approaches to estimating uncertainty. Therefore, we derived consistent uncertainty estimates for these data over all surfaces in order to use them together effectively (Ghent et al. 2019; Nielsen-Englyst et al. 2019a).
3) Estimating air temperature from satellite data: while in some locations air temperature records can span periods of a century or more, in many areas there is a lack of information. EUSTACE has helped to provide daily air temperature information by using temperature estimates from satellite measurements to boost the amount of information beyond that already available from weather station records and ships (Nielsen-Englyst et al. 2019; Høyer et al. 2018; Kennedy and Kent, 2019).

4) Understanding the role of lakes: a number of EUSTACE studies explored various aspects of the relationship between lake surface water temperature and air temperature, demonstrating the place of lakes in the global climate system, their response to climate change and the importance of using spatially-resolved data to explore aspects of the response of lakes to climate change (Woolway and Merchant, 2017; 2018; Woolway et al. 2017b, c, d; 2018b).
5) Estimating complete fields: EUSTACE used cutting-edge statistical methods to exploit the links between air temperature in different places and through time to estimate daily air temperatures in places and at times when neither direct measurements, nor estimates from satellite were available

Hereafter, we will briefly discuss these activities, together with the independent validation of EUSTACE products.

Detecting and correcting for non-climatic discontinuities in weather station series

Most instrumental temperature series suffer from non-climatic artefacts (i.e. discontinuities or "breaks"; e.g., due to the relocation of weather stations, changes in the instrument shelter, changes in observation practices) which often result in sudden changes in the time series (e.g. Peterson et al., 1998; Brandsma and Können, 2006). Changes like this are not often adequately documented, so we need to use an automated method to detect them that we can apply to a global dataset. Correcting for these changes is termed “homogenisation”. Until recently, homogenisation efforts have mostly addressed the monthly or annual time scales and have only adjusted shifts in the mean value. This is not sufficient when dealing with daily data as inhomogeneities can affect not just the mean, but the entire distribution of variables (Trewin, 2013). The effects of, for example, shelter changes on temperature depend non-linearly on the ambient weather conditions such as sunshine and wind.

Homogenisation of daily and sub-daily data has received more attention in recent years (e.g. Aguilar et al. 2008), but efforts are still rare compared to work on monthly data (Venema et al. 2012). Existing methods correcting daily or sub-daily temperature data can be grouped into three basic categories:

1) Corrections of the mean: Methods that start from monthly mean break sizes (i.e. sizes of non-climatic discontinuities), which are then distributed to individual days. Daily corrections are computed by fitting a spline or piecewise linear function between monthly mean corrections (e.g. Vincent et al. 2002). This is the easiest approach, but comes with a risk that the tails of the distribution would not be properly corrected.
2) Corrections of higher order moments of the distribution: Methods that directly adjust the distribution of daily temperature based on a daily reference series (e.g. Trewin, 2013). This is better suited for extremes, but it requires longer and better correlated reference series than method 1).

3) Methods that incorporate basic physics such as the effects of radiation and ventilation on the temperature shield (e.g., Auchmann and Brönnimann 2012). This requires detailed metadata that are not usually available for large datasets.

Until quite recently, no global dataset of homogenised daily land surface air temperature was available. Corresponding homogenisation work was restricted to a few regions such as Canada (Vincent et al. 2002), the Mediterranean region (e.g., Brunet et al. 2006, Kuglitsch et al. 2009), Australia (Trewin, 2013) and China (Xu et al. 2013).

Most break-detection methods require highly correlated reference series. However, a non-climatic network-wide break point (e.g., the simultaneous introduction of new instruments) can be difficult to detect if reference series are from the same network. For global studies, only unhomogenised daily temperature data have been available through GHCN-Daily and other sources, which are not suitable in all locations for analysing trends in extremes, for example. Berkeley Earth have produced an experimental gridded daily temperature product over land (see a description of their method in Rohde et al. (2013a; b)), but their homogenised daily station series are not available and the analysis was constructed without directly homogenising daily data. Rather, Rohde et al. (2013 a; b) constructed fields of daily anomalies (from their monthly mean values) and added them to the existing homogenised monthly dataset.

EUSTACE has combined multiple break-detection algorithms (those of Caussinus and Mestre (2004), Toreti et al. (2012), and Wang (2008)). We applied them either to annual and semiannual averages of differences between each station and neighboring reference series (our relative tests; all methods used), or to the averages of the target station alone (our absolute test; Wang (2008) only used), in the absence of neighboring stations or if available reference series are not suitable (Brugnara et al. (2019) provides details). Using multiple methods of detecting discontinuities provides an ability to assess the robustness of the results. Figure 2 illustrates the coverage of the EUSTACE station dataset and indicates the type of break detection method applied to each station (relative or absolute) and also where application of the break detection methods has not been possible because of insufficient record length (i.e., less than 10 years). A simple likelihood index is formed from a 50-member break detection ensemble and users of the EUSTACE global station dataset can select a likelihood threshold appropriate to their needs, such that the detection power is maximised whilst minimising the false alarm rate. This is the first global daily station dataset with estimated locations of non-climatic discontinuities and their likelihood, together with valuable metadata, e.g. on resolution of measurements.

In addition to break detection, the EUSTACE global station dataset has undergone other quality checks both on the air temperature measurements themselves and on reported station altitudes (Brugnara et al. 2019). Appendix C provides a link to the resulting dataset of daily mean, maximum and minimum temperature.

For European weather station series, EUSTACE has made adjustments, where possible, to reduce the impact of non-climatic discontinuities. Briefly, we used an iterated quantile-matching approach (an example of method type 2 above) to adjust the distributions of the measurements, not just their means, by comparing to the measurement distributions at nearby reference stations (Squintu et al. (2019a; b) give details). The homogenisation brings the distributions before and after each station change much closer together, adjusting for the non-climatic effects of such discontinuities.

Applying the quantile matching to the whole European station dataset has an impact on the apparent trends in temperature over Europe (see Squintu et al., 2019a). Sometimes, the EUSTACE corrections increase the trend and sometimes they decrease it. Where stations previously showed negative trends since 1951, they show positive trends in most cases after homogenisation; in all cases making them more consistent with their neighbouring stations.

This is the first time that a pan-European station dataset of daily data has been homogenised to reduce the impact of non-climatic discontinuities. The homogenised European station dataset is provided separately from the global station dataset and comprises part of the European Climate Assessment and Dataset (ECA&D) product. A gridded 100-member ensemble dataset available either on a 0.1^{o} latitude by 0.1^{o} longitude grid or a 0.25^{o} latitude by 0.25^{o} longitude grid, based on the homogenised station records has also been developed as a contribution to the next version of the E-OBS dataset (Cornes et al., 2018). A two-step method (documented in Cornes et al., 2018) was used to create the ensemble: (i) the daily values were fitted with a Generalised Additive Model, to capture large-scale spatial trends and (ii) the residuals from this were then interpolated using stochastic Gaussian Random Field simulation. Appendix C provides a link to the CEDA catalogue record for these datasets of daily mean, maximum and minimum temperature.

Estimating consistent skin temperature uncertainties

EUSTACE uses surface temperature retrievals over land, ocean and ice based on information gathered by infra-red satellite sensors. One of our key aims is to estimate the uncertainty in our air temperature products, so first we addressed the inconsistency in the availability of uncertainty estimates for skin temperature retrievals over different surfaces. Here skin temperature is the temperature at a few microns below the top-most surface of the land, ocean or ice.

Uncertainty in surface skin temperature retrieved from satellites arises from various sources (Merchant et al., 2015):

1) Radiometric noise in the measurements made by the satellite sensor. This is usually the simplest component of uncertainty, and a standard "uncertainty propagation" can be applied to derive the surface skin temperature uncertainty associated with any surface skin temperature retrieval, given information about the radiometric noise. There is usually no or negligible correlation of error from this source between different surface skin temperature retrievals.
2) Limitations of the retrieval process would introduce uncertainty into the surface skin temperature even if the actual radiometric measurements made had zero error. For physically-derived retrievals, this component can be isolated and estimated if representative simulations of the retrieval process are available; this is not the case

where purely empirical relationships are used. An important aspect of this component of uncertainty is that the errors are likely to be correlated in space and time, and therefore may not "average out" in a simple way when transforming data from finer to coarser spatio-temporal scales.

3) Effects that are more systematic, principally: sensor calibration (which may drift over time) and radiative transfer simulation (including the effects of imperfect instrument characterisation and incorrect surface emissivity assumptions, although sub-pixel emissivity variability over land is usually considered random despite having local, coherent structure. See Ghent et al. 2019 for further discussion of uncertainties arising from misspecification of emissivity).

In addition to the above, error is introduced into surface skin temperature estimates because of imperfect cloud detection (when infrared sensors are used, as in EUSTACE; see Bulgin et al. 2018), unrecognised atmospheric aerosol, sensor anomalies, signal contamination, geo-location error, corrupted data streams, etc. Errors arising from these contributing sources are often far from Gaussian in their distributions, with complex effects on surface skin temperature uncertainty. These uncertainties have not been quantified in EUSTACE.

For all surfaces, EUSTACE estimated uncertainties partitioned according to the correlation structure of the different contributing error sources, following the method developed by Merchant et al. (2014) and expanded in Merchant et al (2015). Uncertainties are split into those arising from uncorrelated random effects, from effects which are locally correlated (these arise from atmospheric effects and/or from uncertainties in the specification of emissivity) and from effects which are correlated over large space and time scales. The derivation of uncertainties in land surface temperature is documented in Ghent et al. (2019) and in Nielsen-Englyst et al. (2019a) for ice surface temperature. Uncertainties in sea surface temperature are as calculated by Merchant et al. (2014).

Links to EUSTACE products containing these consistently-estimated uncertainties are given in Appendix C.

Estimating air temperature from satellite skin temperature

Before we can use the satellite data to estimate air temperature, we have to understand the relationship between surface air temperature and surface skin temperature and how it varies throughout the day, by surface type and through the seasons. The challenges are different in each domain, so EUSTACE explored the relationship separately over land, ocean and ice. Based on our understanding of the factors influencing the relationship in each case, we developed multiple linear regression relationships. As well as *in situ* measurements and satellite skin temperature estimates, these use extra information to help to categorise the way the skin/air temperature relationship behaves, such as vegetation, latitude and snow cover. Inclusion of altitude was found to provide no additional skill due to a lack of high altitude weather stations, although it does affect the relationship. Wind speed has a clear influence on the relationship (Good 2016), but use of wind speed information (from a dynamical reanalysis) in the regression provided no additional skill. The changing vegetation fraction information used also acts as a proxy for some other relevant surface effects, such as urbanisation, but there was no explicit attempt here to model the impact of urbanisation. The uncertainty arising from excluded effects is also not dealt with explicitly in the error model. We withheld a pre-defined set of *in situ* measurements from the regression to use in validation of the results. We then used the regression relationships to estimate air temperature when and wherever a satellite skin temperature retrieval is available, i.e. in clear-sky conditions over the period of record.

The relationship between skin and air temperature is not straightforward; Good (2016) explores this over land. Simultaneously-measured air and skin temperature vary relative to each other over the course of a day. Depending on conditions, the skin temperature can become much warmer than the air temperature when the sky is clear, but when cloud is present, the skin temperature quickly decreases to a value close to the air temperature. The daily maxima and minima in the skin and air temperatures usually occur at different times of day and the amplitudes of their diurnal cycles are often quite different. These differences also vary with season and with location. Nielsen-Englyst et al. (2019b) found a very different relationship over ice-covered surfaces in Greenland with the closest coupling between skin and air temperature occurring at noon in the summer under clear skies, when the sun warms the surface. At other times, particularly in darkness, the surface is often colder than the air above it through radiative cooling and the formation of a surface inversion layer. Under overcast skies, the surface can become warmer than the overlying air during more of the day. Spatial mismatches between satellite retrievals and *in situ* measurements mean that care needs to be taken on the resolution of satellite data used to develop the relationships. Consequently, we train our regression over land on skin temperature at 0.05^{o} latitude by 0.05^{o} longitude resolution, as the relationship with air temperature has been shown to peak at this resolution (Sohrabinia et al. 2014). Weather stations were preferentially selected for model training if their land cover type matched the dominant land cover type in the surrounding 5^{o} latitude by longitude area. Retrievals from infrared sensors are only available in clear sky conditions, so we might expect that to bias our understanding of the relationship. By using *in situ* measurements from both clear and cloudy conditions, we mitigate the impact of this (see Høyer et al. 2015; Nielsen-Englyst et al., 2019a; Kennedy and Kent, 2019 for details on the relationships between skin and air temperature across different surfaces).

Once a regression relationship has been derived, that relationship is used to estimate air temperature where we have skin temperature retrievals. We perform this procedure separately over land, ocean and ice and build up a global picture of air temperature based on the available satellite measurements (see an example in Figure 3). Global regression coefficients are used over land. Here, the estimation is most challenging, largely due to a lack of representative station measurements, in high altitude regions (for both daily minimum and maximum temperature) and at high latitudes and/or with high snow cover (for daily maximum).

Since we previously estimated our skin temperature retrieval uncertainties arising from components with different correlation structures, when we propagate those through the regression-based air temperature estimation together with the uncertainties inherent in the estimation, we can also derive components of uncertainty in the air temperature estimates arising from random, locally-correlated and systematic effects. This means that the uncertainties in our air temperature estimates are also estimated consistently across the different surfaces and can be propagated appropriately through an application.

EUSTACE air temperature estimates from satellite are provided on a 0.25^{o} latitude by 0.25^{o} longitude grid in separate files for each surface (land, ocean and ice). Daily mean temperatures are provided over ocean and ice and daily maximum and minimum is provided over land. Appendix C provides access information.

Understanding the role of lakes

EUSTACE has undertaken work using both lake surface water temperature from satellites and from *in situ* measurements gathered by the project to better understand the relationship between lake surface water temperature and near surface air temperature.

Lakes can show an amplified response of summer surface water temperature to near surface air temperature variability over the lake. This amplification of response is variable, but greater for cold lakes (e.g., those situated at high latitude and high elevation) and for deep lakes (Woolway and Merchant, 2017). Over-lake atmospheric boundary-layer stability is found to be more frequently unstable, with over-lake air temperature lower than lake surface water temperature, at lower latitudes (Woolway et al., 2017b). In summer, the frequency of unstable conditions decreases with increasing lake area, as a result of an increase in wind speed with lake size, affecting heat and carbon fluxes between the atmosphere and the lake. A study of Central European lakes shows variable warming rates across the year, but these lakes have warmed most in spring with significant trends seen over the last few decades (Woolway et al., 2017c). Abrupt changes seen in these lakes in the 1980s are consistent with abrupt changes in air temperature at the same time. Warming trends seen across nineteen large Northern Hemisphere lakes (Woolway and Merchant, 2018) vary significantly across lakes as well as between them. Deeper areas of large lakes exhibit longer correlation time scales of lake surface water temperature anomalies and a shorter stratified warming season. Deep areas of large lakes consequently display higher rates of increase of summer lake surface water temperature.

Wind speed has a substantial impact on stratification of lakes, which can have a greater influence than air temperature (Woolway et al. 2017d), and is a controlling factor on lake-air turbulent heat fluxes. Variations in turbulent heat fluxes over lakes have a marked seasonal cycle in some cases, with heat loss higher over large lakes and at low latitudes (Woolway et al., 2018b). The relative contribution of latent and sensible heat fluxes to the total heat flux differs between lakes and with latitude.

The relationship between lake surface water temperature and near surface air temperature is a two-way interaction. Air temperature influences lake temperature (via its role in turbulent fluxes) and the presence of a lake has an impact on the air temperature in its vicinity; an impact that metaphorically has some "memory" of earlier air temperature anomalies by virtue of the thermal inertia of the lake. The lake influence can be substantial, and in some instances be in excess of 2°C. In some regions, in particular where lakes are abundant (e.g., Northern Europe), their influence on the surrounding climate needs to be considered. For EUSTACE, the key question is how the lake modifies the dynamics over time of the daily minimum, maximum, and mean air temperature in its vicinity. EUSTACE has estimated the region of influence of lakes globally, provided in the Supplemental material to facilitate the inclusion of this effect in future air temperature analyses.

Estimating more-complete fields

Having used surface skin temperature retrievals over all surfaces of Earth to estimate near surface air temperature, we have global, but not globally-complete, fields covering the last few decades. Gaps remain due to the impact of clouds on the satellite estimates, for example. We also have over a century and a half of spatially-incomplete data from ships and weather stations. Night-only ship data were used, to avoid daytime biases, and adjusted to represent air temperature at 2 m following Kent et al., 2013. To try to complete the picture, we needed to use statistical modelling to capture information on how temperature covaries between locations. This information is contained in both the satellite estimates from the recent past and the weather station and ship measurements (Woodruff et al. 2011). The statistical modelling helps us understand unobserved regions on any given day.

The state-of-the-art in the spatial statistics research community was previously far ahead of the methods that had been introduced to the Earth sciences, both in terms of generality and computational efficiency. In particular, methods capable of propagating uncertainty from multiple input data sources and realistic modelling of uncertainty due to spatial variability had seen only very limited use in the Earth sciences.

Current methods for spatial interpolation in Earth sciences that also include statistical uncertainty estimates fall mainly into two categories: low-dimensional function representations (e.g. Banerjee et al., 2008, Wikle, 2010), and local covariance-based kriging methods (e.g. Furrer et al., 2006). Given a realistic computational effort, none of these approaches provide full quantification of uncertainties on long and short spatial and temporal scales simultaneously; low-dimensional basis methods cannot capture small-scale variability and dealing with statistical non-stationarity is challenging for covariance-based methods. New techniques for statistical spatio-temporal models have been developed recently by combining numerical methods for stochastic partial differential equations (SPDEs) with efficient Bayesian computations for Markov random fields. When combined with methods for fast computations for hierarchical statistical models (e.g., Rue et al., 2013) they can handle multiple scales as well as non-stationarity (Lindgren et al., 2011, Bolin and Lindgren, 2011), for a cost similar to that of low-dimensional models. Previously, these methods have successfully been used in ecology, epidemiology, and geology, but not until now for datasets of the size and resolution of global historical daily temperature datasets. EUSTACE development has made extensive use of these methods to create a global daily mean air temperature analysis on a 0.25° latitude by 0.25° longitude grid.

We model daily mean air temperature measurements, first, as an average of each day's maximum and minimum temperature and, second, as a combination of the true temperature plus bias terms (including accounting for locally-correlated biases in the air temperature estimates from satellite) and other errors affecting each measurement type. We then assume that the true daily mean air temperature can be modelled as a linear combination of three different components: a moving long-term average climatology; a large-scale component representing inter-annual variability and a daily, weather-related component. Each component is modelled as a linear combination of Gaussian variables and is solved conditioned on the other components, starting with the climatology. The solution is improved iteratively starting with the climatology, followed by the large-scale and then the local component, moving from the broadest and slowest scales, to the shortest and fastest. The process is then repeated. The estimation of the climatology component benefits directly from the inclusion of satellite-derived data. The time-variation of the large-scale component is informed largely by the long-term *in situ* measurements from ships and weather stations. The correlations captured by the local component benefit from both the satellite-derived and *in situ* data. Different types of errors in the input measurements are associated with the individual component to which they are most relevant. For example, station biases arising from non-climatic discontinuities are associated with and estimated as part of the large-scale component, because breaks in the station series are identified at an annual resolution. To make the computation tractable, we use a combination of local linear basis functions. These basis functions combine to describe variation in space (for the daily component) and, in some cases, also in time (for the large-scale component). The basis functions are defined on a nested triangular mesh which also helps to speed up the computation. This Bayesian method allows us to represent uncertainty in the process by drawing samples from the posterior distributions of the model components. Figure 4 illustrates the additional information this generates and the uncertainty in different components of the process for 1 January 2006.

We generate ten samples of possible representations of mean near surface air temperature for each day from 1 January 1850. The usefulness of the complete field is determined strongly by the availability of measurements to constrain the analysis. Therefore, where we have estimated values which add no additional information (as defined by climatology or large-scale uncertainties greater than a threshold), we mask these out of the analysis (white areas in top right panel of Figure 4). In addition, in a few limited areas the statistical model produced extreme climatological values; these were also masked. Consequently, the analysis is not globally-complete.

The purpose of EUSTACE is to provide information on daily near surface air temperature to enable assessments of vulnerability to its daily variations, rather than for monitoring of large-scale changes on longer timescales. Nonetheless, it is important to know how the global analysis compares to data sets developed for large-scale monitoring. The upper panels of Figure 5 shows regional annual average near surface air temperature anomaly in the EUSTACE global analysis v1.0 since 1850 for Europe and North America, together with the same quantity in: a blend of CRUTEM4 (Jones et al., 2012) and HadNMAT2 (Kent et al., 2013); NOAAGlobalTemp (Smith et al., 2008; Vose et al., 2012); GISTEMP (Hansen, 2010); and Berkley Earth (Rohde et al., 2013a and b). From 1895 onwards, the data sets agree closely. Prior to 1895, there are very few daily station measurements in the EUSTACE global station data set, so the EUSTACE analysis v1.0 relies on night marine air temperature to infer values over Europe. This causes a discrepancy in the EUSTACE analysis when compared to the global surface temperature monitoring data sets, which are themselves instead based on monthly weather station values. Monthly average data are more plentiful for the late nineteenth century, having been digitised separately from daily values. Over North America, the agreement is good back to 1870.

More pertinent to the aims of EUSTACE is the ability of the global analysis v1.0 to represent the evolution of daily near surface air temperature at a particular location. Having withheld a large number of station records from the development of the analysis, we can examine how the analysis compares to these records over the course of example years. The lower panels of Figure 5 show this for Cimbaj, Uzbekistan in 1975 and for Fort Nelson, Canada in 2003. The station records for these locations were not included in the analysis so provide an independent comparator. The uncertainty in the analysis is larger for Cimbaj than for Fort Nelson (shown by the envelope around the EUSTACE analysis v1.0 time series). Nonetheless, in both locations, the analysis compares well on a day-to-day basis with the record of daily mean near surface air temperature from GHCN-D v3.26. In particular, we see that the gaps in the Fort Nelson record for 2003 are completed by the EUSTACE analysis method, which uses information from other weather station records and air temperature estimated from satellite to infer the missing values.

The EUSTACE prototype global daily air temperature ensemble is openly available via the CEDA archive (see Appendix C).

Validation

The EUSTACE daily air temperature estimates (both the air temperatures estimated from satellite and the global analysis) were matched with withheld validation measurements from land stations, ice stations, moored buoys, ships and ice buoys. These data were excluded from both the derivation of regression relationships between skin temperature retrievals from satellite and air temperature and from the production of the global daily analysis fields. Veal et al. (2019a) presents the full evaluation, but Figure 6 summarises the results for the EUSTACE global analysis.

Over ocean, the EUSTACE global analysis v1.0 performs well over the period 1850-2015, with a global median discrepancy (robust standard deviation, RSD) of +0.00 K (1.76 K) against withheld ship measurements (Woodruff et al., 2011) adjusted to a height of 2 m. The highest discrepancies (analysis minus validation data) are found in the Southern Ocean, although matchups are sparse here. The global analysis also performs well in most land regions with a global median discrepancy (RSD) against weather station measurements of -0.23 K (1.76 K), however seasonal median discrepancies over central Asia are high, 6-10 K in winter at some stations (these most erroneous data have been masked out of the final product). Over permanent ice domains, the global analysis performs less well, especially over sea-ice: regional median discrepancies (RSDs) against ice buoy data are +1.19 K (4.60 K) in the Arctic and +4.76 K (6.81 K) in the Antarctic; note that these latter two statistics are affected by the sparsity of *in situ* measurements against which to compare the EUSTACE analysis in these regions, but are dominated by a drift over the Poles in the analysis which has largely been masked out of the final product. The regional median discrepancies (RSDs) over land-ice (including the Antarctic ice-shelf) against weather station data are lower: +0.37 K (4.04 K) in the Arctic and +0.47 K (2.68 K) in the Antarctic.

In addition, estimates of uncertainty are also evaluated using the withheld data. The uncertainty estimates are assessed by first binning the matchup discrepancies by the value of the uncertainty on the EUSTACE temperature estimate. Matchup statistics (median and RSD of the matchup discrepancies) are calculated for each bin. The matchup discrepancy has contributions from the uncertainty in the *in situ* reference data as well as the uncertainty on the EUSTACE temperature estimate. There is also a contribution from matching two different spatial scales, i.e. a point *in situ* value with the EUSTACE 0.25^{o} grid box estimate. The expected match up variance can be modelled as the sum of the squares of these contributions. The actual and modelled matchup discrepancy variances are plotted in Figure 7. Assuming our estimates of the uncertainty in the reference data and the matchup process are good then, if the EUSTACE uncertainty estimates are also good, for each bin the matchup RSD (blue bar) should match the modelled value (dashed line). If the blue bars are higher than the dashed line then the matchup discrepancy RSD exceeds the modelled value, indicating that the EUSTACE uncertainty estimate is too low. The uncertainty estimates for the EUSTACE global analysis v1.0 show little agreement with expectation over ocean (overestimated and showing little variation with actual discrepancy), but good agreement over land. Since the EUSTACE analysis validates extremely well in comparison to withheld data over the ocean, this mitigates the impact of the less-effective uncertainty estimates here. Analysis uncertainties are underestimated over ice regions, particularly in the Northern Hemisphere and over Southern Hemisphere land ice; here, this arises from assumptions in the analysis method about the correlation structure of errors in the over-sampled air temperature estimates from satellite.

The EUSTACE matchup data base is available for non-commercial use (see Appendix C for details).

Priorities for future work

EUSTACE relies on good retrievals of surface skin temperature from infrared satellite instruments. Adequate removal of values contaminated by cloud between the surface and the sensor is crucial for accurate skin temperature retrieval, but also for correct estimation of uncertainties and for accurate estimation of air temperature from skin temperature. The skin temperature datasets currently used in EUSTACE are sporadically contaminated by uncleared clouds. Use of improved satellite retrievals will improve the EUSTACE products.

As a proof-of-concept, EUSTACE has demonstrated that inclusion of air temperatures estimated from satellite enables the more-stable estimation of the climatological component of the global analysis (where biases in air temperature estimates from satellite are not large or there are sufficient *in situ* measurements to inform their correction), as compared to use of *in situ* measurements alone. Use of longer satellite datasets would improve the amount of information available to the analysis and improve results further. Since the inputs to the EUSTACE analysis were fixed, more satellite data have become available (i.e. version 2 of the Arctic and Antarctic Ice Surface Temperatures from thermal infrared satellite sensors (AASTI) dataset over ice, Globtemperature land surface skin temperature from a further Moderate Resolution Imaging Spectroradiometer sensor, and stable sea surface temperatures from the Advanced Very High Resolution Radiometer series in the ESA SST CCI v2.1 dataset).

With more satellite skin temperature information would come the possibility of developing and applying regionally-varying regression relationships over land. EUSTACE air temperature estimates from satellite over land currently employ a global relationship determined by latitude, snow cover and fractional vegetation cover; this results in some (sometimes large) regionally-varying biases in the resultant air temperature estimates, which are reduced in the global analysis through the additional statistical modelling undertaken there and the inclusion of measurements made *in situ*.

Interactions with users have demonstrated that information on daily maximum and minimum temperatures are needed in addition to the daily mean. Although EUSTACE undertook modelling work to enable the production of a global analysis of maximum and minimum via the mean and the diurnal temperature range, it proved impossible to pull it through into production within the timeframe of the project. Methods developed demonstrate promise and have applicability beyond surface temperature diurnal temperature range to other non-Gaussian variables. These prototyped methods would also enable full propagation of components of uncertainty with different correlation length scales through to the final analysis; the current EUSTACE global analysis simplifies the assumptions made to enable the calculations, but consequently results in underestimated uncertainties, especially over polar regions where satellite data are plentiful.

Pull-through of the lake influence mask (see Supplemental material) as a covariate (as distance from coast or altitude are currently specified) in the EUSTACE global analysis has the potential to improve the air temperature fields local to large lakes (with an influence on the scale of the EUSTACE grid box or larger, i.e. 0.25^{o} in latitude and longitude).

The availability of daily measurements made *in situ* could be increased substantially by continuing the current international data rescue and digitisation efforts (see Brönnimann et al. (2018), for example) and by making these and other daily measurements openly available. Each new set of digitised data has the potential to improve a global analysis of air temperature by better constraining the statistical modelling, particularly when targeted to regions currently under-represented in the EUSTACE global station dataset (see Figure 2) or in under-sampled areas of the ocean, such as the Southern Ocean (Brönnimann et al. (2018)).

In the course of our work, we have identified the following needs to extend the current observing system: more simultaneous Voluntary Observing Ship measurements of sea-surface and near-surface air temperature (because the network is declining and provides the only means of measuring near-surface air temperature over ocean globally) and more weather station measurements of near-surface air temperature in certain surface regimes (e.g. desert, deep forest, ice, high elevation, high latitude) to both better define the relationship between skin and near-surface air temperature there and provide more data for validation.

Summary and conclusions

The potential for future improvements outlined above notwithstanding, EUSTACE has produced a number of novel outcomes:

- a global daily station dataset with estimated locations of non-climatic discontinuities and their likelihood;
- a pan-European station dataset homogenised to reduce the impact of non-climatic discontinuities and gridded ensemble analyses for Europe;
- consistently-estimated components of uncertainty in satellite skin temperature retrievals over different surfaces of Earth;
- air temperature estimates from satellite for each surface (land, ocean and ice) with propagated uncertainty components;

- a deeper understanding of the role of lakes in responding to and influencing surrounding surface air temperature;
- a global, multi-decadal daily analysis of surface air temperature incorporating both measurements made *in situ* and estimated from satellite data; and
- validation of products using withheld reference data.

These data have been made publicly available, where not restricted by source data licenses, both for direct use and to form the basis of future onward developments (see Appendix C for details).

APPENDIX A

The EUSTACE team

The EUSTACE consortium included 9 organisations:

1) Met Office (United Kingdom)
2) The University of Reading (United Kingdom)
3) Science and Technology Facilities Council (United Kingdom)
4) University of Leicester (United Kingdom)
5) Koninklijk Nederlands Meteorologisch Instituut-KNMI (Netherlands)
6) University of Bern (Switzerland)
7) University of Bath (United Kingdom)
8) Danmarks Meteorologiske Institut (Denmark)
9) University of Edinburgh (United Kingdom)

An External Expert Advisory Board comprised: Prof. Peter Thorne (University of Ireland, Maynooth); Dr. Elizabeth Kent (National Oceanography Centre, Southampton); and Prof. Doug Nychka (National Centers for Atmospheric Research and Colorado School of Mines).

APPENDIX B

EUSTACE input data

The EUSTACE data products are based on a number of input data sources, summarised in Tables A1-A3.

Table A1 here

Table A2 here

Table A3 here

APPENDIX C

EUSTACE products

The EUSTACE data products have been catalogued in the Centre for Environmental Data Analysis (CEDA) archive, with individual download pages pointing to the data. Two products, the European homogenised data and the gridded European dataset, which also form part of the European Climate Assessment & Dataset (ECA&D) are made available separately via ECA&D.

The EUSTACE data products and their availability and licenses are summarised in the table below.

Table A4 here

Data are made available on an open license (Open Government Licence http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/) where possible. For the station datasets and the matchup data base, this was not possible due to the licensing conditions of the input datasets, which meant they could only be made available for non-commercial use. These have been made available under a non-commercial license (Non-Commercial Government http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/version/2/).

In addition, EUSTACE has produced:

- User requirements reports;
- Product user guides, including detailed guidance on uncertainties and information content in the products; and
- Peer-reviewed journal articles.

Links to all of these can be found on the EUSTACE website (https://www.eustaceproject.org).

Acknowledgments

The EUSTACE project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 640171. We thank Professor Doug Nychka for his advice throughout EUSTACE as part of the External Expert Advisory Board and Gary Corlett for validation of interim versions of the EUSTACE air temperature estimates over ocean. We also thank Esther Conway for her work on the data management plan and Roy Mandemakers for provision of a step-by step-guide for using the Climate4Impact portal as part of the EUSTACE user guides. ECK contribution was funded under NERC grant NE/J020788/1. We thank three anonymous reviewers for their comments which improved the manuscript.

References

Aguilar, E., M. Brunet, J. Sigró, F. S. Rodrigo, Y. L. Rico, and D. R. Alvarez, 2008: Homogenization of Spanish temperature series on a daily resolution: A step forward towards an analysis of extremes in the Iberian Peninsula. Proc. Seventh European Conf. on Applied Climatology, Amsterdam, the Netherlands, European Meteor. Soc., A-00-697.

Auchmann, R and S. Brönnimann, 2012: A physics-based correction model for homogenizing sub-daily temperature series. J. Geophys. Res. 117, D17119, https://doi.org/10.1029/2012JD018067

Banerjee, S., Gelfand, A. E., Finley, A. O., Sang, H., 2008: Gaussian Predictive Process Models for Large Spatial Datasets. J. Roy. Statist. Soc. B, 70(4), 825–848.

Bessembinder, J., 2016: User requirement specification for product design, EUSTACE Deliverable 4.1, https://www.eustaceproject.org/eustace/static/media/uploads/Deliverables/eustace_d4-1.pdf

Bessembinder, J., N. Rayner, D. Ghent, K. Madsen and J. Mitchelson, 2017: Report on user requirements: results from second round of user consultations, EUSTACE Deliverable 4.9, https://www.eustaceproject.org/eustace/static/media/uploads/Deliverables/eustace_d4-9.pdf

Bolin D., Lindgren F., 2011: Spatial Models Generated by Nested Stochastic Partial Differential Equations, with an Application to Global Ozone Mapping, The Annals of Applied Statistics, 5(1), 523–550.

Brandsma, T. and G.P. Konnen, 2006: Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level, Int. J. Climatol., 26, 75-89, https://doi.org/10.1002/joc.1236.

Brönnimann, S., and Coauthors, 2018: Observations for Reanalyses, Bull. Amer. Meteorol. Soc., 99, 1851–1866, https://doi.org/10.1175/BAMS-D-17-0229.1

Brugnara Y., R. Auchmann, S. Brönnimann, A. Bozzo, D. Cat Berro and L. Mercalli, 2016: Trends of mean and extreme temperature indices since 1874 at low-elevation sites in the Southern Alps. J. Geophys. Res., 121, 3304–3325, https://doi.org/10.1002/2015JD024582

Brugnara, Y., E. Good, A. A. Squintu, G. van der Schrier, S. Brönnimann, 2019: The EUSTACE global land station daily air temperature dataset. Geoscience Data Journal, https://doi.org/10.1002/gdj3.81

Brunet, M., and Coauthors, 2006: The development of a new dataset of Spanish daily adjusted temperature series (SDATS) (1850– 2003), Int. J. Climatol., 26, 1777-1802, https://doi.org/ 10.1002/joc.1338

Buizza, R. and Coauthors, 2018: The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of Earth System climate reanalyses, Bull. Amer. Meteor. Soc., 99, 1003-1014, https://doi.org/10.1175/BAMS-D-17-0199.1

Bulgin, C. E., C. J., Merchant, D., Ghent, L., Klueser, T., Popp, C. Poulsen, and L. Sogacheva, 2018: Quantifying uncertainty in satellite-retrieved land surface temperature from cloud detection errors, Remote Sensing, 10 (4). 616. ISSN 2072-4292, https://doi.org/10.3390/rs10040616

Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set, J. Geophys. Res., 111, D05101, https://doi.org/10.1029/2005JD006280

Carrea, L., and Coauthors, 2019: Lake surface temperature [in “State of the Climate 2018”]. Bull. Amer. Meteor. Soc., 100 (9), S13-S14, doi:10.1175/2019BAMSStateoftheClimate.1

Caussinus, H. and O. Mestre, 2004: Detection and correction of artificial shifts in climate series, J. R. Stat. Soc., 53(3), 405-425.

Cornes, R. C., G. van der Schrier, E. J. M. Besselaar, and P. D. Jones, 2018: An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res. (Atmospheres), 123, 9391– 9409, https://doi.org/10.1029/2017JD028200

Cristóbal, J., M. Ninyerola, and X. Pons, 2008: Modeling air temperature through a combination of remote sensing and GIS data, J. Geophys. Res., 113, D13106, https://doi.org/10.1029/2007JD009318.

Dunn, R. J. H., K. M. Willett, D. E. Parker, and L. Mitchell, 2016: Expanding HadISD: quality-controlled, sub-daily station data from 1931, Geosci. Instrum. Method. Data Syst., 5, 473-491, https://doi.org/10.5194/gi-5-473-2016

EUMETSAT, 2014: CORE-CLIMAX System Maturity Matrix Instruction Manual, 34pp.

Furrer, R., Genton, M.G., Nychka, D., 2006: Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Statist, 15, 502–523.

Ghent D., K. Veal, T. Trent, E. Dodd, H. Sembhi and J. Remedios, 2019: A new approach to defining uncertainties for MODIS land surface temperature, Remote Sens., 11, https://doi.org/10.3390/rs11091021

Good, E. J., 2016: An in situ-based analysis of the relationship between land surface "skin" and screen-level air temperatures, J. Geophys. Res. Atmos., 121, 8801–8819, https://doi.org/10.1002/2016JD025318.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change, Rev. Geophys., 48, RG4004, https://doi.org/10.1029/2010RG000345.

Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H., 2013: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset, Int. J. Climatol., 34, 623-642, https://doi.org/10.1002/joc.3711.

Hartmann, D.L., A.M.G. Klein Tank, M. Rusticucci, L.V. Alexander, S. Brönnimann, Y. Charabi, F.J. Dentener, E.J. Dlugokencky, D.R. Easterling, A. Kaplan, B.J. Soden, P.W. Thorne, M. Wild and P.M. Zhai, 2013: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Høyer, J. L., E. Good, P. Nielsen-Englyst, K. S. Madsen, I. Woolway, J. Kennedy, 2018: Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes, EUSTACE Deliverable 1.5, https://www.eustaceproject.org/eustace/static/media/uploads/d1.5_revised.pdf

Hunziker, S., and Coauthors, 2017: Identifying, attributing, and overcoming common data quality issues of manned station observations. Int. J. Climatol., 37, 4131-4143.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. J. Geophys. Res., 117, doi:10.1029/2011JD017139

Kennedy, J. J. and E. C. Kent, 2019: Estimating the climatological mean and spatial covariance of air-sea temperature differences, in prep for Geophys. Res. Lett.

Kent, E.C., Rayner, N.A., Berry, D.I., Saunby, M., Moat, B.I., Kennedy, J.J., Parker, D.E., 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. J. Geophys. Res., 118, 1281–1298, https://doi.org/10.1002/jgrd.50152

Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat, 2014: Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res., 119, 2294–2313, https://doi.org/10.1002/2013JD020803.

Klein-Tank, A. M. G., and Coauthors, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology, 22, 1441-1453.

Kuglitsch, F.G., A. Toreti, E. Xoplaki, P. M. Della-Marta, J. Luterbacher, and H. Wanner, 2009: Homogenization of daily maximum temperature series in the Mediterranean. J. Geophys. Res., 114, https://doi.org/10.1029/2008JD011606.

Kushnir, Y., A.A. Scaife and Coauthors, 2019: Towards operational predictions of the near-term climate, Nature Climate Change, 9, 94-101, http://doi.org/10.1038/s41558-018-0359-7

Jones, G.S. and J.J. Kennedy, 2017: Sensitivity of Attribution of Anthropogenic Near-Surface Warming to Observational Uncertainty, J. Climate, 30, 4677-4691, http://doi.org/10.1175/JCLI-D-16-0628.1

Lindgren, F., H. Rue, and J. Lindström, 2011: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, J. Roy. Statist. Soc. B, 73(4), 423–498, https://doi.org/10.1111/j.1467-9868.2011.00777.x

Menne, M.J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An Overview of the Global Historical Climatology Network-Daily Database. J. Tech., 29, 897-910, http://doi.org/10.1175/JTECH-D-11-00103.1

Merchant C. J., and Coauthors, 2013: The surface temperatures of Earth: steps towards integrated understanding of variability and change, Geosci. Instrum. Method. Data Syst., 2, 305–321, https://doi:10.5194/gi-2-305-2013

Merchant, C. J., and Coauthors, 2014: Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI), Geosci. Data Journal, https://doi.org/10.1002/gdj3.20

Merchant, C., D. Ghent, J. Kennedy, E. Good, and J. Høyer, 2015: Common approach to providing uncertainty estimates across all surfaces, EUSTACE Deliverable 1.2, 20 pp, https://www.eustaceproject.org/eustace/static/media/uploads/Deliverables/eustace_d1-2.pdf

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, J. Geophys. Res., 117, http://doi.org/10.1029/2011JD017187.

Nielsen-Englyst, P., Høyer J. L., Madsen, K. S., Tonboe, R., and Dybkjær, G., 2019a: Deriving Arctic 2-m air temperatures from satellite, submitted to The Cryosphere, https://doi.org/10.5194/tc-2019-126

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R., Dybkjær, G., and Alerskans, E., 2019b: In situ observed relationships between skin temperatures and 2 m air temperatures in the Arctic, The Cryosphere, http://doi.org/10.5194/tc-13-1005-2019

Osborn, T.J., P.D. Jones and M. Joshi, 2017: Recent United Kingdom and global temperature variations, Weather, 72, 323-329, http://doi.org/10.1002/wea.3174

Peterson, T.C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. Int. J. Climatol., 18, 1493-1517, http://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T

Rayner, N.A., 2019: Final verified products delivered to CEMS for verification. EUSTACE Deliverable 2.6, https://www.eustaceproject.org/eustace/static/media/uploads/d_2.6_final.pdf

Rennie, J. J., and Coauthors (2014). The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. Geoscience Data Journal, 1, 75-102.

Rohde R., R. A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, and C. Wickham, 2013a: A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinfor Geostat: An Overview 1:1, http://doi.org/10.4172/2327-4581.1000101

Rohde R., R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickham, and S. Mosher, 2013b: Berkeley Earth Temperature Averaging Process. Geoinfor Geostat: An Overview 1:2, http://doi.org/10.4172/2327-4581.1000103

Rue H., S. Martino, F. Lindgren, D. Simpson, A. Riebler, 2013: R-INLA: Approximate Bayesian Inference using Integrated Nested Laplace Approximations. Trondheim, Norway. URL http://www.r-inla.org/.

Sánchez-Lugo, A., P. Berrisford, C. Morice and J.P. Nicolas, 2019: Global surface temperature [in "State of the Climate 2018"]. Bull. Amer. Meteor. Soc., 100 (9), S11-S12, doi:10.1175/2019BAMSStateoftheClimate.1

Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperatures analysis (1880–2006). J. Climate, 21, 2283–2296, http://doi.org/10.1175/2007JCLI2100.1.

Sohrabinia, M. , P. Zawar-Reza and W. Rack, 2014: Spatio-temporal analysis of the relationship between LST from MODIS and air temperature in New Zealand, Theor. Appl. Climatol., DOI 10.1007/s00704-014-1106-2.

Squintu, A.A., G. van der Schrier, Yuri Brugnara, Albert Klein Tank, 2019a: Homogenization of daily temperature series in the European Climate Assessment & Dataset, Int. J. Climatol., 39:1243-1261. http://doi.org/10.1002/joc.5874

Squintu, A. A., G. van der Schrier, E. J. M. van den Besselaar, R. C. Cornes, A. Klein Tank, 2019b: Building long homogeneous temperature series across Europe: a new approach for the blending of neighboring series, submitted to J. Appl. Meteorol. Clim.

Stickler, A., and Coauthors, 2014: ERA-CLIM: historical surface and upper-air data for future reanalyses. Bull. Amer. Meteorol. Soc., 95, 1419-1430.

Toreti, A., F. G. Kuglitsch, E. Xoplaki, and J. Luterbacher, 2012: A novel approach for the detection of inhomogeneities affecting climate time series. J. Appl. Meteorol. Clim., 51, 317-326, http://doi.org/10.1175/JAMC-D-10-05033.1.

Trewin, B., 2013: A daily homogenized temperature data set for Australia. Int. J. Climatology, 33, 1510-1529.

Veal, K., 2019a: Validation report for the final in-filled EUSTACE surface air temperature product, EUSTACE Deliverable 3.5

Veal, K., 2019b: Intercomparison report for the final in-filled EUSTACE surface air temperature product, EUSTACE Deliverable 3.4

Venema, V., et al, 2012: Detecting and repairing inhomogeneities in datasets, assessing current capabilities. Bull. Am. Meteorol. Soc., 93, no. 7, pp. 951-954.

Vincent, L., X. Zhang, B. R. Bonsall, W. D. Hogg, 2002: Homogenization of Daily Temperatures over Canada. J. Climate, 15, 1322-1334.

Vose, R.S., and Coauthors, 2012: NOAA's merged land-ocean surface temperature analysis. Bull. Amer. Meteor. Soc., 93, 1677–1685, http://doi.org/10.1175/BAMS-D-11-00241.1.

Walters, D. and Coauthors, 2019: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. Geoscientific Model Development, 12, 1909-1963, http://doi.org/10.5194/gmd-12-1909-2019

Xu, X. and Coauthors 2019: The effects of temperature on human mortality in a Chinese city: burden of disease calculation, attributable risk exploration, and vulnerability identification. Int. J. Biometeor., 63, 1319-1329, http://doi.org/10.1007/s00484-019-01746-6

Wang, X. L., 2008: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or f test. J. Appl. Meteorol. Clim., 47, 2423-2444, http://doi.org/10.1175/2008JAMC1741.1.

Weedon, G.P., and Coauthors, 2011: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, J Hydromet, 12, 823-848, http://doi.org/10.1175/2011JHM1369.1

Wikle, C. K., 2010: In Handbook of Spatial Statistics, chapter Low-rank Representations for Spatial Processes, pp. 107–118. Chapman and Hall/CRC, Boca Raton, FL.

Woodruff, S. D., and Coauthors, 2011: ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol., 31, 951-967, http://doi.org/10.1002/joc.2103

Woolway, R. I., and Coauthors, 2016: Lake surface temperatures. [in "State of the Climate in 2015"]. Bull. Amer. Meteor. Soc., 97 (8), S17-S18, https://doi.org/10.1175/2016BAMSStateoftheClimate.1

Woolway, R.I. and C. Merchant, 2017: Amplified surface temperature response of cold, deep lakes to inter-annual air temperature variability. Nature Scientific Reports, 7, http://doi.org/10.1038/s41598-017-04058-0.

Woolway, R. I., et al., 2017a: Lake surface temperature [in "State of the Climate in 2016"]. Bull. Amer. Meteor. Soc., 98 (8), S13–S14, http://doi.org/10.1175/2017BAMSStateoftheClimate.1.

Woolway, R. I., and Coauthors, 2017b: Latitude and lake size are important predictors of over-lake atmospheric stability, Geophys. Res. Lett., 44, http://doi.org/10.1002/2017GL073941

Woolway, R.I., Dokulil, M. T., Marszelewski, W., Schmid, M., Bouffard, D., Merchant, C. J., 2017c: Warming of Central European lakes and their response to the 1980s climate regime shift. Climatic Change, 142, 505-520, http://doi.org/10.1007/s10584-017-1966-4.

Woolway, R. I., P. Meinson, P. Nõges, I. D. Jones and A. Laas, 2017d: Atmospheric stilling leads to prolonged thermal stratification in a large shallow polymictic lake, Climatic Change, 141, 759–773, http://doi.org/10.1007/s10584-017-1909-0

Woolway, R. I. and Merchant, C. J., 2018: Intra-lake heterogeneity of thermal responses to climate change: A study of large Northern Hemisphere lakes. J. Geophys. Res., 123, http://doi.org/10.1002/2017JD027661

Woolway, R. I., and Coauthors, 2018a: Lake surface temperature [in “State of the Climate in 2017”]. Bull. Amer. Meteor. Soc., 99(8), S13–S15, http://doi.org/10.1175/2018BAMSStateoftheClimate.

Woolway, R. I., and Coauthors, 2018b: Geographic and temporal variation in turbulent heat loss from lakes: A global analysis across 45 lakes. Limnol. Oceanogr., 63, 2436-2449, http://doi.org/10.1002/lno.10950

Xu, W., Q. Li, X. L. Wang, S. Yang, L. Cao, and Y. Feng, 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices, J. Geophys. Res., 118, http://doi.org/10.1002/jgrd.50791.

Tables.

Table A1. Satellite data on which EUSTACE products are based and period of data used.

Satellite instrument	Satellite programme	Variables used	Data producers
Along Track Scanning Radiometer (ATSR) series, 1991-2012	ESA	Sea surface temperature at 0.2m depth on 0.25° latitude by 0.25° longitude grid	ESA CCI SST, experimental v1.2 (A)ATSR Level 3C data product. See Appendix C for data access.
Advanced Very High Resolution Radiometer (AVHRR) series, 2000-2009	NOAA	Ice surface skin temperature on instrument swath	AASTI v1.0 dataset generated by Met Norway and DMI within the NORMAPP and the NACLIM projects. See Appendix C for data access.

Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua + Terra, 2000-2016	NASA	Land surface skin temperature on instrument swath	USGS/NASA (via ESA GlobTemperature). MODIS Collection 6 radiances downloaded from the NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center [https://ladsweb.modaps.eosdis.nasa.gov/]. See Appendix C for data access.

Table A2. Weather station air temperature measurements on which EUSTACE products are based and period of data used.

Dataset	Link	Reference
Global Historical Climatology Network – Daily (GHCN-D), version 3.22, 1850-2015	http://doi.org/10.7289/V5D21VHZ	Menne et al., 2012
International Surface Temperature Initiative (ISTI), v1.00 stage 2, 1850-2015	http://www.surfacetemperatures.org/databank	Rennie et al., 2014
European Climate Assessment & Dataset (ECA&D), 1950-2015	https://www.ecad.eu/	Klein-Tank et al., 2002
Data rescued by ERA-CLIM project, various		Stickler et al., 2014
DECADE project, 1931 onwards	http://www.geography.unibe.ch/research/climatology_group/research_projects/decade/index_eng.html	Hunziker et al., 2017
Southern Alps homogenized, 1871-2015		Brugnara et al 2016
Data from the national weather service of Argentina	Servicio Meteorologico Nacional Argentina	

Table A3. Marine *in situ* measurements on which EUSTACE products are based and period of data used.

Dataset	Link	Reference
HadNMAT2 observations, derived from ICOADS release 2.5.1, 1850-2010	http://www.metoffice.gov.uk/hadobs/hadnmat2/	Kent et al., 2013

Table A4. EUSTACE products and their access and licensing information

Short name	Descriptive name	Dataset link	License
Satellite skin temperatures			
Global satellite land surface temperature, v2.1	EUSTACE / GlobTemperature: Global clear-sky land surface temperature from MODIS Aqua on the satellite swath with estimates of uncertainty	http://catalogue.ceda.ac.uk/uuid/0f1a958a130547febd40057f5ec1c837	Open

	components, v2.1, 2002-2016		
	EUSTACE / GlobTemperature: Global clear-sky land surface temperature from MODIS Terra on the satellite swath with estimates of uncertainty components, v2.1, 2000-2016	http://catalogue.ceda.ac.uk/uuid/655866af94cd4fa6af67809657b275c3	Open
Global satellite ice surface temperature, v1.1	EUSTACE / AASTI: Global clear-sky ice surface temperature from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009	https://catalogue.ceda.ac.uk/uuid/60b820fa10804fca9c3f1ddfa5ef42a1	Open
Global satellite sea surface	EUSTACE / CCI: Global clear-sky sea surface temperature from the (A)ATSR series at 0.25	https://catalogue.ceda.ac.uk/uuid/b8285969426a4e00b74814342	Open

temperature, v1.2	degrees with estimates of uncertainty components, v1.2, 1991-2012		
Surface air temperatures from *in situ* measurements			
European station measure-ments	EUSTACE/ECA&D: European land station daily air temperature measurements, homogenised	https://catalogue.ceda.ac.uk/uuid/81784e3642bd465aa69c7fd40ffe1b1b	Non-commercial use only
Global Station Measure-ments	EUSTACE: Global land station daily air temperature measurements with non-climatic discontinuities identified, for 1850-2015	http://catalogue.ceda.ac.uk/uuid/7925ded722d743fa8259a93acc7073f2	Non commercial use only
Validation match up database, v1.0	EUSTACE: coincident daily air temperature estimates and reference measurements, for validation, 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/4b34a2c6890f4e518cacc88911193354	Non-commercial use only

E-OBS	EUSTACE / E-OBS: Gridded European surface air temperature based on homogenised land station records since 1950	https://catalogue.ceda.ac.uk/uuid/b2670fb9d6e14733b303865c85c65d	Non commercial use only
Surface air temperature estimates from statistical analysis			
Air temperature estimates from satellite, v1.0	EUSTACE: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016	https://catalogue.ceda.ac.uk/uuid/f883e197594f4fbaae6edebafb3fddb3	Open
Global air temperature estimates, v1.0	EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0	https://catalogue.ceda.ac.uk/uuid/468abcf18372425791a31d15a41348d9	Open

1256

1257

Figures

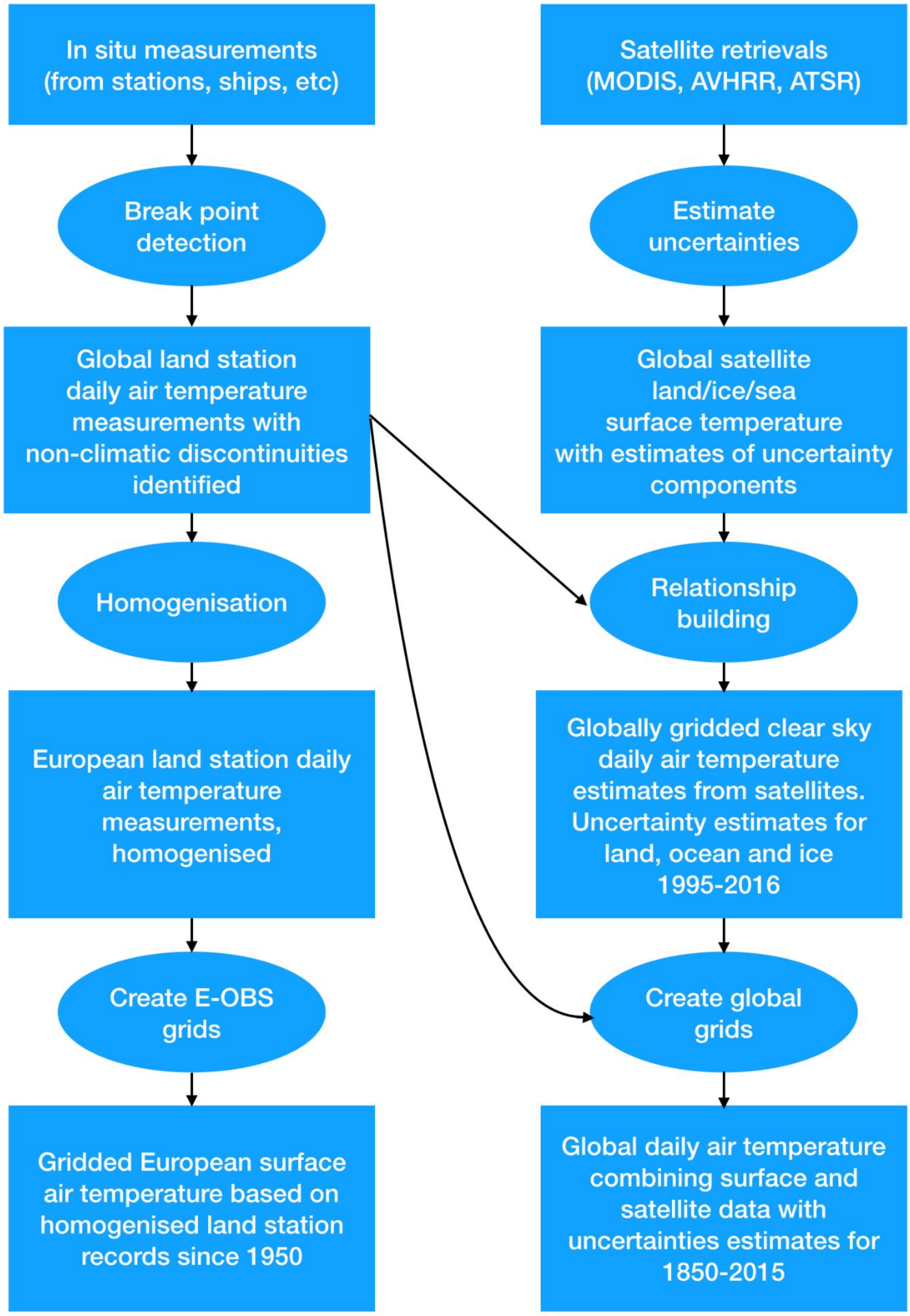
In situ measurements (from stations, ships, etc)
Break point detection
Global land station daily air temperature measurements with non-climatic discontinuities identified
Homogenisation
European land station daily air temperature measurements, homogenised
Create E-OBS grids
Gridded European surface air temperature based on homogenised land station records since 1950
Satellite retrievals (MODIS, AVHRR, ATSR)
Estimate uncertainties
Global satellite land/ice/sea surface temperature with estimates of uncertainty components
Relationship building
Globally gridded clear sky daily air temperature estimates from satellites. Uncertainty estimates for land, ocean and ice 1995-2016
Create global grids
Global daily air temperature combining surface and satellite data with uncertainties estimates for 1850-2015

Figure 1. Schematic of work undertaken in the EUSTACE project. Top-most boxes denote input data. Ovals denote new development. Other boxes denote EUSTACE products (see also Appendix C). Connections between different components are indicated by arrows.

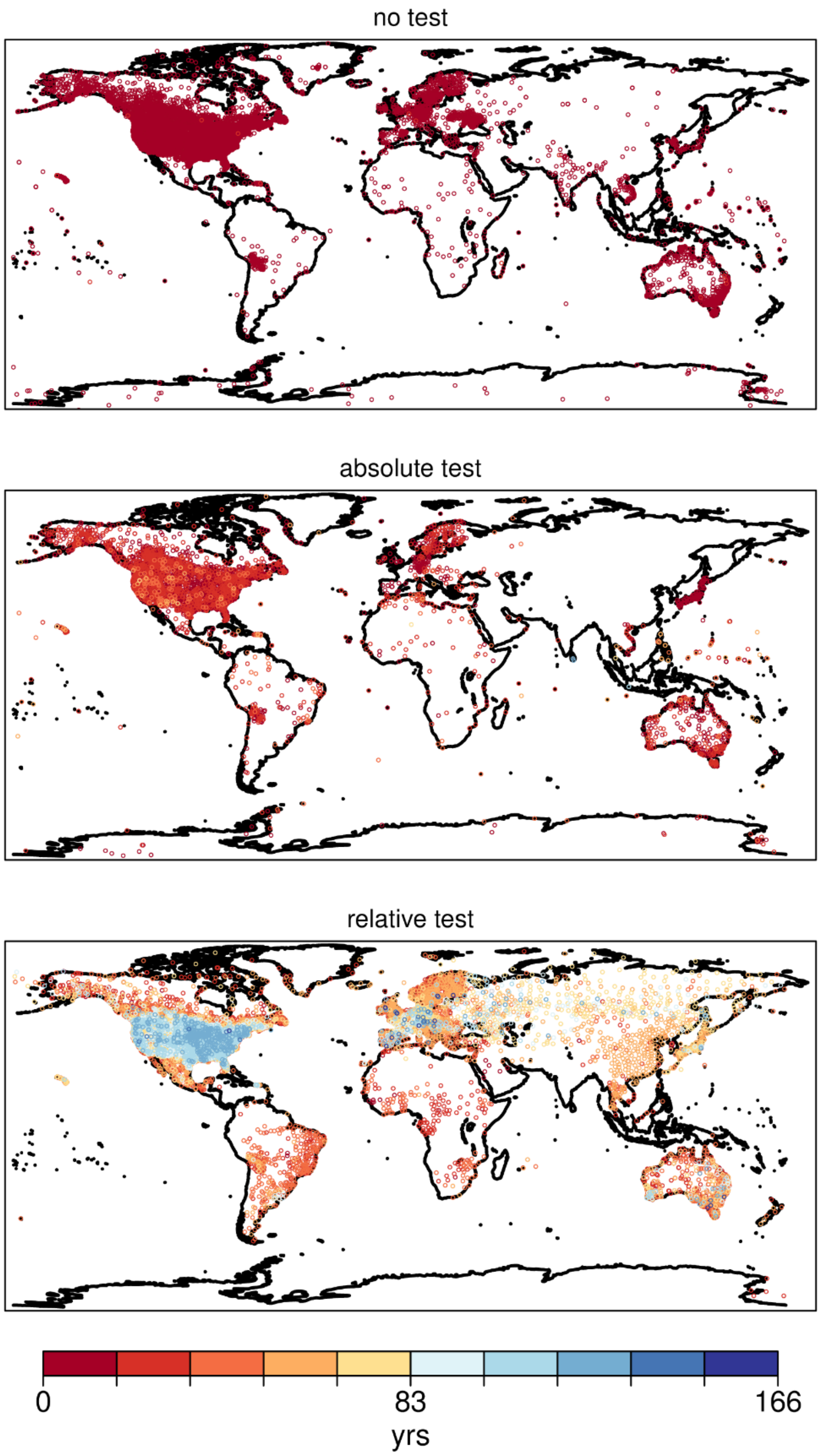

Figure 2. Map of weather stations included in the EUSTACE global station air temperature data set and break-detection tests applied (see text). Color of symbols represents length of daily surface air temperature record available. Top: no test applied. These stations are those

which have records shorter than 10 years. Middle: only absolute test applied. Bottom: relative test applied.

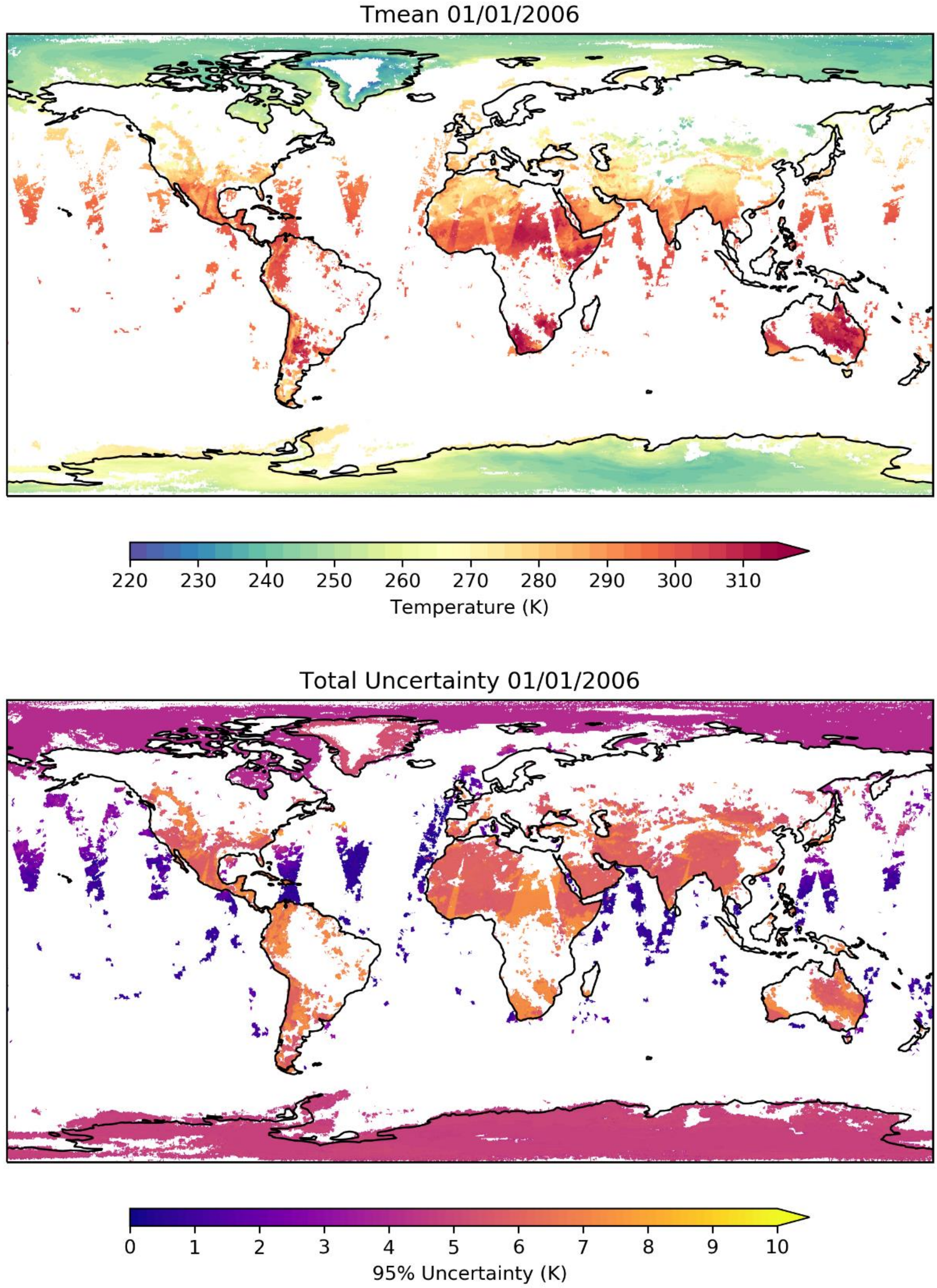

Figure 3. EUSTACE air temperature estimates from satellite. (Top) daily mean air temperatures (K) estimated for 01 01 2006. (Bottom) combined uncertainty (K).

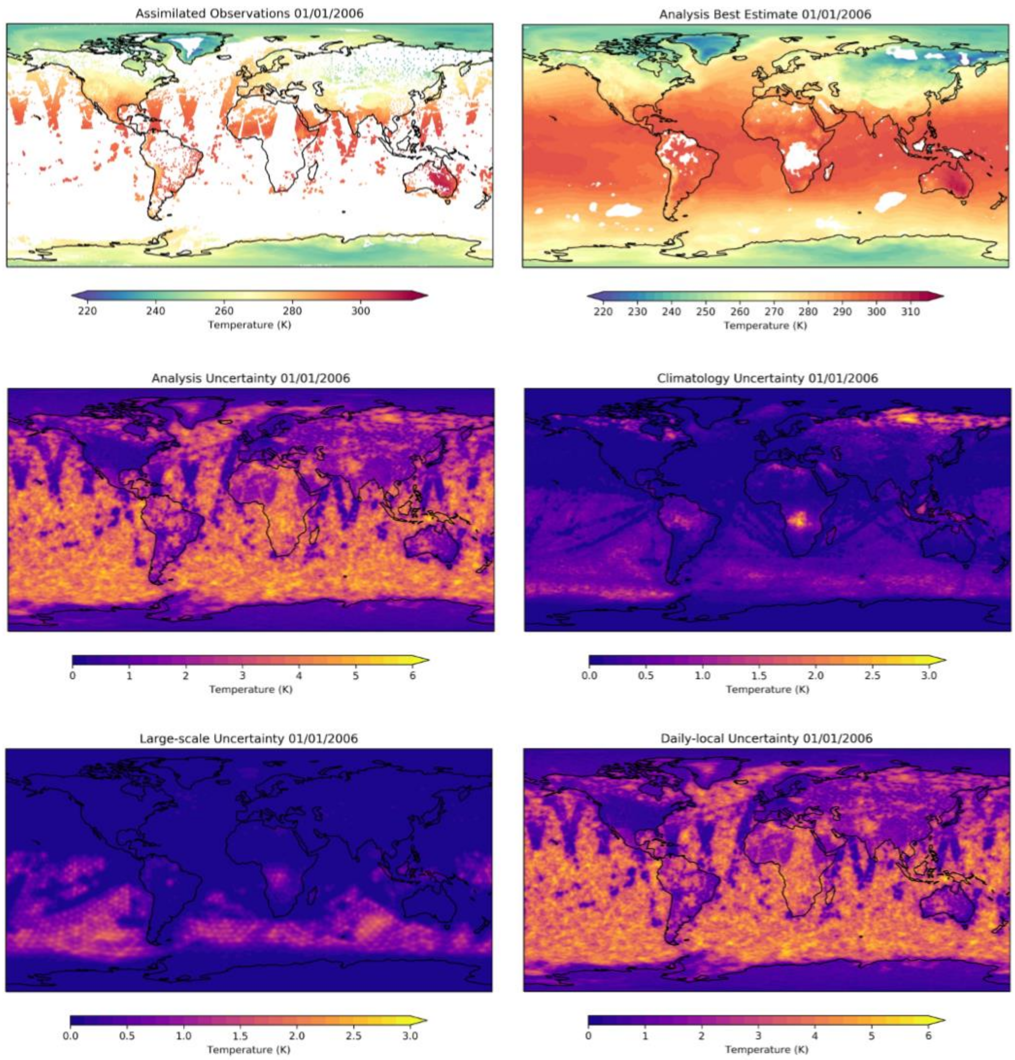

Figure 4. Air temperature (K) for 01 01 2006. Top left: input observations of air temperature (K). Top right: best guess combined *in situ* and satellite measurements from EUSTACE statistical infilling (K). Areas with climatology or large-scale component uncertainty above a threshold are masked. Middle left: total uncertainty (K) in the infilled analysis. Middle right: uncertainty (K) in the climatology component. Bottom left: uncertainty in the large-scale component (K). Bottom right: uncertainty in the local component (K).

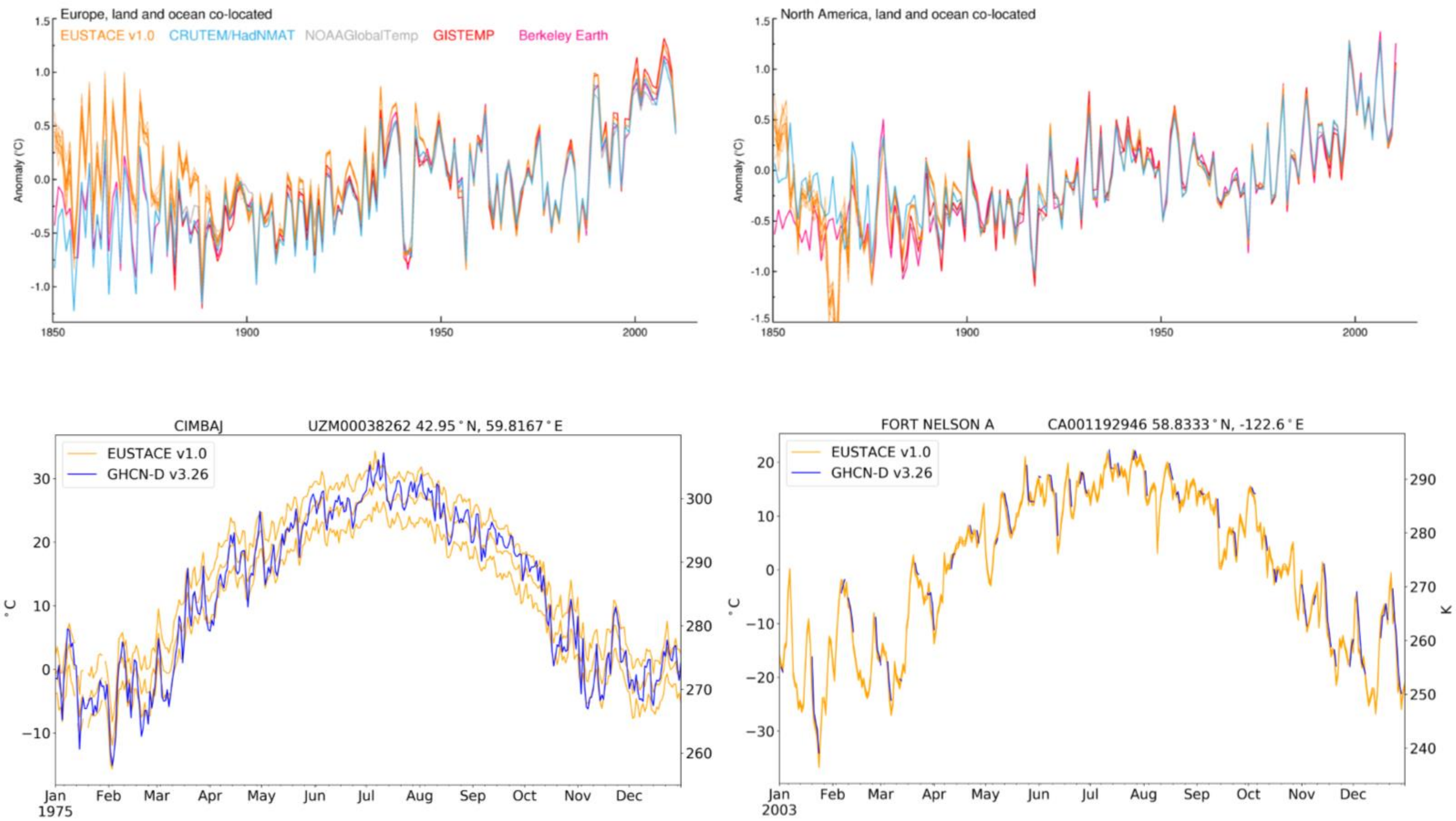

Figure 5. (Top) Annual regional average near surface air temperature anomaly (relative to 1961-1990) in a number of global surface temperature data sets, 1850-2015 (left: Europe; right: North America). Orange: EUSTACE global analysis v1.0; cyan: a blend of CRUTEM4 and HadNMAT2; grey: NOAAGlobalTemp; red: GISTEMP; pink: Berkley Earth. (Bottom) Daily near surface air temperature (K and ^{o}C) over the

course of a year (left: Cimbaj, Uzbekistan in 1975; right: Fort Nelson, Canada in 2003). Orange: EUSTACE global analysis v1.0 (ensemble mean and range); royal blue: GHCN-D v3.26 station measurements.

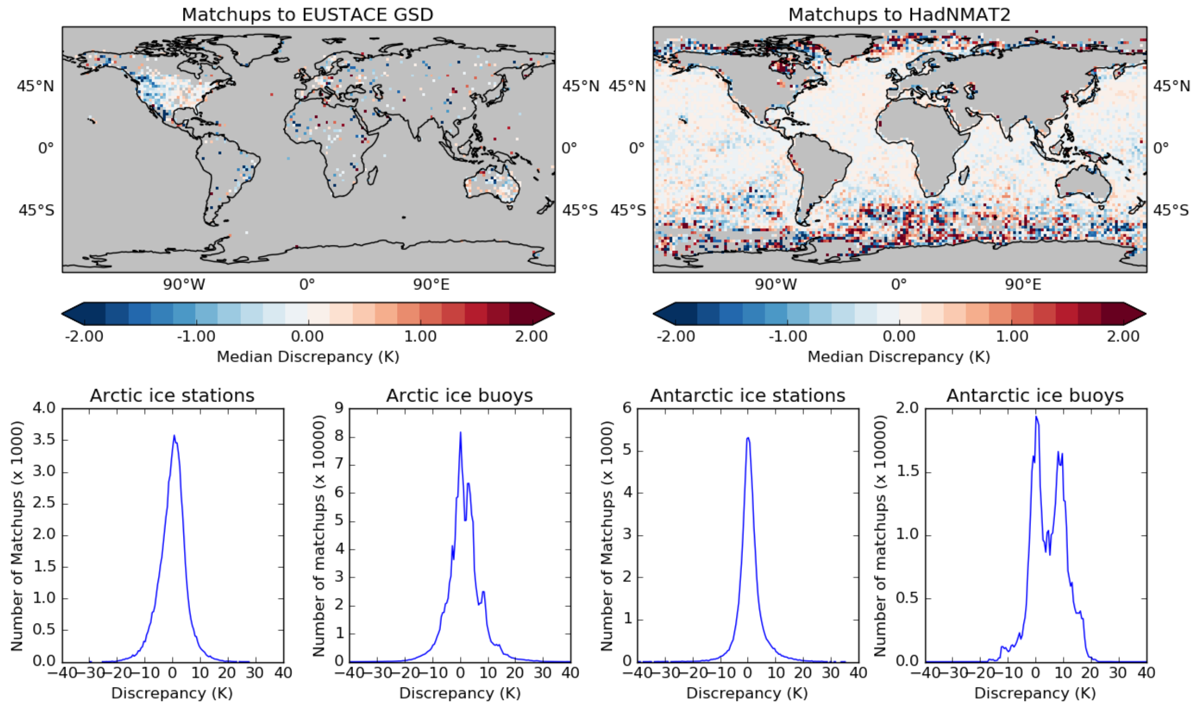

Figure 6. Validation of the EUSTACE global analysis v1.0, 1850-2015 against independent reference data. (Top left) median discrepancy (K) over land, compared to withheld station measurements. (Top right) median discrepancy (K) over ocean, compared to withheld ship measurements corrected to 2m. (Bottom row, left to right) discrepancy (K) between EUSTACE analysis and withheld reference data over ice-covered regions: Arctic land; Arctic sea ice; Antarctic land and Antarctic sea ice.

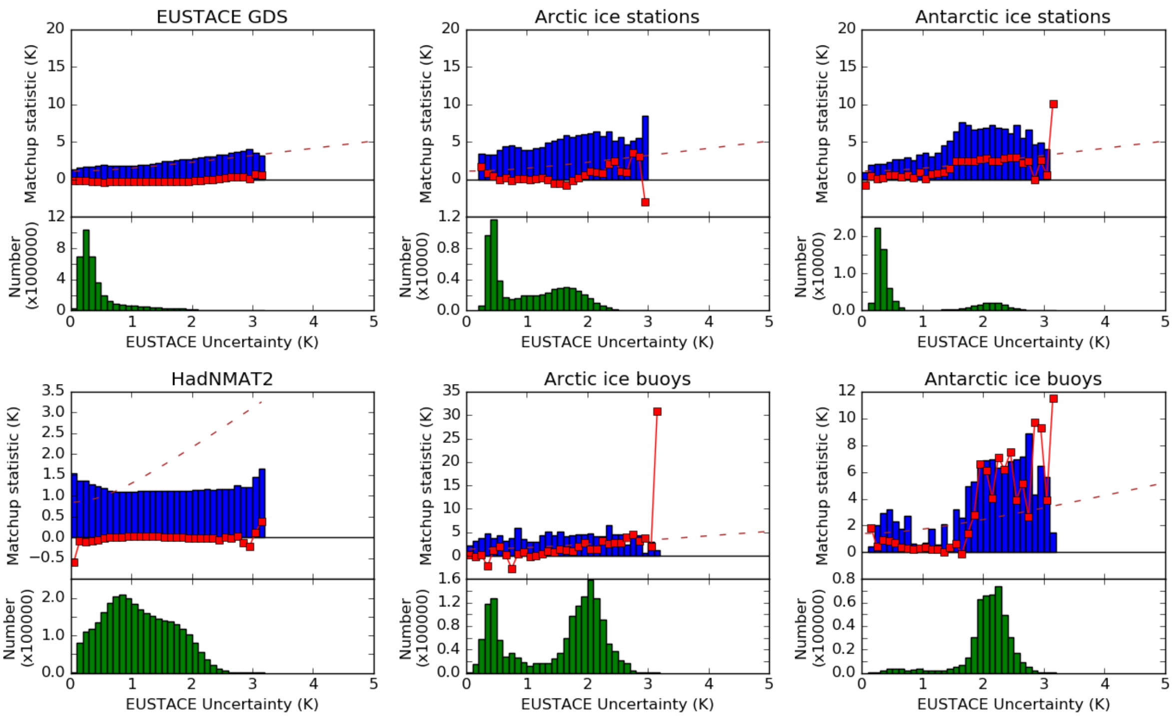

Figure 7. Validation of the uncertainty estimates for the EUSTACE global analysis v1.0, 1850-2015, against independent reference data. Top left: land; top middle: Arctic land ice; top right: Antarctic land ice; bottom left: ocean; bottom middle: Arctic sea ice; bottom right: Antarctic sea ice. Dashed line: modelled discrepancy; combined EUSTACE uncertainty and uncertainty in the validation data (K). Blue bars: robust standard deviation of discrepancies between the analysis and the validation data (K). Red line: median discrepancy (K). Green bars: number of matchups.