JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D21101, doi:10.1029/2009JD011799, 2009

Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature

Nynke Hofstra,[1,2] Malcolm Haylock,[3] Mark New,[1] and Phil D. Jones[3]

Received 26 January 2009; revised 2 July 2009; accepted 29 July 2009; published 4 November 2009.

[1] Gridded data sets derived through interpolation of station data have a number of potential inaccuracies and errors. These errors can be introduced either by the propagation of errors in the station data into derived gridded data or by limitations in the ability of the interpolation method to estimate grid values from the underlying station network. Recently, Haylock et al. (2008) reported on the development of a new high-resolution gridded data set of daily climate over Europe (termed E-OBS). E-OBS is based on the largest available pan-European data set, and the interpolation methods used were chosen after careful evaluation of a number of alternatives, yet the data set will inevitably have errors and uncertainties. In this paper we assess the E-OBS data set with respect to: (1) homogeneity of the gridded data; (2) evaluation of inaccuracies arising from available network density, through comparison with existing data sets that have been developed with much denser station networks; and (3) the accuracy of the estimates of interpolation uncertainty that are provided as part of E-OBS. We find many inhomogeneities in the gridded data that are primarily caused by inhomogeneities in the underlying station data. In the comparison of existing data with E-OBS, we find that while correlations overall are high, relative differences in precipitation are large, and usually biased toward lower values in E-OBS. From the analysis of the interpolation uncertainties provided as part of E-OBS, we conclude that the interpolation standard deviation provided with the data significantly underestimates the true interpolation error when cross validated using station data, and therefore will similarly underestimate the interpolation error in the gridded E-OBS data. While E-OBS represents a valuable new resource for climate research in Europe, users of the data need to be aware of the limitations in the data set and use the data appropriately.

Citation: Hofstra, N., M. Haylock, M. New, and P. D. Jones (2009), Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res.*, *114*, D21101, doi:10.1029/2009JD011799.

1. Introduction

[2] Gridded climate data derived from meteorological station measurements underpin a wide range of applications and research in climate science, including evaluation of global and regional climate models, the construction of bias-corrected climate change scenarios and driving many applications in climate impacts assessments [*Haylock et al.*, 2008]. Increasingly, there has been a need for gridded data at higher spatial and temporal resolutions, as the focus of climate change research has shifted from global to regional and local scales. Recently, *Haylock et al.* [2008] described the development of the first high-resolution gridded data set of daily climate over Europe (termed E-OBS), as part of the EU funded ENSEMBLES project. The data set, comprising daily mean, minimum and maximum temperature and precipitation, was constructed through interpolation of the most complete collection of station data over wider Europe [*Klok and Klein Tank*, 2008]. The data are available on four different Regional Climate Model (RCM) grids (0.25 and 0.5 degree regular latitude-longitude and 0.22 and 0.44 degree rotated pole) and cover the period 1950–2006. Additionally, estimates of interpolation uncertainties are included as part of the data set [*Haylock et al.*, 2008].

[3] Gridded data sets derived through interpolation of station data have a number of potential inaccuracies and errors. Errors in the underlying station data can be propagated into the gridded data; typical sources of error include incorrect station location information and individual erroneous values or nonclimatic breaks (inhomogeneities) in the station time series. A second source of uncertainty relates to the ability of the interpolation method to estimate grid values from the underlying station network. In general, interpolation accuracy decreases as the network density

[1]School of Geography and the Environment, Oxford University Centre for the Environment, University of Oxford, Oxford, UK.
[2]Now at Environmental Systems Analysis Group, Wageningen University, Wageningen, Netherlands.
[3]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.

0148-0227/09/2009JD011799$09.00

decreases, is less accurate for variables with more variable spatial characteristics (e.g., precipitation) and degrades in areas of complex terrain (e.g., mountain areas). While E-OBS is based on the largest available pan-European data set and the interpolation methods used were chosen after careful evaluation of a number of alternatives [*Hofstra et al.*, 2008], the data set will inevitably have errors and uncertainties.

[4] The aim of this paper is to assess the E-OBS data set with respect to some of the potential errors that may be present. Users can then familiarize themselves with the strengths and weaknesses of the data and use them responsibly. We choose three important properties of E-OBS to analyze in this paper: (1) homogeneity of the gridded data; (2) inaccuracies due to the underlying station network density, through comparison with existing data sets that have been developed with much denser station networks; and (3) the accuracy of the estimates of interpolation uncertainty that are provided as part of E-OBS.

[5] Long-term station data are often influenced by nonclimatic factors, such as changes in station location or environment, instruments and observing practices. These so-called inhomogeneities can often lead to misinterpretations of the climate data analyzed [*Peterson et al.*, 1998]. The station data used for E-OBS are not fully homogenized. Individual station series may have been homogenized by the original custodians of each series, but the series provided by partner organizations have been used directly, meaning potentially inhomogeneous stations may be contributing to the interpolated grids. As station density strongly influences the interpolation [*Hofstra et al.*, 2008], E-OBS was constructed using many potentially inhomogeneous stations, as their exclusion would degrade the station network density and hence accuracy of the interpolation. In addition, several studies explain that, for area averages of relatively large areas, inhomogeneities balance out during interpolation [*Dai et al.*, 1997; *New*, 1999; *Peterson et al.*, 1998]. However, that may not be the case for the E-OBS high-resolution grids. Therefore, the first out of three properties tested is the homogeneity of the data set.

[6] The second topic is a comparison with other gridded data sets that have been developed with much denser station networks. These data sets are available, in the case of precipitation, for long periods for the UK and the Alps and for the period October 1999 to December 2000 for Europe as a whole. For temperature, unfortunately, we have only been able to secure data for the UK. Data sets developed with denser station networks are assumed to be a better approximation of the true area averages. So if the E-OBS gridded data set produces grid area averages that are close to those calculated from the higher-quality grids, the E-OBS data set can be deemed to be a reasonable representation of the true area-average gridded values.

[7] Because of the inevitable interpolation uncertainties, the E-OBS data set has been provided with information on the interpolation uncertainty for each grid box and each day [*Haylock et al.*, 2008]. E-OBS interpolation uncertainty has been derived by combining the Bayesian standard error estimates of the monthly climatology [*Hutchinson*, 1995] and the interpolation standard deviation for daily anomalies [*Yamamoto*, 2000] (see section 5 for more detail). Here we concentrate on the interpolation standard error estimates, and evaluate the accuracy of the estimates through cross validation against station data. This represents the first evaluation of the *Yamamoto* [2000] standard error method, which has to date only been applied to geological data.

[8] The remainder of the paper is structured as follows. Section 2 provides a more detailed description of the E-OBS data set, including the underlying station data and the interpolation and gridding methodology. We then cover each of the three evaluations in turn: inhomogeneities (section 3), comparison against regional gridded data sets based on denser station networks (section 4) and evaluation of the interpolation standard error estimates (section 5). We conclude with a summary of results and a discussion of the implications of our assessment for use of the E-OBS data set.

2. E-OBS Data Set

[9] The E-OBS gridded data set was derived through interpolation of the ECA&D (European Climate Assessment and Data) station data described by *Klok and Klein Tank* [2008]. The station data set comprises a network of 2316 stations, with the highest station density in Ireland, the Netherlands and Switzerland, and lowest density in Spain, Northern Africa, the Balkans and Northern Scandinavia. The number of stations used for the interpolation differs through time and by variable. The full period of record used for interpolation is 1950–2006, but the period 1961–1990 has the highest density. At any particular time, there are more precipitation than temperature stations. Inhomogeneities in the station time series have been flagged, but potentially inhomogeneous stations have been used for the interpolation, for reasons noted above.

[10] The E-OBS data set was derived through a three stage process [*Haylock et al.*, 2008]. Monthly means (totals) of temperature (precipitation) were first interpolated to a rotated pole 0.1 degree latitude by longitude grid using three-dimensional (latitude, longitude, elevation) thin plate splines. Daily anomalies, defined as the departure from the monthly mean (total) temperature (precipitation), were interpolated to the same 0.1 degree grid, and combined with the monthly mean grid. For temperature, daily anomalies were interpolated using kriging with elevation as an external drift factor. For precipitation indicator kriging has been used to interpolate daily anomalies, where the state (wet/dry) of precipitation was first interpolated, after which the magnitude at "wet" 0.1 degree grid points was interpolated using universal kriging. Finally, the 0.1 degree points have been used to compute area-average values at the four E-OBS grid resolutions (0.25 and 0.5 degree regular latitude-longitude grid and 0.22 and 0.44 degree latitude-longitude rotated pole grids). In this paper, we use the 0.25 degree regular latitude-longitude grid for further evaluation, as results for the other grids are essentially on a similar grid resolution.

[11] Standard error estimates that accompany the gridded data have been derived through combination of the individual standard error estimates for monthly and daily interpolations. Standard errors for the monthly mean or total are the Bayesian standard error estimates, as available in the ANUSPLIN package used for the spline interpolation [*Hutchinson*, 1995; *Wahba*, 1983]. Error estimates for daily

anomalies have been calculated using the method proposed by *Yamamoto* [2000] (see section 5). Both standard error estimates were calculated at the 0.1 degree master grid. For temperature, monthly and daily uncertainties were combined taking the square root of the sum of the squares of the two uncertainties. For precipitation, the relative uncertainty of the daily total (U_{dt}/dt, uncertainty of the daily total, divided by the daily total) is the square root of the sum of the squares of the relative uncertainty of the monthly total (U_{mt}/mt) and the relative uncertainty of the daily proportion of monthly total precipitation (U_{dp}/dp),

$$U_{dt} = dt^* \sqrt{\left(\frac{U_{dp}}{dp}\right)^2 * \left(\frac{U_{mt}}{mt}\right)^2}. \quad (1)$$

Uncertainties at the 0.1 degree grid have been averaged over the target grids allowing for spatial autocorrelation. Details on the interpolation methods, their implementation and the calculation of uncertainties are available from *Haylock et al.* [2008].

3. Homogeneity Assessment

3.1. Homogeneity Testing

[12] To analyze the influence of inhomogeneities in station data on gridded time series and to inform the user about possible inhomogeneous areas within the data set, we apply a homogeneity test to the gridded data set and compare results to the same test for station data. Numerous tests could be used [e.g., *Peterson et al.*, 1998], but for this study we use the Wijngaard method [*Wijngaard et al.*, 2003], which is the same test that was applied to the ECA&D station data used to construct the E-OBS, where 39% of all available precipitation and 25% of all available temperature station series were found likely to be homogeneous over the period 1961–2006 [*Klok and Klein Tank*, 2008]. It has to be noted that the Wijngaard method [*Wijngaard et al.*, 2003] only identifies abrupt shifts and does not detect slow changes, for example, in vegetation or land use, which can also cause data problems. These slow changes have not been discussed in this paper.

[13] The Wijngaard method is an absolute test, as it does not use a supposedly homogeneous reference series. This was appropriate for the version of the ECA&D data set before the ENSEMBLES project started, because of its sparse network [*Wijngaard et al.*, 2003]. It comprises four stepwise homogeneity tests: the standard normal homogeneity test (SNHT) for a single break [*Alexandersson*, 1986], the Buishand range test [*Buishand*, 1981], the Pettitt test [*Pettitt*, 1979] and the Von Neumann test [*Von Neumann*, 1941]. These tests have different characteristics; for example the SNHT, Buishand and Pettitt tests are location-specific, but the Von Neumann test is not. Moreover, the SNHT test is more sensitive to inhomogeneities earlier or later in the time series, whereas the Buishand and Pettitt tests work better for breaks near the middle of the series. If zero or one of the tests detects a break at the 1% significance level the time series is classified "useful"; if a break is detected by two tests the series is classified "doubtful" and if three or four tests find a break, the series is classified "suspect." Homogeneity testing is uncertain; some inhomogeneities found may not be real breaks and other breaks and changes in variance of the data may be overlooked.

[14] For precipitation the annual wet day count is used for the analysis of breaks, as this statistic generally has lower variance than annual total precipitation, enabling a better signal-to-noise ratio for significance testing. For temperature, the annual mean diurnal temperature range (mDTR) and the annual mean of the absolute day-to-day differences of DTR (vDTR) are used for homogeneity detection. DTR is used in preference to mean, maximum or minimum temperature, as tests on DTR are often more sensitive: breaks that are mainly radiation related have different effects on minimum and maximum temperature and are, therefore, only weakly apparent in these variables, but do appear clearly in DTR [*Jones and Lister*, 2009; *Wijngaard et al.*, 2003]. As the homogeneity tests are applied to both mDTR and vDTR, a temperature station is classified according to the worst outcome for the two variables.

[15] We apply the Wijngaard tests to both station and E-OBS gridded data and compare the results. We calculate the annual wet day count, mDTR and vDTR for each year if for each month no more than 20% of the data are missing. If less than 80% of the years in the period 1950–2006 are present, the homogeneity test for that station or grid box is not performed, although these stations may have been used for the interpolation. The consequence of using only stations containing more than 80% of years is that there may be more stations with potential inhomogeneities in the data set. However, a problem with using a smaller percentage is that the Wijngaard method may not be sensitive enough to find breaks at the 0.01 significance level due to the short period used for the test. *Wijngaard et al.* [2003] concluded that, although the exact value of the threshold is arbitrary, a 1 mm threshold should be applied to define a wet day because otherwise too many breaks were detected, and we accordingly adopt this threshold.

3.2. Results and Discussion

[16] Figure 1 shows the stations and grid boxes that are potentially useful (green), doubtful (blue) or suspect (red), according the Wijngaard classification. For precipitation there are many more useful stations and grid boxes than suspect ones. Suspect areas are mainly located in northern Norway, Scotland, Italy, the Balkan, parts of central Europe and in northern Russia. For temperature most of Europe has a statistical significant inhomogeneity at some point in the gridded data, indicated by breaks in mDTR or vDTR (or both). However, if we only look at mDTR there are major differences (see auxiliary material Figure S1), with many more potential inhomogeneities in coastal areas, with remaining areas of central France, UK, Netherlands, parts of Spain and major parts of Ukraine, Northern Russia, Finland, southern Sweden, Czech Republic, Baltic states and former Yugoslavia classified as useful in that case.[1] That we find breaks in mDTR along the coast may be explained by a reduced variability in those areas due to the influence of the sea, making it easier to detect a break in mDTR. Inhomogeneities are much more widespread in

[1]Auxiliary materials are available with the full article. doi:10.1029/2009JD011799.

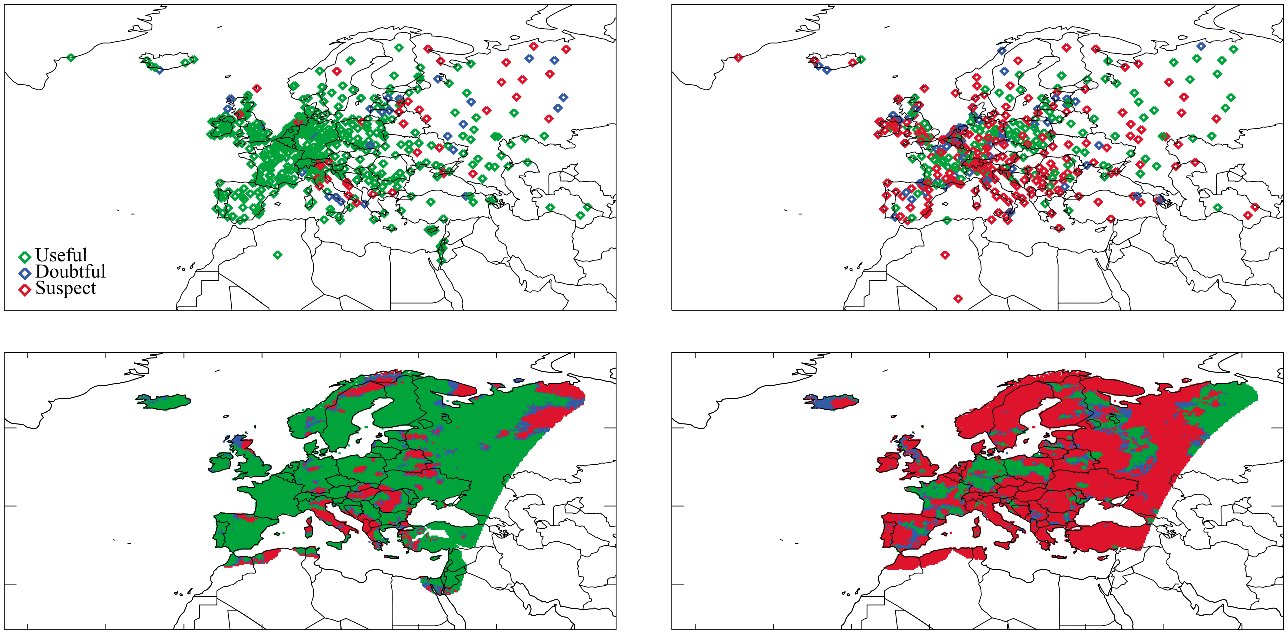

Figure 1. Overall homogeneity, according to the Wijngard test, of (top) the station network and (bottom) the gridded data for (left) precipitation and (right) temperature for the period 1950–2006. For temperature, mDTR and vDTR are combined, with the most negative outcome for the two variables used.

vDTR with no clear difference between coastal and non-coastal areas.

[17] Figure 1 also shows that the areas that have the most suspect stations often also have suspect grids, but sometimes even one suspect station may influence a whole area. An example of the latter is precipitation in northern Sweden where only one station is suspect, but has an influence over many grid boxes. Conversely, some stations have a smaller influence on the area, as, for example, in Russia where many stations are inhomogeneous, but only small areas are influenced. Many stations in this area have breaks in different years and these may be canceled out in the gridded values. In addition, some areas, for example, in Northern Spain, show inhomogeneities in the gridded data in areas where all stations are classified useful. These inhomogeneities may be introduced because of a variable density of the station network. For temperature, inhomogeneous stations are present across the whole of Europe, which is reflected in the inhomogeneities of the gridded data.

[18] In the case of precipitation many more areas of the grids are classified as potentially useful than for temperature (78% for the wet day count versus 46% for mDTR and 28% for vDTR for the grids, and 89% versus 49% and 56% for the stations; see Table 1), which is related to the fact that the homogeneity test is less sensitive for the wet day count. The percentage of stations that are qualified useful in this study is comparable to the percentage of stations that are qualified useful in the study of *Klok and Klein Tank* [2008] (89% for the wet day count in this study versus 94% in the Klok and Klein Tank study and 49% versus 54% for temperature). The differences can be explained by the different time period used for the studies (1950–2006 in our study compared to 1961–2006 from *Klok and Klein Tank* [2008]). The mDTR has a much higher percentage of useful grids than vDTR, whereas vDTR has a higher percentage of useful stations than mDTR. This indicates that for the stations, breaks are more strongly manifested in the mean of the data, whereas in the grids, breaks are more strongly manifested in the standard deviation. That may be due to the fact that the variability of the grid values are dependent on the station density of the network used for the interpolation and the distance to the grid center (N. Hofstra et al., The influence of interpolation and station network density on the distribution and extreme trends of climate variables in gridded data, submitted to *Climate Dynamics*, 2009). A station network that does not have a constant density in time may introduce inhomogeneities.

[19] We also assess the distribution of breaks in time and compare these between gridded and station data (Figure 2). The Von Neumann test is not included in this assessment, as this test is not location-specific. As expected, the SNHT detects more inhomogeneities near the beginning and end of the period than the Buishand and Pettitt tests. SNHT also detects more breaks for any one variable than the other tests (Table 1). For wet day count the inhomogeneity in 1965 detected in the station data by the Pettitt test is also visible in the gridded data. Breaks in the 1975–1985 period in the station data are mainly reflected in the gridded data close to 1980. For mDTR the breaks in station and gridded data do not show a specific pattern. However, where for vDTR the largest inhomogeneities in the station data are found around 1970, the largest breaks in the gridded data are found in the early 1990s. The latter breaks may be due to a declining station density around this time. We investigated whether inhomogeneities could be determined on a decadal basis, by analyzing each of the six decades separately, but the

Table 1. Fraction of Stations or Grids That Are Useful, Doubtful, or Suspect and the Inhomogeneous Fraction for Each Statistical Test

			Overall Fraction			Fraction With Breaks			
	Stations or Grids	Number of Stations or Grids	Useful	Doubtful	Suspect	Standard Normal Homogeneity Test	Buishand	Pettitt	Von Neumann
Wet day fraction	Stations	836	0.892	0.044	0.064	0.123	0.072	0.114	0.087
	Grids	22176	0.781	0.078	0.141	0.219	0.164	0.216	0.166
mDTR	Stations	472	0.492	0.114	0.394	0.477	0.422	0.432	0.468
	Grids	21970	0.464	0.099	0.437	0.515	0.470	0.460	0.485
vDTR	Stations	472	0.555	0.097	0.348	0.434	0.388	0.400	0.381
	Grids	21970	0.275	0.113	0.612	0.738	0.630	0.580	0.697

Wijngaard method is not sensitive enough to find any inhomogeneities in these shorter periods at the 0.01 significance level.

[20] We also divide the calculated potential breaks for all three methods of the 57 year period into six decadal groups and assess the inhomogeneities spatially (see auxiliary material Figures S2–S5). We can conclude, for example for precipitation, that most Italian and former Yugoslavian stations around the Adriatic Sea with a break have this break in the period 1980–1990 for all three tests; these breaks are also propagated through into the gridded data. For precipitation, for all three tests in general, the timing of the breaks in the gridded and station data compares quite well. For temperature, the agreement in timing of breaks between the station and gridded data is smaller. For example, for vDTR a large part of Russia and the Ukraine have the largest significant break between 1990 and 2000 for all three tests, whereas most stations in this area suggest the largest break exists between 1960 and 1980. This indicates that there may be multiple breaks in the station time series of which one becomes more important in the gridded data.

[21] The inhomogeneities within the gridded data are important to keep in mind during any use of the data set. For example, when studying trends in the data, the results within the areas that are suspect may not be meaningful. For those who require more detail, we have prepared a file on homogeneity of the data, which can be downloaded from the E-OBS download site (http://eca.knmi.nl/download/ensembles/ensembles.php).

4. Comparison With Existing Data Sets

4.1. Existing Data Sets

[22] In the second test of the data set we compare E-OBS to existing data sets developed with much denser station networks. Since station density is a very important factor in the interpolation and the interpolation errors are smaller in areas with a dense station network [*Hofstra et al.*, 2008], these existing data sets are deemed close to the 'true' areal average, and provide a useful reference against which to judge the E-OBS data set. The differences in interpolation methods used for the development of the different data sets are assumed to be of minor importance. The three existing data sets used are the UK, Alps and ELDAS data sets. ELDAS and the Alps data sets only comprise precipitation data. The UK data set contains all four variables. We were unable to find or not allowed access to additional data sets in other regions.

4.1.1. UK

[23] The UK data set, supplied by the UK Met Office, comprises a 5 × 5 km equal-area grid, covering the period 1958–2002 for precipitation, 1995–2002 for minimum and maximum temperature and 1995–2006 for mean temperature [*Perry and Hollis*, 2005]. This data set has been compiled from a station network of 4400 stations for precipitation and 540 stations for temperature using multiple regression with geographic factors as the independent variables, followed by inverse distance weighting (IDW) of the residuals. In comparison, the ECA&D station network has 138 stations within this area, of which most had 70–85% of the data available for all variables. To allow comparison with the E-OBS interpolations all grid points

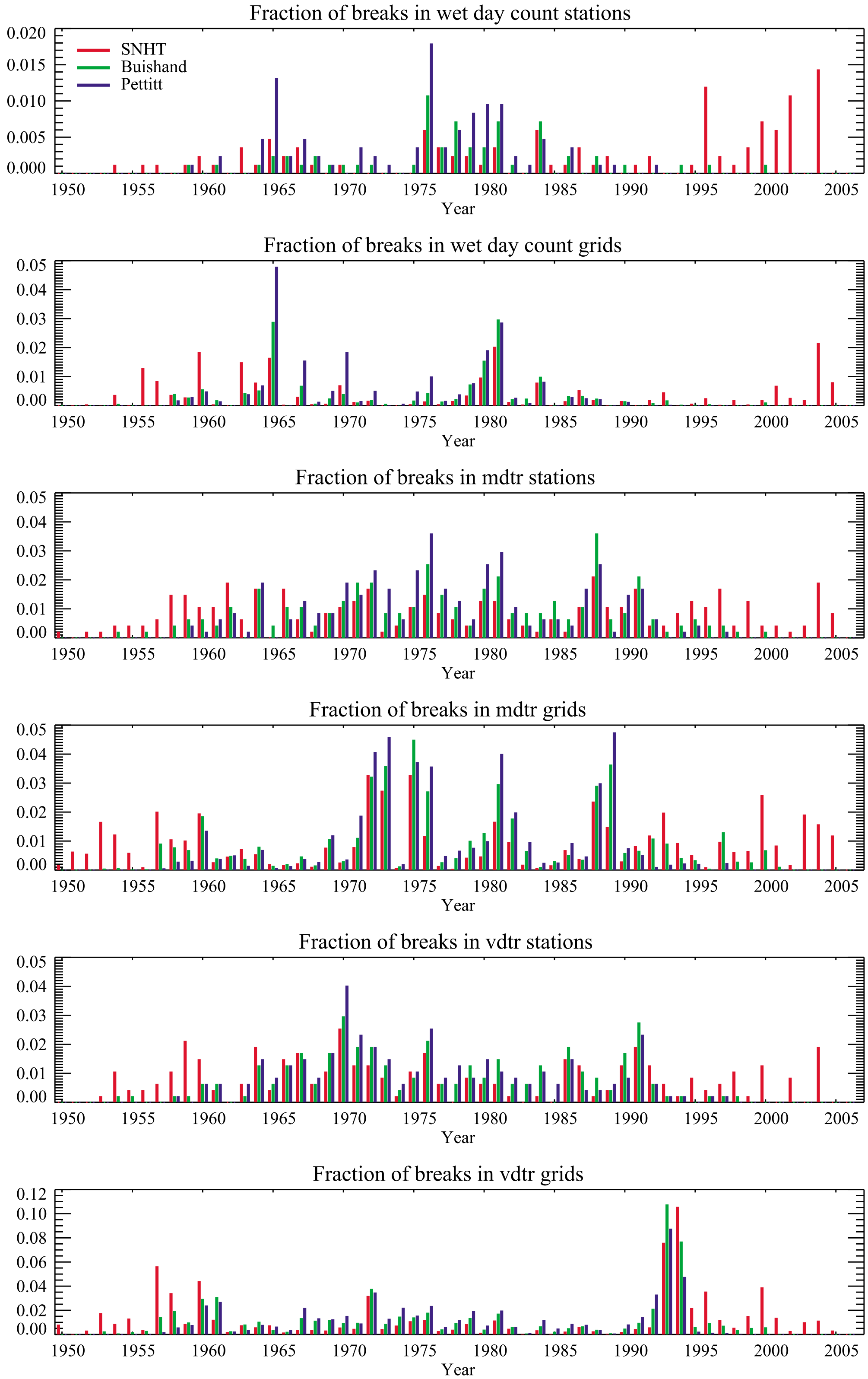

Figure 2. The fraction of stations and grid points with a statistically significant (0.01) inhomogeneity in each year of the data set. Inhomogeneities are calculated for the full 1950–2006 period. The Von Neumann test is not location-specific, so we do not show the results of this test.

within each 0.25 degree grid used for the interpolation were averaged. We also compare this data set to ELDAS (see section 4.1.3), for which a 1 degree grid is used.

4.1.2. Alps

[24] The Alps data set, comprising precipitation only, is an updated version of the climatology and daily data described by *Frei and Schär* [1998] and *Schwarb et al.* [2001], described in more detail by *Hofstra et al.* [2008]. The data are available on a 0.25×0.1667 degree grid and cover the period 1966–1999. For the period 1966–1970 there are no data available over Austria and after 1990 there are data quality issues with many of the Italian stations, so in our comparison, we use the period 1966–1990, except for Austria, where the period 1971–1990 has been used. The data set was constructed through addition of daily anomalies to the long-term climatological mean. Anomalies were interpolated from station data using a modified version of the Shepard algorithm (an ADW technique) [*Frei and Schär*, 1998; *Shepard*, 1984] and the long-term climatology was derived with a local regression approach (PRISM) [*Daly et al.*, 2002] specifically calibrated for the Alps [*Schwarb et al.*, 2001]. The data set is based on over 6500 station records. In comparison, the E-OBS station network has 341 stations available within this area, with majority having over 70% data presence. To allow comparison with E-OBS on a common grid, both data sets were averaged to a 0.25×0.25 degree grid. Since 1.5 Alps grids make up one 0.25 degree grid in latitudinal direction, the average is taken as 2/3 times the value of the grid that falls fully within the 0.25 degree grid and 1/3 times the value of the grid that falls half in the 0.25 degree grid. This might introduce a small uncertainty, which is deemed negligible.

4.1.3. ELDAS

[25] The ELDAS daily precipitation data set has been developed by *Rubel et al.* [2004] for the Development of a European Land Data Assimilation System to predict Floods and Droughts (ELDAS) project. It covers central and northern Europe at 0.2 degree latitude by longitude and covers the relatively short period of October 1999 to December 2000. Some 21,600 stations were used for the interpolation, compared to 2000 for E-OBS over the ELDAS domain. Station density is reasonably homogeneous, but areas such as Portugal, Belgium, Italy, the Balkans, Czech Republic, the Baltic states and Scandinavia have a lower density than Spain, France, the Netherlands, the UK, Denmark, Germany, Poland, Switzerland and Austria. Interpolation was done via the Precipitation Correction and Analysis method [*Rubel and Hantel*, 2001]; this comprises a dynamical bias correction combined with an ordinary block kriging algorithm. To enable comparison, we averaged ELDAS and E-OBS to a common 1 degree latitude by longitude grid.

4.2. Comparison

[26] We compare E-OBS to the high-quality grids using five skill scores for temperature and six for precipitation. Skill scores are often used in forecasting, but we use them here to identify the skill of our data set to reproduce data that have been developed with a much denser station network, which are assumed to provide a good representation of the "true" area averages. We calculate the skill scores for all data together to obtain overall scores, and also on a grid point basis to explore the spatial patterns in difference between grids. We use the mean absolute error (MAE), root mean squared error (RMSE), compound relative error (CRE) and Pearson correlation (R) to assess temperature and the precipitation amount. The Critical Success Index (CSI) and Percent Correct (PC) are used to study precipitation state (wet or dry, where a wet day is defined as having precipitation ≥ 0.5 mm). The skill scores are described in detail elsewhere [*Hofstra et al.*, 2008], but we include an explanation of each score in the auxiliary material. For precipitation we also divide the MAE and RMSE by the mean precipitation for the grids in order to remove the influence of the amount of precipitation on these two skill scores in each grid.

[27] We note that the high-quality data are not true areal averages. However, given they are based on an order of magnitude denser networks than E-OBS, we expect them to be subject to smaller interpolation errors. Thus we can only quantify differences between the data sets, which provide a qualitative indication of potential errors in E-OBS, but should not be interpreted as errors in the data set.

4.3. Results and Discussion

[28] Table 2 provides an overview of the results of the skill scores, calculated "globally" for each grid pairing, as well as for each standard season. At first sight, the data sets compare very well: correlations, CSIs and PCs are high (for example, the correlation coefficient over the full area for temperature (UK only) is approximately 0.99 and for precipitation 0.85–0.92), the CREs are small and RMSEs are fairly small (for example, CRE is 0.02–0.04 and 0.18–0.36 for temperature and precipitation, respectively). However, the mean differences between data sets are quite large. RMSE is 0.70–0.90°C for temperature and 2.17–2.42 mm for precipitation, apart from the Alps where it is larger, at 5.77 mm. MAE shows similar, but smaller differences. For precipitation, the relative RMSE varies between 0.73 (UK) to 1.33 over the Alps. Relative difference between E-OBS precipitation and the other data sets are smaller in winter (UK and ALPS) and autumn (ELDAS). The main reason for larger differences between the data sets in summer is that in summer precipitation is mainly convective rather than frontal. During this season the correlation between stations is lower than in the other seasons. Interpolation with a larger station density will then produce better areal averages than interpolation using a less dense network. For mean and minimum temperature the data sets are closer to each other in spring, whereas they compare better in winter for maximum temperature.

[29] Figure 3 presents the results for precipitation spatially. E-OBS compares best to the UK data set, as does the ELDAS data set, suggesting that over the UK E-OBS is fairly reliable. The differences are generally larger over the west of Scotland, where topography is an important contributing factor to spatial variability in rainfall. E-OBS does not agree as well with the Alps data set, where the topographic complexity means that the sparse E-OBS network does not result in the same gridded data as the denser Alps network; although absolute errors are large because precipitation is on average higher in the Alps, relative errors are also larger than in the UK. Similarly, E-OBS compares poorly to ELDAS over Norway, due to

Table 2. Skill Scores for the Comparison of the E-OBS Gridded Data Set With the UK, Alps, and ELDAS Gridded Data Sets for the Four Variables Minimum, Maximum, and Mean Temperature and Precipitation[a]

		R	MAE (mm or °C)	MAE/Mean	RMSE (mm or °C)	RMSE/Mean	CRE	CSI	PC
				Annual					
UK	Minimum temperature	0.984	0.687	n/a	0.895	n/a	0.041	n/a	n/a
	Maximum temperature	0.991	0.597	n/a	0.780	n/a	0.024	n/a	n/a
	Mean temperature	0.991	0.517	n/a	0.695	n/a	0.023	n/a	n/a
	Precipitation	0.916	1.081	0.355	2.170	0.729	0.182	0.836	0.909
Alps	Precipitation	0.880	2.253	0.514	5.766	1.325	0.357	0.769	0.897
Eldas	Precipitation	0.846	1.159	0.457	2.419	1.009	0.316	0.744	0.874
				Winter					
UK	Minimum temperature	0.971	0.700	n/a	0.918	n/a	0.082	n/a	n/a
	Maximum temperature	0.977	0.507	n/a	0.680	n/a	0.056	n/a	n/a
	Mean temperature	0.974	0.533	n/a	0.718	n/a	0.068	n/a	n/a
	Precipitation	0.925	1.187	0.331	2.227	0.627	0.176	0.856	0.914
Alps	Precipitation	0.894	2.013	0.505	5.031	1.274	0.346	0.784	0.906
Eldas	Precipitation	0.848	1.256	0.458	2.360	0.926	0.373	0.759	0.869
				Spring					
UK	Minimum temperature	0.973	0.663	n/a	0.860	n/a	0.069	n/a	n/a
	Maximum temperature	0.981	0.640	n/a	0.822	n/a	0.051	n/a	n/a
	Mean temperature	0.984	0.491	n/a	0.631	n/a	0.039	n/a	n/a
	Precipitation	0.916	0.893	0.359	1.803	0.730	0.181	0.828	0.908
Alps	Precipitation	0.881	2.237	0.514	5.345	1.231	0.365	0.775	0.888
Eldas	Precipitation	0.853	1.039	0.465	2.103	0.992	0.338	0.742	0.875
				Summer					
UK	Minimum temperature	0.955	0.668	n/a	0.866	n/a	0.116	n/a	n/a
	Maximum temperature	0.970	0.709	n/a	0.896	n/a	0.087	n/a	n/a
	Mean temperature	0.965	0.520	n/a	0.700	n/a	0.082	n/a	n/a
	Precipitation	0.898	1.004	0.402	2.136	0.874	0.207	0.807	0.903
Alps	Precipitation	0.852	2.531	0.546	6.088	1.385	0.392	0.732	0.878
Eldas	Precipitation	0.826	1.026	0.514	2.003	1.334	0.577	0.690	0.870
				Autumn					
UK	Minimum temperature	0.976	0.720	n/a	0.928	n/a	0.067	n/a	n/a
	Maximum temperature	0.987	0.518	n/a	0.667	n/a	0.035	n/a	n/a
	Mean temperature	0.983	0.526	n/a	0.709	n/a	0.042	n/a	n/a
	Precipitation	0.921	1.243	0.341	2.408	0.681	0.173	0.849	0.912
Alps	Precipitation	0.899	2.228	0.495	6.196	1.368	0.326	0.783	0.914
Eldas	Precipitation	0.863	1.226	0.431	2.511	0.911	0.306	0.765	0.879

[a]Skill scores have been calculated for each grid point and are then averaged.

the greater station density for the ELDAS data set in this topographically complex area. Finally, the E-OBS precipitation data set has virtually no stations available in northern Africa, which causes the poor agreement in this area. Figure 4 shows the spatial pattern of skill for temperature over the UK. In general, the agreement is good for all three temperature elements. Differences are greatest over Scotland compared to the rest of the UK. That may be a result of the higher station density of the UK network, which may have had more station data available at higher elevations in Scotland. Differences in agreement between the grids are generally larger than differences between the four seasons.

[30] We also evaluate whether E-OBS shows a bias compared to the high-density data sets, by counting the frequency of days where E-OBS is more than ± 0.1 standard deviations from the high-density data set (Figure 5). For precipitation, E-OBS shows a negative bias at nearly all grid boxes relative to the Alps and ELDAS data sets. Compared to the ELDAS data set, E-OBS is positively biased over parts of Norway and at scattered locations elsewhere in Europe. Over the UK, E-OBS rainfall tends to be negatively biased in areas of higher rainfall in the west, apart from Northern Ireland where there is a positive bias (and also compared to ELDAS). For temperature there are areas with a positive (too warm) and a negative (too cold) bias. One striking feature is that areas such as Devon/Cornwall and southern Wales, that are too warm for minimum temperature, are often too cold for maximum temperature. The bias for temperature is not consistent over the whole of the UK.

[31] In Figure 6 we assess the difference between E-OBS and the high-density data sets across the distribution of precipitation amount and temperature. For this we calculate for each grid deciles of temperature and precipitation (for all wet days) for the full available time period. We then calculate for each day and each grid the absolute difference between the E-OBS and the other data sets and plot the median, 5th, 25th, 75th and 95th percentiles of these differences in each decile (Figure 6). While precipitation is biased toward smaller values in all deciles of the data set, the bias is larger for more extreme precipitation. In the comparison of the 10th decile for the Alps the error between the two data sets can be as high as 16 mm, which is the median of the error when E-OBS is compared to the Alps data set (see median of 9–10th decile of E-OBS versus Alps comparison in Figure 6). The reason for this relates to the

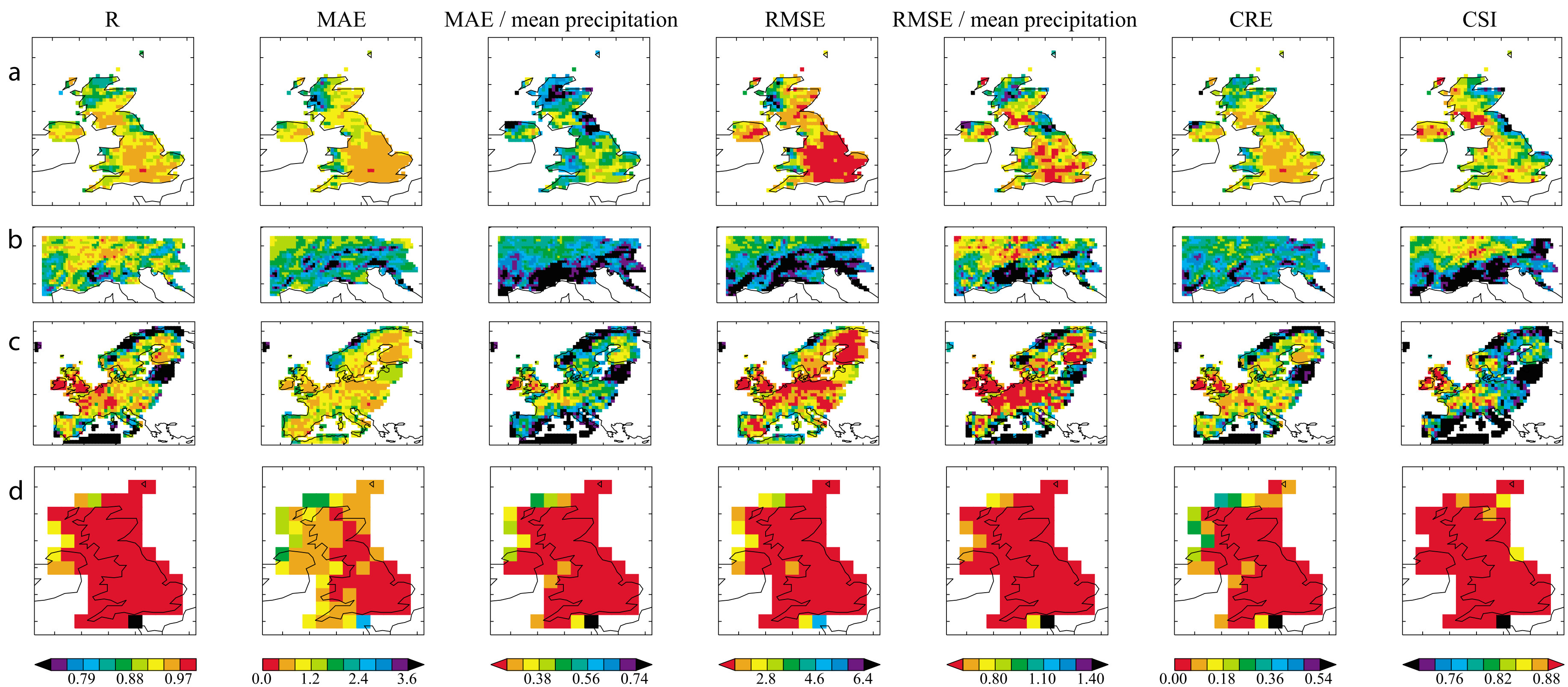

Figure 3. A spatial overview of the skill scores R, MAE (mm), RMSE (mm), CRE, and CSI for precipitation for the comparison of the E-OBS data set with the data sets of the (a) UK, (b) Alps and (c) ELDAS and (d) the UK versus ELDAS. MAE/mean precipitation and RMSE/mean precipitation are added to remove the influence of the average amount of precipitation in a grid cell on the skill score.

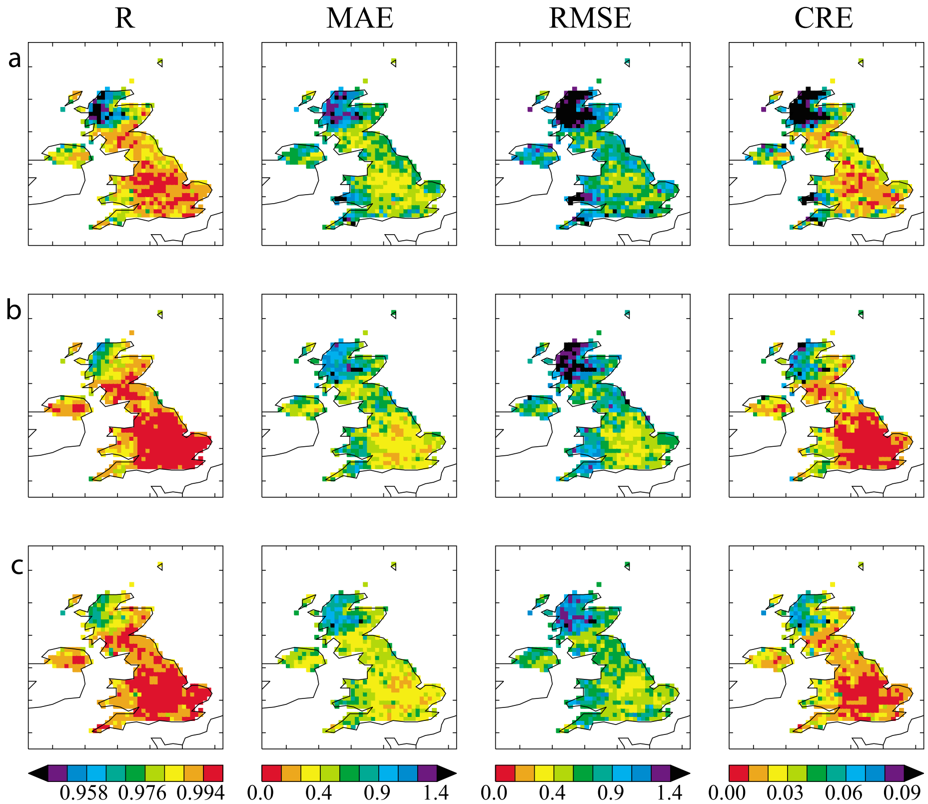

Figure 4. Same as Figure 3 but for the skill scores R, MAE (°C), RMSE (°C), and CRE for (a) minimum, (b) maximum, and (c) mean temperature for the comparison with the UK data set.

much higher station density in the other data sets. For E-OBS, interpolation typically occurs from more distant stations compared to the high-density data sets; as extreme precipitation events are usually more localized, they will be oversmoothed if a sparse network is used. For temperature, differences in error are similar for all deciles, with an average of around 0.5°C. The errors are slightly larger in the first decile for minimum temperature and the tenth decile for maximum temperature, which means that there are slightly larger errors in the extremes, but overall extreme temperature events will be quite well represented (see also the discussion of extremes by *Haylock et al.* [2008]).

[32] We can conclude that E-OBS exhibits quite large differences from existing data sets based on higher-density networks. While correlations overall, and on a grid-by-grid basis, are high, relative differences in precipitation are large, and usually biased toward an underestimation. For temperature (UK only), mean absolute differences are at least 0.5°C. The fact that the ELDAS precipitation data set shows a much better spatial match to the UK data set than E-OBS underlines the fact that E-OBS is fundamentally limited by its underlying station network. As the E-OBS network density over the UK is above average compared to density over the rest of Europe, we can conclude that this issue is likely to be pervasive across much of the E-OBS domain. Assessment of the agreement with existing data sets for all deciles of precipitation and temperature shows that the errors are larger in the extremes than in the more average amounts of precipitation or temperature. There seem to be significant problems with the underestimation of precipitation extremes. Comparability is much higher for temperature than for precipitation, due to the fact that temperature is a continuous variable as opposed to precipitation.

5. Assessment of the Uncertainties

5.1. Calculation of Uncertainties

[33] *Brohan et al.* [2006] give an overview of all sources of known and calculable uncertainty in their HadCRUT3 gridded global monthly temperature data set. Three groups of uncertainties have been identified: (1) station error, (2) sampling error, and (3) bias error. Station error includes errors made during thermometer reading, possible adjustment of homogeneities, calculation of the station normal, and processing of raw data. The sampling error is the difference between the "true" spatial average and the interpolated estimate. It depends on, among others, the number of stations in the grid box, the distribution of those stations and on the variability of the climate in the grid box. The gridding method used by *Brohan et al.* [2006] is a simple area average of the stations within a grid, which is different from the kriging method that we use, but the sampling error of our gridding method will depend on the same factors. Two sources of bias error are summarized by *Folland et al.* [2001]: urbanization effects [*Jones et al.*, 1990] and thermometer exposure changes [*Parker*, 1994]. For precipitation a similar list of sources of uncertainty can be made. Here we focus on sampling error as it is expected to be the largest contributor to overall error. The objective here is to evaluate the accuracy of the estimates of interpolation sampling error for daily anomalies used in E-OBS. As

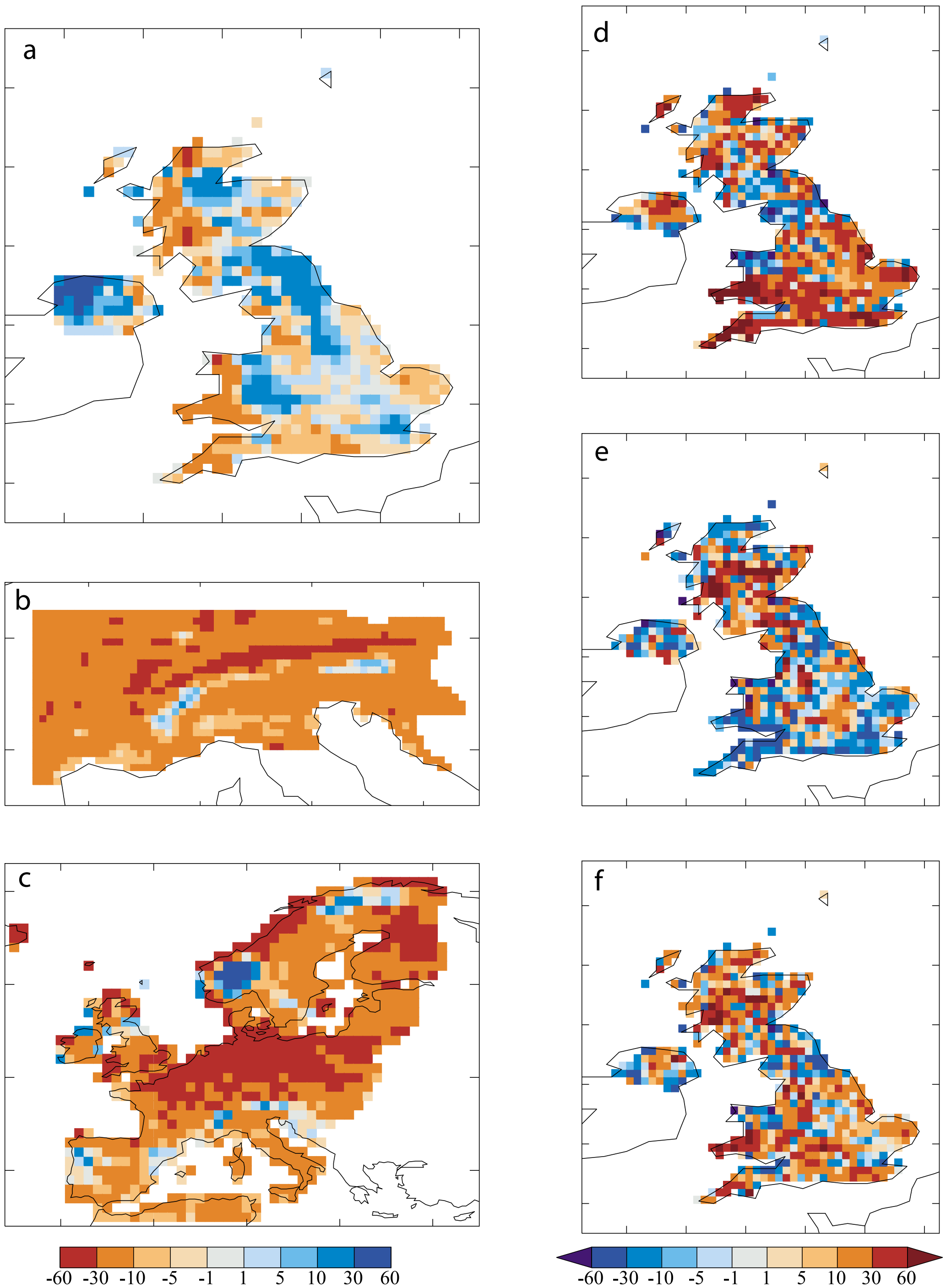

Figure 5. Spatial pattern of bias in the E-OBS data set compared to higher-quality data over the Alps, ELDAS domain, and UK, expressed as the percentage of days that E-OBS data are more than 0.1 standard deviations below the higher-quality data, subtracted from the percentage of days the E-OBS data are more than 0.1 standard deviation above the higher-quality data. Thus a positive value indicates that E-OBS data tend to be biased greater than the higher-quality data, and vice versa. Precipitation is shown for (a) UK, (b) Alps, and (c) ELDAS. For precipitation, red means that the area is estimated too dry in E-OBS compared to the other data sets, and blue means too wet. Temperature (UK only) is shown for (d) minimum temperature, (e) maximum temperature, and (f) mean temperature. For temperature blue means that the area is estimated too cold in E-OBS compared to the other data sets, red means too warm.

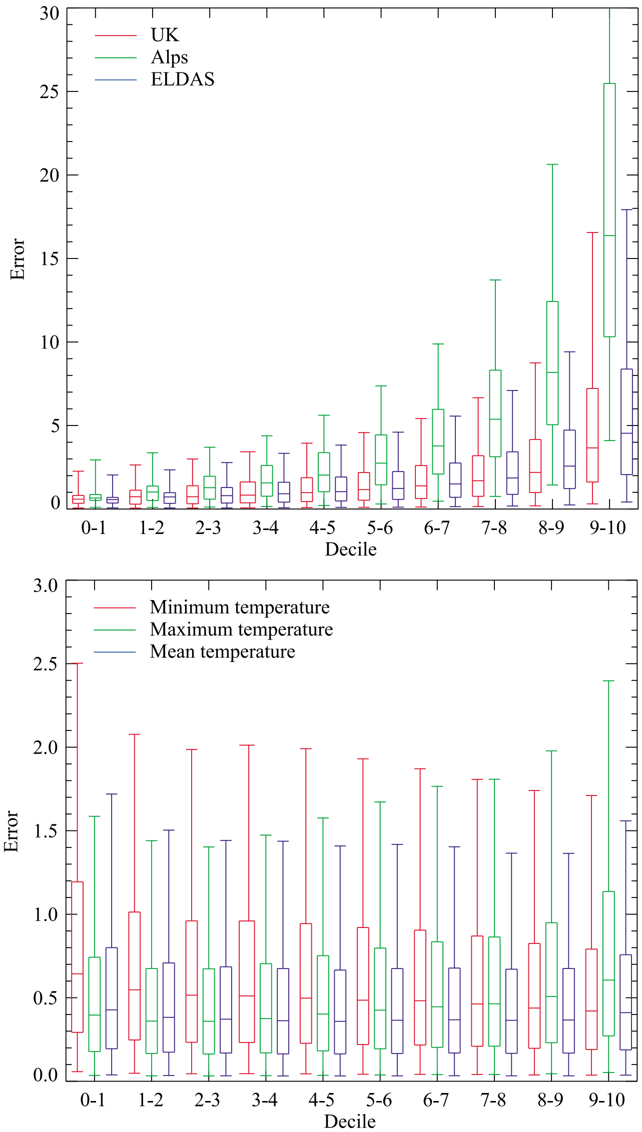

Figure 6. Absolute error in different deciles for each comparison with existing data sets for (top) precipitation (in millimeters) and (bottom) temperature (in °C). In the top plot, red is for the UK, green is for the Alps, and blue is for ELDAS; in the bottom plot, red is for minimum temperature, green is for maximum temperature, and blue is for mean temperature. The box of absolute error shows the 0.25th median and 0.75th percentile, and the whiskers show the 0.05th and 0.95th percentile. Deciles are calculated for each grid separately. The whisker of the 9th to 10th decile for the comparison with the Alps is cut off and runs through to 41 mm.

explained in the introduction, these daily errors are estimated using the method proposed by *Yamamoto* [2000].

[34] *Yamamoto* [2000] estimates the so-called "interpolation standard deviation" at each grid point as the weighted average of the squared differences between station and interpolated values as follows:

$$s_0 = \sqrt{\sum_{i-1}^{n} \lambda_i [z(x_i) - z^*(x_0)]^2}, \qquad (2)$$

where x_i (i = 1,n) are the locations of the stations used for the interpolation and λ_i are the weights used in the kriging interpolation and z are the observed values at the i stations used for the interpolation (x_i) and z* is the interpolated value at the location for the interpolation (x_0). The kriging weights are a function of distance to x_0 and their calculation is described by *Haylock et al.* [2008].

[35] *Yamamoto* [2000] compared his interpolation standard deviation to the kriging standard deviation and cross-validation error. The kriging standard deviation is a standard by-product of kriging and used widely as a measure of reliability of the kriging procedure. The interpolation standard deviation has much larger correlation with cross-validation error than with the kriging standard deviation. The reason for that is that the kriging standard deviation is not a true estimate of uncertainty [*Journel and Rossi*, 1989; *Monteiro da Rocha and Yamamoto*, 2000], as it cannot properly measure local data dispersion [*Yamamoto*, 2000].

[36] As we do not have the true grid values for evaluation, we adopt station cross validation to test the accuracy of the *Yamamoto* [2000] interpolation standard deviation. We estimate the daily anomaly at each station in the ECA&D data set used to construct E-OBS, using the same interpolation approach used for E-OBS gridded data. Interpolation standard deviation is calculated using equation (1) above and cross-validation error as the absolute difference between the interpolated station value and the observed value,

$$cve_0 = |z^*(x_0) - z(x_0)|. \qquad (3)$$

[37] We next transform the interpolation standard deviations into 95% confidence intervals by multiplication with 1.96 (assuming a normal distribution, which is justified, but likely not fully correct, because we have normalized the data by using anomalies for the interpolation rather than actual daily data) and addition to and subtraction from the interpolated daily values for each station. We then count the number of times the observed station value falls within the 95% confidence interval for the interpolated value, with the expectation that if the confidence interval is an accurate estimate of interpolation uncertainty we would expect the station value to fall outside the confidence interval approximately 5% of the time.

5.2. Results and Discussion

[38] We first compare the cross-validation error (CVE) and interpolation standard deviation (ISD) through scatterplots. Results are similar for all temperature variables, so we only show figures for precipitation and minimum temperature.

[39] Correlation between the CVE and ISD for both temperature and precipitation is positive (Figure 7). The relationship between CVE and ISD is stronger for precipitation (r = 0.57) than minimum temperature (r = 0.33), which provides confidence that the spatial distribution of ISD will reflect the spatial variability in interpolation error. The relationship is also closer to one-to-one for precipitation, whereas for temperature, ISD tends to be too large at smaller CVE and vice versa.

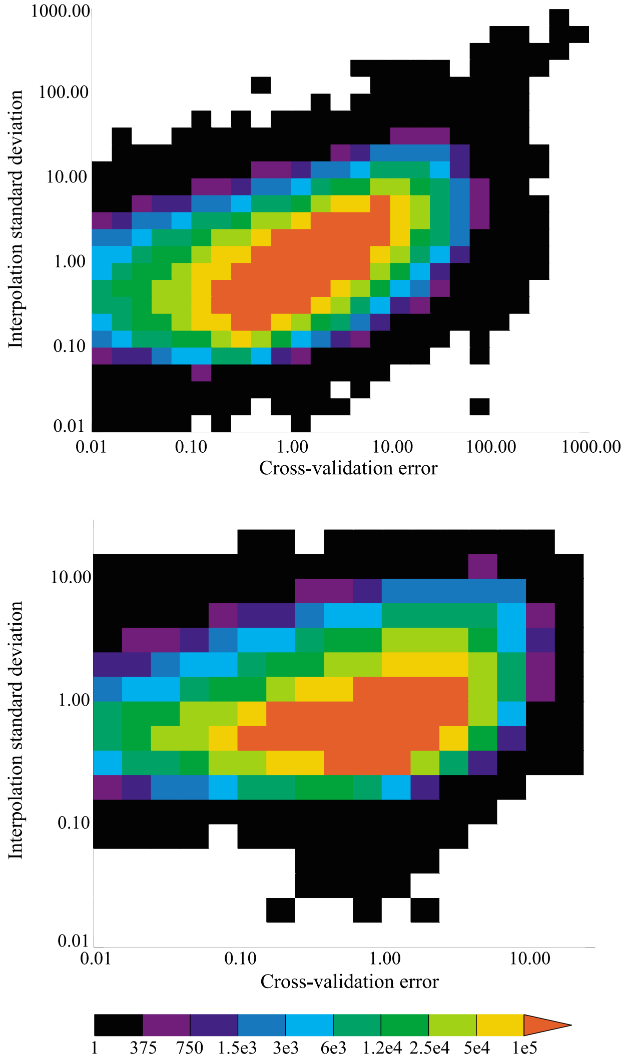

Figure 7. Bivariate histograms showing the joint frequency distribution of cross validation error and interpolation standard deviation for (top) precipitation and (bottom) minimum temperature. Both plots are on a log-log scale.

[40] However, a better test of the accuracy of the ISD is the count of the percentage of station values falling outside the interpolation 95% confidence interval derived from the ISD (Figure 8). For precipitation, the upper 95% limit is mostly exceeded between 5 and 10% of the time, while values fall below the lower limit 10–25% of the time, indicating that while the upper limit is a reasonable estimate, the lower limit is poorly defined, and that precipitation is frequently significantly underestimated. For temperature, there are roughly equal numbers of values falling above and below the 95% confidence interval, but as with precipitation, the number exceeds that expected. Most stations have at least 10% of data falling outside the confidence interval, with many stations having more than 25% of values outside the interval. There is also a clear north-south gradient in the percentage of the precipitation values falling outside the confidence limits, with the CI underestimation being much larger in the north. The main reason for this is the fact that there are fewer rain days in the south of Europe, compared to the north. The error is smaller when no or little precipitation is observed, compared to a situation when a lot of precipitation is observed.

[41] From this analysis, we can conclude that the interpolation standard deviation provided with the data is a strong underestimation of the actual interpolation error and should be used with care. Moreover, it has to be taken into account, that the confidence intervals available with the gridded data only include interpolation sampling error and no station and bias errors.

6. Summary and Conclusions

[42] We have analyzed the new E-OBS European high-resolution gridded data set of daily minimum, maximum and mean temperature and precipitation in three ways. First, we assessed the homogeneity of the gridded data and related this to the homogeneity of the station data. Second, we compared the data set to existing gridded data sets developed with denser station networks. And finally, we evaluated the accuracy of the interpolation standard deviation, a measure of interpolation error that is provided with the data set. While the three issues we assess do not give a complete overview of the reliability of the data set, they do provide important additional information for users of the data set.

[43] The results of the *Wijngaard et al.* [2003] homogeneity tests show that there are many potential inhomogeneities present in the gridded data set. There are more statistically significant breaks present in temperature than precipitation data, and within the temperature data, there are more breaks for vDTR (the annual mean of the absolute day-to-day differences of the diurnal temperature range (DTR)) than mDTR (annual mean DTR) variables. Inhomogeneities in the gridded data are often related to inhomogeneities in the stations contributing to the value of the grid. However, this relation is not the same for all areas. Sometimes an area is inhomogeneous even if there are zero or only one inhomogeneous station in the area (e.g., for precipitation in northern Spain and northern Sweden, respectively) and on other occasions many stations are inhomogeneous, but the area is not effected (e.g., for temperature in southeastern France). The former statement indicates that a station network that varies in time may introduce inhomogeneities in the data. In addition, not all stations could be tested for homogeneity, as many stations did not have data for 80% or more of the years 1950–2006. These stations might have inhomogeneities that we do not find in our study. The year of the break of inhomogeneous grids generally corresponds to the year of the break of stations in the surrounding area, although the correspondence is better for precipitation than for temperature. This information will be critical when, for example, performing analyses of trends in extremes using E-OBS. We would recommend users of the data to only use potentially homogeneous areas for trend analysis. For a future update of the E-OBS data set we recommend that the issue of inhomogeneities is studied thoroughly. A balance will have to be found between the loss of station data and the introduction of inhomogeneities. An alternative, as presented by *Haylock et al.* [2008], would be to incorporate the homogeneity

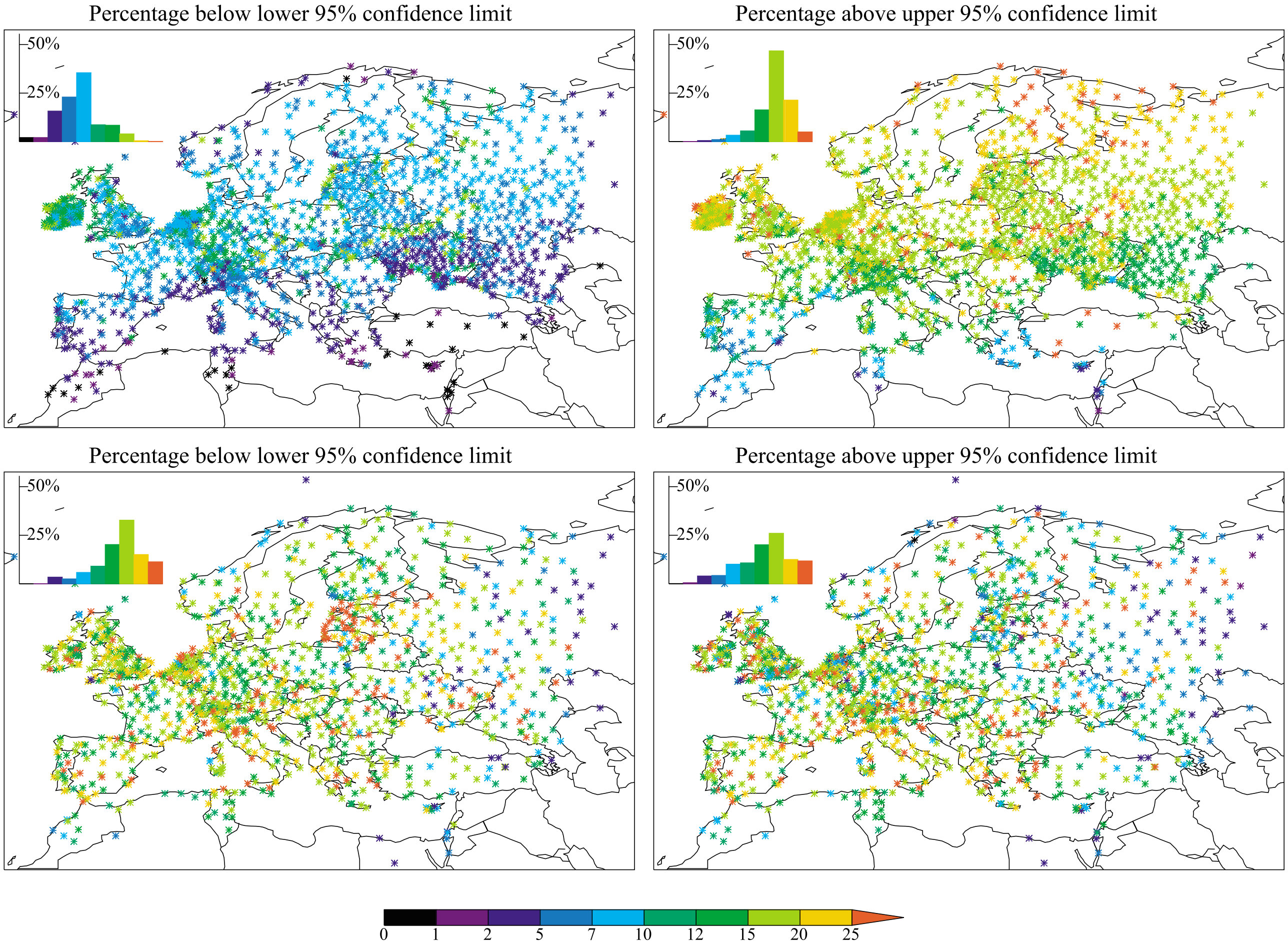

Figure 8. Spatial patterns of the percentage of interpolated data exceeding the (left) lower and (right) upper limits of the 95% confidence interval for (top) precipitation and (bottom) minimum temperature for all stations. Insets display histograms of the frequency of the overestimation or underestimation of the stations.

assessment results in the uncertainty estimates, such as by using stochastic simulations.

[44] As far as the authors are aware this is the first paper that studies the homogeneity of gridded climate data. In this paper we only apply an existing test for stations to the gridded data. We do not test how effective the use of this method is. An additional issue compared to station data is that gridded data, at least in the case of E-OBS, are formed by a station network that varies in density and placement over time, which may introduce inhomogeneities in the actual value and the variance of temperature and precipitation. This requires further analysis.

[45] When compared to existing high-resolution regional gridded data for the UK, Alps and Europe (ELDAS) that are based on much denser station networks, E-OBS shows an excellent correlation. However, mean absolute errors are significant, in the order 0.5°C for temperature and greater than 100% for precipitation. For both variables and all skill scores the data sets compare worse in mountainous areas. For precipitation, agreement is in general better in winter, whereas for temperature agreement is mainly best in spring. In the case of precipitation, E-OBS also shows a negative bias, indicating that E-OBS tends to be oversmoothed relative to the high-density data sets. For temperature, E-OBS shows a small positive bias over quite large areas, but some scattered areas have a stronger negative bias. Moreover, the E-OBS data set compares better to the mean of the variables of the existing data sets than to the extremes, although differences are much larger for precipitation than for temperature. Consequently, the data set should be used with caution in comparison to RCM outputs, especially with respect to evaluation of RCM precipitation extremes.

[46] The uncertainty estimates available with the data only represent sampling, or interpolation, errors. These are calculated by combining errors from both parts of the interpolation process, namely interpolation of the monthly mean (temperature) or totals (precipitation) using thin plate smoothing splines and the interpolation of daily anomalies using versions of kriging (see section 2). We evaluate the daily interpolation error estimates, estimated using *Yamamoto*'s [2000] interpolation standard deviation approach. A comparison of these errors with cross-validation errors shows that for most of Europe cross-validation error is positively correlated with interpolation standard deviation. However, the frequency with which *Yamamoto*'s [2000] 95% interpolation confidence interval is exceeded is much larger than expected, indicating that the

interpolation standard deviation significantly underestimates the actual interpolation error. The 95% confidence limits are on average exceeded 25% and sometimes even over 50% of the time. In a future update of the data we recommend that ensemble stochastic simulations should be considered for the estimation of uncertainties. Such an approach could be used to create a set of interpolated realizations that honor the observations, but vary away from the observing stations by an amount dependent on the distance to the observations as well as the variability of the observations [*Ahrens and Beck*, 2008; *Deutsch and Journel*, 1998]. These have also been mentioned by *Haylock et al.* [2008] but have not been implemented due to time constraints. *Bellerby and Sun* [2005] and *Teo and Grimes* [2007] suggest shortcuts that should reduce the computing time required. That uncertainty estimates are underestimated should not have the consequence that users of the data do not use the estimates. However, the uncertainties should be seen as minimum uncertainties.

[47] The E-OBS data set is the first publicly available data set that covers the whole of Europe at a very high spatial resolution for daily data. However, as this study reveals, there are some potentially important limitations to the data. Inhomogeneities are present within the data, the data show quite large absolute and relative differences and biases to existing data sets that have been developed with very dense station networks, and the standard errors delivered with the data appear to significantly underestimate the true interpolation error. This will have to be taken into account when the data are used, for example, for the evaluation of RCM outputs. Trends analysis may also be affected by potential inhomogeneities in the data. In addition, the underestimation of extremes within the data may, for instance, influence future predictions using RCM outputs regarding flooding. Moreover, when using the standard errors that have been supplied with the data it has to be taken into account that these errors only include interpolation sampling errors and that these sampling errors are underestimated.

[48] The E-OBS data will often be the only available data set for studies of, for example, the comparison of RCM outputs for the whole of Europe. With the collation of more data, reconsideration of how to deal with inhomogeneities in station data, and the improvement of the uncertainty estimates the data will improve in the future. However, users of the data should take notice of the weaknesses mentioned in this paper and use the data appropriately.

[49] **Acknowledgments.** We would like to thank all institutes (see Appendix 1 of *Klok and Klein Tank* [2008]) that made meteorological station data available for the study. This study was funded by the EU project ENSEMBLES (WP 5.1 contract GOCE-CT-2004-50539). N.H. is also funded by the Dutch Prins Bernhard Cultuurfondsbeurs and the Dutch talentenbeurs.

References

Ahrens, B., and A. Beck (2008), On upscaling of rain-gauge data for evaluating numerical weather forecasts, *Meteorol. Atmos. Phys.*, *99*, 155–167, doi:10.1007/s00703-007-0261-8.

Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, *6*, 661–675, doi:10.1002/joc.3370060607.

Bellerby, T. J., and J. Sun (2005), Probabilistic and ensemble representations of the uncertainty in an IR/Microwave satellite precipitation product, *J. Appl. Meteorol.*, *6*, 1032–1044.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Buishand, T. A. (1981), The analysis of homogeneity of long-term rainfall records in the Netherlands, *KNMI Sci. Rep. WR 81–7*, 42 pp., R. Neth. Meteorol. Inst., De Bilt, Netherlands.

Dai, A., et al. (1997), Surface observed global land precipitation variations during 1900–88, *J. Clim.*, *10*, 2943–2962, doi:10.1175/1520-0442(1997)010<2943:SOGLPV>2.0.CO;2.

Daly, C., et al. (2002), A knowledge-based approach to the statistical mapping of climate, *Clim. Res.*, *22*, 99–113, doi:10.3354/cr022099.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 369 pp., Oxford Univ. Press, Oxford, U. K.

Folland, C. K., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, *28*(13), 2621–2624, doi:10.1029/2001GL012877.

Frei, C., and C. Schär (1998), A precipitation climatology of the alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, *18*, 873–900, doi:10.1002/ (SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.

Haylock, M., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded dataset of surface temperature, precipitation and sea-level pressure, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), The comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, *113*, D21110, doi:10.1029/2008JD010100.

Hutchinson, M. F. (1995), Interpolating mean rainfall using thin plate smoothing splines, *Int. J. Geogr. Inf. Syst.*, *9*(4), 385–403, doi:10.1080/02693799508902045.

Jones, P. D., and D. H. Lister (2009), The influence of the circulation on surface temperature and precipitation patterns over Europe, *Clim. Past Discuss.*, *5*, 535–555.

Jones, P. D., et al. (1990), Assessment of urbanization effects in time series of surface air temperature of land, *Nature*, *347*, 169–172, doi:10.1038/347169a0.

Journel, A. G., and M. E. Rossi (1989), When do we need a trend model in kriging?, *Math. Geol.*, *21*, 715–739, doi:10.1007/BF00893318.

Klok, L., and A. M. G. Klein Tank (2008), Updated and extended European dataset of daily climate observations, *Int. J. Climatol.*, *29*, 1182–1191, doi:10.1002/joc.1779.

Monteiro da Rocha, M., and J. K. Yamamoto (2000), Comparison between kriging variance and interpolation variance as uncertainty measurements in the Capanema iron mine, state of Minas Gerais—Brazil, *Nat. Resour. Res.*, *9*, 223–235, doi:10.1023/A:1010195701968.

New, M. (1999), Uncertainty in representing observed climate, in *Representing Uncertainty in Climate Change Scenarios and Impact Studies*, edited by T. Carter et al., pp. 59–66, Clim. Res. Unit, Univ. Of East Anglia, Norwich, U. K.

Parker, D. E. (1994), Effects of changing exposure of thermometers at land stations, *Int. J. Climatol.*, *14*, 1–31, doi:10.1002/joc.3370140102.

Perry, M., and D. Hollis (2005), The generation of monthly gridded datasets for a range of climate variables over the UK, *Int. J. Climatol.*, *25*, 1041–1054, doi:10.1002/joc.1161.

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

Pettitt, A. N. (1979), A non-parametric approach to the change-point detection, *Appl. Stat.*, *28*, 126–135, doi:10.2307/2346729.

Rubel, F., and M. Hantel (2001), BALTEX 1/6-degree daily precipitation climatology 1996–1998, *Meteorol. Atmos. Phys.*, *77*, 155–166, doi:10.1007/s007030170024.

Rubel, F., et al. (2004), Daily and 3-hourly quantitative precipitation estimates for ELDAS, report, 32 pp., Biometeorol. Group, Univ. Vet. Med., Vienna.

Schwarb, M. C., et al. (2001), Mean annual and seasonal precipitation in the European Alps 1971–1990, in *Hydrological Atlas of Switzerland*, plates 2.6 and 2.7, Landeshydrol. und Geol., Bern.

Shepard, D. S. (1984), *Computer mapping: The SYMAP interpolation algorithm*, in Spatial Statistics and Models, edited by G. L. Gaile and J. C. Willmott, pp. 133–145, Springer, New York.

Teo, C.-K., and D. I. F. Grimes (2007), Stochastic modelling of rainfall from satellite data, *J. Hydrol.*, *346*, 33–50, doi:10.1016/j.jhydrol.2007.08.014.

Von Neumann, J. (1941), Distribution of the ratio of the mean square successive difference to the variance, *Ann. Math. Stat.*, *12*, 367–395, doi:10.1214/aoms/1177731677.

Wahba, G. (1983), Bayesian "confidence intervals" for the cross-validated smoothing spline, *J. R. Stat. Soc., Ser. B*, *45*(1), 133–150.

Wijngaard, J. B., et al. (2003), Homogeneity of 20th century European daily temperature and precipitation series, *Int. J. Climatol.*, *23*, 679–692, doi:10.1002/joc.906.

Yamamoto, J. K. (2000), An alternative measure of the reliability of ordinary kriging estimates, *Math. Geol.*, *32*(4), 489–509, doi:10.1023/A:1007577916868.

M. Haylock and P. D. Jones, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

N. Hofstra, Environmental Systems Analysis Group, Wageningen University, P.O. Box 47, NL-6700 AA Wageningen, Netherlands. (nynke.hofstra@wur.nl)

M. New, School of Geography and the Environment, Oxford University Centre for the Environment, University of Oxford, South Parks Road, Oxford OX2 8JU, UK.