



A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006

M. R. Haylock,^{1,2} N. Hofstra,³ A. M. G. Klein Tank,⁴ E. J. Klok,⁵
P. D. Jones,¹ and M. New³

Received 1 April 2008; revised 30 June 2008; accepted 28 July 2008; published 30 October 2008.

[1] We present a European land-only daily high-resolution gridded data set for precipitation and minimum, maximum, and mean surface temperature for the period 1950–2006. This data set improves on previous products in its spatial resolution and extent, time period, number of contributing stations, and attention to finding the most appropriate method for spatial interpolation of daily climate observations. The gridded data are delivered on four spatial resolutions to match the grids used in previous products as well as many of the rotated pole Regional Climate Models (RCMs) currently in use. Each data set has been designed to provide the best estimate of grid box averages rather than point values to enable direct comparison with RCMs. We employ a three-step process of interpolation, by first interpolating the monthly precipitation totals and monthly mean temperature using three-dimensional thin-plate splines, then interpolating the daily anomalies using indicator and universal kriging for precipitation and kriging with an external drift for temperature, then combining the monthly and daily estimates. Interpolation uncertainty is quantified by the provision of daily standard errors for every grid square. The daily uncertainty averaged across the entire region is shown to be largely dependent on the season and number of contributing observations. We examine the effect that interpolation has on the magnitude of the extremes in the observations by calculating areal reduction factors for daily maximum temperature and precipitation events with return periods up to 10 years.

Citation: Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.

1. Introduction

[2] Data sets of spatially irregular meteorological observations interpolated to a regular grid are important for climate analyses. Such gridded data sets have been used extensively in the past and will continue to be important for many reasons. First, such interpolated data sets allow best estimates of climate variables at locations away from observing stations, thereby allowing studies of local climate in data-sparse regions.

[3] Second, for monitoring of climate change at the regional and larger scale we frequently utilize indices of area averages. Such indices range in scale, from those representing local regions such as the Central England temperature record [Parker and Horton, 2005] up to indices

of global change such as the mean global temperature [Brohan *et al.*, 2006]. Such area averages require data at an equal area grid (or an averaging scheme incorporating this implicitly) so as not to bias the average to regions with a higher spatial density of observing stations.

[4] Third, climate variability studies often seek regional patterns of coherent variability and therefore employ multivariate eigenvalue techniques, such as principal component analysis, canonical correlation analysis and singular value decomposition. Such techniques prefer regularly spaced observations so as not to bias the eigenvalues to regions with a higher density of observations.

[5] Fourth, validation of Regional Climate Models (RCMs) is becoming more important as such models gain increased popularity for regional climate change studies. Such a direct comparison between models and interpolated data assumes that the observations and model are indicative of processes at the same spatial scale. Models are generally agreed to represent area averaged rather than point processes [Osborn and Hulme, 1998], especially their representation of the hydrological cycle. Therefore construction of a gridded data set where each grid value is a best estimate average of the grid square observations is the most appropriate data set for validation of the model, rather

¹Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.

²Now at PartnerRe, Zurich, Switzerland.

³Climate Research Lab, Oxford University Centre for the Environment, Oxford, UK.

⁴Royal Netherlands Meteorological Institute, De Bilt, Netherlands.

⁵TNO Built Environment and Geosciences, Utrecht, Netherlands.

than comparison between the model and point observations directly.

[6] Finally, impacts models are important for determining the possible consequences of climate change, such as changes to water supply or crop yields. Such models often require regularly spaced data, and are much easier to implement with temporally complete time series which gridding achieves, with no need to consider how to deal with missing data.

[7] In this paper, we present a European high-resolution gridded daily data set of precipitation and surface temperature (mean, minimum and maximum). The data set was developed as part of the European Union Framework 6 ENSEMBLES project, with the aim being to use it for validation of RCMs and for climate change studies.

[8] There are several similar daily gridded data sets already available for Europe, however none can compare to the set presented here in terms of the length of record, the spatial resolution, the incorporation of daily uncertainty estimates or the attention devoted to finding the best interpolation method. The currently existing gridded data sets have either a coarser grid resolution, a shorter time span, or do not cover all of Europe. Additionally, none of the data sets includes error estimates. HadGHCND [Caesar *et al.*, 2006] is a global gridded daily data set based upon near-surface maximum and minimum temperature observations. It spans the years 1946–2000 but is available on a much coarser 2.5° latitude by 3.75° longitude grid. The European Commission Joint Research Centre in Ispra, Italy houses the MARS-STAT database containing European meteorological observations interpolated onto a 50 km grid, but only from 1975 up to present. The Alpine precipitation gridded analysis [Frei and Shär, 1998] is based on 6700 daily precipitation series and covers the period 1966–1995. The spatial resolution of this data set is 25 km and the region encompasses just the Alpine countries. Daily observations at point locations, for the present study, have been available through the European Climate Assessment and Data set (ECA&D; <http://eca.knmi.nl/>). Most of the station data used in the current study have also helped to enhance the ECA&D station data set to its current status of over 2,000 stations [Klok and Klein Tank, 2008]. All aspects of the data used are discussed in section 2.

[9] A major part of the gridding exercise has been to select the most appropriate methodology for interpolating the point observations to a regular grid. In this process we investigated in detail current best-of-class methods and carried out a detailed comparison of the skill of such interpolation methods. The details of this comparison are presented by Hofstra *et al.* [2008]. While we will not revisit the comparison here, we will focus more on the details of the kriging method that was shown by Hofstra *et al.* [2008] to be the best method. A discussion of the gridding method is the focus of section 3.

[10] An important additional product of the data set is the estimation of interpolation uncertainty of the daily grid square estimates. The methodology for calculating the uncertainty is explained in section 3. Quantifying uncertainty has been an important focus of the exercise to enable users of the data set to gain a better understanding of the temporal and spatial evolution of data quality. We hope that products derived from the data set, such as analyses of past

climate change and comparisons with model data, will be able to incorporate and effectively use these uncertainty measures.

[11] To measure the impact that interpolation has had on the daily grid square extremes, we have undertaken a simple analysis comparing extremes in the raw station data and in data interpolated to the station locations. This is presented in section 4.

[12] We conclude the study with a discussion of the shortcomings of the data set and a summary of our methodology and findings in section 5.

2. Data

2.1. Data Collection

[13] Daily observations were compiled for precipitation, and minimum, maximum and mean surface temperature covering the time period 1950–2006. The collection of data was primarily carried out by the Royal Netherlands Meteorological Institute (KNMI), which also hosts the European Climate Assessment and Data set (ECA&D). The ECA&D set of observing stations served as the starting point for the ENSEMBLES data set, and the ECA&D database infrastructure was also used for ENSEMBLES. At the start of this project (February 2005), only data from the ECA&D data set were at our disposal. At that time, this data set included about 250 stations having data over 50 years, which is insufficient for the purpose of high-resolution gridding. Ideally, the preferred station density for high-resolution (25 km) gridding would be at least one station per 25×25 km. Since Europe's surface area is approximately 10,000,000 km², we would need around 16,000 stations. It was clear early in the project that we would not be able to achieve this, as that many stations do not exist. Therefore, it was a high priority that the data set include interpolation uncertainty estimates based on the spatial correlation structure of the data.

[14] Additional station series were gained from other research projects, such as STARDEX [Haylock *et al.*, 2006], or by petitioning various National Meteorological Services directly. Contacting the services directly was carried out in collaboration with the ECA&D. Other existing data sets provided further stations, such as the Global Climate Observing System (GCOS) Surface Network [Peterson *et al.*, 1997], the Global Historical Climatology Network (GHCND) [Gleason, 2002] and the Mesoscale Alpine Programme (MAP) [Bougeault *et al.*, 2001]. These efforts resulted in an increase in the number of stations from the original 250 to 2316 stations (the exact number varies over time), still well below the ideal number but an order of magnitude improvement on the initial set. Further details on the data collection and a list of contributing institutions are provided by Klok and Klein Tank [2008].

[15] The map of the station network (Figure 1) reveals uneven station coverage, with the highest station density in the UK, Netherlands and Switzerland. The station networks for minimum and maximum temperature are very similar to mean temperature (Figure 1b). Figure 1 shows the complete gridding region. While we have provided grids for northern Africa to match the coverage of the ENSEMBLES RCMs, the very poor station coverage (reflected in very high uncertainty) reduces the utility of the data in these regions.

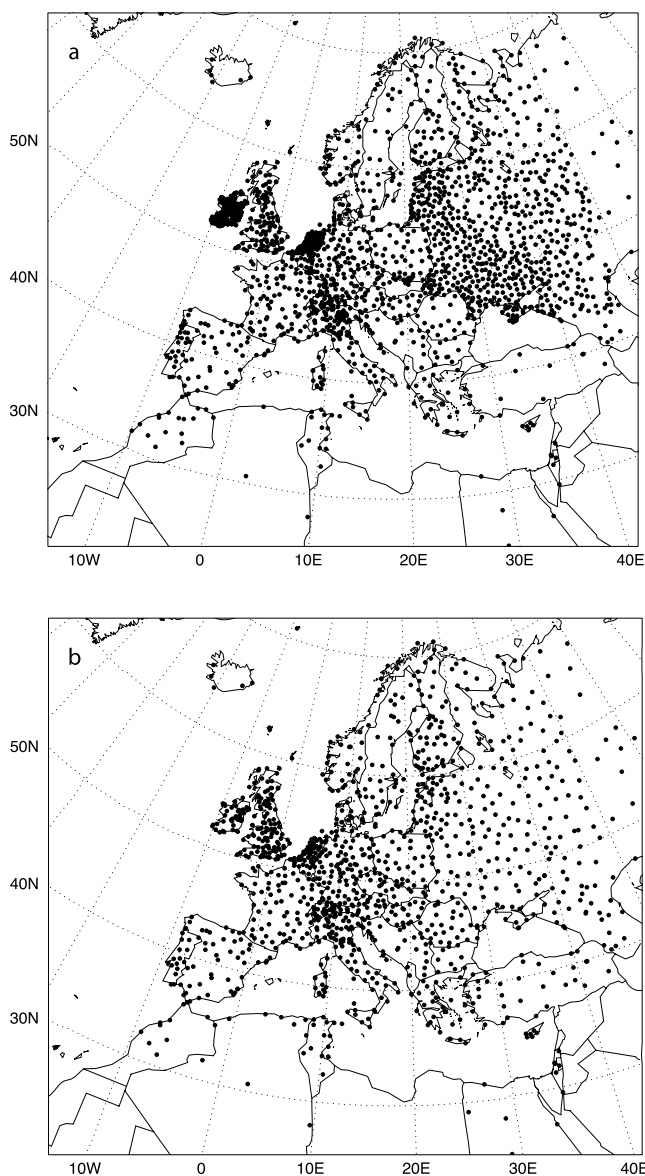


Figure 1. The complete gridding region (land-only), showing the station network for (a) precipitation and (b) mean temperature.

2.2. Data Quality

[16] Raw station observations underwent a series of quality tests to identify obvious problems and remove suspicious values. This included: precipitation less than zero or greater than 300 mm; temperatures higher than 60°C; minimum temperature greater than maximum; and more than 10 days with the same (nonzero) precipitation. Flagged observations of excessive precipitation were checked manually in regions where such amounts might occur (e.g., excessive precipitation in Alpine regions).

[17] Temperature outliers are removed by identifying those days that were more than five standard deviations from the mean with reference to all days within five days of that calendar day over all years. For example, to test the observation on 12 January 1970, we calculate the mean and standard deviation using observations from 7 January to 17

January for all years. In each test we excluded our candidate observation so as not to influence the calculation of the mean and standard deviation in case this is a very large outlier. Since each removal of a value will influence the detection of other outliers, we ran this test repeatedly through the data until no more outliers were detected.

[18] In a preliminary analysis of the spatial correlation structure of the raw station data, we discovered that by shifting some stations forward or backward in time by one day, we obtained much higher correlations between stations. This revealed the problem that the method of assigning dates to each observation by the different National Meteorological Services possibly varied between countries. For example, it is unclear whether the 24 h precipitation total recorded as 1960-01-02 corresponds to the precipitation that fell on that date, or this was the date the observation was made (often at 0900). This holds similarly for the maximum temperature reading. There is also the possibility that this methodology changed through time. This problem has not been detailed before in the literature, but is nevertheless critical to producing a daily gridded data set, and so needed to be addressed.

[19] With limited time and resources available, we addressed the issue by finding the shift for each station (−1, 0, or 1 day) that produced the highest correlation with the nearest grid square in the ERA40 data set. While the ERA40 data could have possibly assimilated the incorrect data, it does at least maintain dynamical consistency between the variables. To calculate the ERA40 daily observation from the four six-hourly time steps the method depended on the variable. For precipitation, we used the 0600–0600 total, which assumes that our precipitation totals were measured at 0900. For example, for 1960-01-01 we added the 1200 and 1800 values of 1960-01-01 to the 0000 and 0600 values of 1960-01-02. For maximum temperature, we took the maximum of the four time steps 0000, 0600, 1200, and 1800 for each day and for minimum temperature the minimum of these four time steps. Each variable was treated independently, so a different shift could be applied to different variables at the same station. Most stations were either unshifted or moved one day earlier (the value that is recorded as 1960-01-02 was moved to 1960-01-01).

3. Gridding Methodology

3.1. A Three-Step Process

[20] Kriging involves solving a set of linear equations to minimize the variance of the observations around the interpolating surface. This least squares problem therefore assumes that the station data being interpolated are homogeneous in space. This is not the case when we have stations across Europe from many climate zones. In regions with higher precipitation, our interpolation should allow for higher interpolation error. The daily data therefore need to be made homogeneous across the region.

[21] We addressed this problem by adopting a three-step methodology of interpolating the daily data: interpolating the monthly mean using thin-plate splines to define the underlying spatial “trend” of the data; kriging the anomalies with regard to the monthly mean; and applying the interpolated anomaly to the interpolated monthly mean to

create the final result. This is similar to universal kriging [Journal and Huijbrechts, 1978], whereby a polynomial is fit to the underlying spatial trend. In such a large and complex region, thin-plate splines are more appropriate method for trend estimation than polynomials. For temperature we calculated anomalies as the difference between the daily observation and the monthly mean. For precipitation we calculated the daily anomaly as the quotient of the daily precipitation total and the monthly total. The precipitation anomaly is therefore the proportion of the monthly total that fell on that day. This has the effect of transforming all the precipitation anomalies to the range [0,1].

[22] As well as homogenising the data across stations, there were other reasons to perform two separate interpolations. First, it enables us to use two separate methods for interpolating the monthly and daily data. We performed cross validation exercises to compare the skill of various methods in interpolating monthly means/totals and daily anomalies to determine the best method. The results of the daily comparison are given by Hofstra *et al.* [2008], which revealed kriging to be the best method. A similar cross validation was carried out for monthly data which showed thin-plate-splines to be the best method. For monthly precipitation and temperature we use three-dimensional splines, taking into account the station elevation.

3.2. Monthly Temperature Means and Precipitation Totals

[23] Thin-plate smoothing splines can be thought of as an extension of linear regression, where the parametric linear function in linear regression is replaced by a smooth nonparametric function. The function is determined from the data, in particular the degree of smoothness is determined by minimizing the predictive error, given implicitly by a cross validation error. The method was originally described by Craven and Wahba [1978]. We utilized the ANUSPLIN package based on the work of Hutchinson [1995]. Note that thin-plate splines are a stochastic method (see section 3.3) and are a very different approach to interpolation than the commonly used cubic (or higher order) splines which fit polynomials to each section of the plane to pass through all the data points and maintain continuity of the derivative (slope).

[24] Monthly precipitation totals and monthly means of temperature were calculated for stations with less than 20% missing data in that month. For months with fewer than 20% missing days, precipitation totals were adjusted by dividing the totals by the proportion of nonmissing observations. This was to account for the possibility of rain on the missing days. A threshold of 20% was considered as high enough so as not to reject too many stations and low enough so as not to create large uncertainty in the monthly total or average. Stations that have a large amount of missing data might also be expected to have other problems, such as incorrect flagging of precipitation accumulation over several days. The number of stations selected by this methodology (Figure 2) shows a sharp rise up until 1960, then increases slightly until 1990, after which there is a large drop. The precipitation stations show a sharp dip in 1976. The number of stations with at least one observation (or less than 99% missing data) per month (Figure 2) also shows a decline in 1976, suggesting that the dip is caused by a

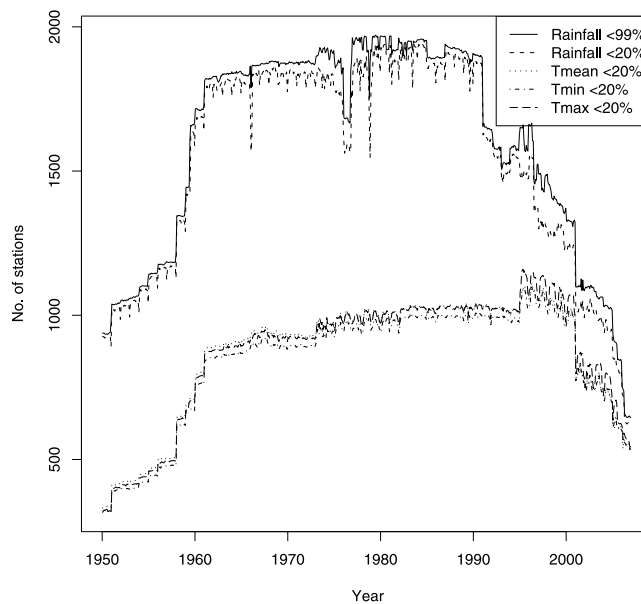


Figure 2. The number of stations with less than 99% and 20% missing observations for each month.

reduction in the number of stations rather than an increase in the amount of missing data. The number of precipitation stations with less than 20% missing data shows a decline every winter, which is not reflected in the number of stations with at least one observation, suggesting that this annual decline is due to an increase in the amount of missing data. A closer examination of the stations that exhibit an increase in the proportion of missing observations every winter (not shown) does not reveal any regional or elevation dependence. This suggests the increase in missing precipitation observations in winter cannot be explained by, for example, snowfall upsetting gauge measurements. The number of temperature stations does not show such a decline every winter.

[25] The monthly means or totals were then interpolated with thin-plate splines using elevation to a high resolution 0.1° by 0.1° rotated pole grid, with the “North Pole” at 162°W , 39.25°N . We chose to use a rotated pole grid so as to allow quasi-equal area grid spacing over the study region. This enabled the largest spatial coverage with the minimum number of interpolated grid squares to increase computational efficiency. Using an unrotated grid would have resulted in a higher grid density in the north of the region compared to the south. The rotated pole was chosen to match the grid used by many of the regional climate models in ENSEMBLES, including models run by the Community Climate Change Consortium for Ireland, the Danish Meteorological Institute, GKSS Forschungszentrum Geesthacht GmbH, the Royal Netherlands Meteorological Institute, the Norwegian Meteorological Institute, the UK Met Office and the Swedish Meteorological and Hydrological Institute.

[26] Although the final grids produced in the data set were spaced at 25 and 50 km, we chose to produce the high resolution 0.1° rotated pole master grid, which we would use to area average to create the final grids. The reason for this is that the interpolation methods were tuned to reproduce as accurately as possible a point observation, whereas

the aim of the gridding exercise was to produce grid square averages. Although the interpolated data calculated directly at the center of a grid square using thin-plate splines or kriging would be very similar to the result obtained after interpolating to many points in the grid square then averaging these, this may not be the case for precipitation occurrence. We wanted to have grid cells with a precipitation occurrence distribution more like an area average (as in a regional climate model) than an observing station which has generally fewer rain days and higher totals on these days than a climate model. This can best be achieved by interpolating to a finer grid with each point having a precipitation occurrence distribution similar to that of an observing station, then averaging these to create a coarser grid.

3.3. Daily Anomalies

[27] Kriging is an interpolation method that has been developed extensively in the geosciences for the application of mapping ore reserves using sparsely sampled drill cores. The methodology has a long history of development, beginning with the pioneering mathematical formalization of Kolmogorov in the 1930s [Kolmogorov, 1941]. The popularity of kriging as a tool for spatial interpolation increased substantially through the efforts of Matheron [1963] and the later work of Journel and Huijbregts [1978]. It is now used extensively in many fields of geoscience because of its skill as an interpolator and its powerful application to other problems such as estimating uncertainty. The term kriging was adopted by Matheron in recognition of the work of Krige [1951].

[28] Kriging, in various forms, has been applied to precipitation interpolation. Barancourt and Creutin [1992] used indicator kriging (see below) for interpolation of daily rainfall in a high station density region in southern France. Ali *et al.* [2005] adopted universal kriging for Sahelian rainfall at different timescales. Their study included a cross validation of several interpolation methods to prove the superiority of regression kriging (defining the spatial trend using multivariate regression) compared with universal and ordinary kriging. In the far larger and more heterogeneous region of our study we prefer to define the spatial trend with thin-plate splines. Dinku *et al.* [2007] created a station-based gridded data set of daily rainfall over East Africa for the validation of satellite products. Although they used kriging for the mean values, anomalies were interpolated with angular distance weighting.

[29] Kriging is a stochastic interpolation method. It assumes that the interpolated surface is just one of many possible solutions, all of which could have produced the observed data points. Stochastic methods use probability theory to model the observations as random functions. The aim of the interpolation is to produce a surface that fits to the expected mean of the random function at unsampled locations. Kriging forms part of a class of interpolators known as best linear unbiased estimators (BLUE): the “estimated” (interpolated) value is a linear combination of the predictors (nearby stations) such that the sum of the predictor weights is 1 (unbiased) and the mean squared error of the residuals from the interpolating surface is minimized (best estimate). The interpolating surface is therefore a local

function of the neighboring data, but conditional on the data obeying a particular model of the spatial variability.

[30] The key to kriging is deciding which statistical model best describes the spatial variation of the data. This is determined by fitting a theoretical function to the experimental “variogram”: the absolute difference between stations as a function of their distance. For this comparison we have selected the best of five models: Gaussian, exponential, spherical, hole effect, and power. These are the most common functions used for variogram modeling and their mathematical description can be found in most geostatistical texts [e.g., Kitandis, 1997; Webster and Oliver, 2001]. The most appropriate function was determined by fitting each of these nonlinear functions to the experimental variogram using the method of Marquardt [1963], and selecting the model with the lowest Chi-square statistic. The model fitting gave more weight to those points in the experimental variogram that were calculated from more station pairs. All variogram models contained three parameters: the range, a measure of how rapidly the spatial correlation reduces with distance; the sill, which describes the expected temperature or precipitation anomaly difference between two stations at far separation; and the nugget variance to allow for spatial variation at a scale not resolved by the station network.

[31] We tried calculating the variogram independently for each day of the analysis period, as well as for each calendar month or just a single variogram for every day. A cross validation exercise showed that the best interpolation skill came from using just a single variogram for all days, probably due to greater statistical certainty in model fitting when using the larger amount of data. Lebel *et al.* [1987] adopted a similar approach of using a single “climatological variogram” for interpolating Sahelian rainfall, although they used a scale factor that depended on the time of year. In our case, we are interpolating the rainfall deviation from the monthly mean and so a monthly-dependent variogram did not add skill.

[32] We tested different search radii to select neighboring stations for variogram modeling and interpolation. The highest cross-validation skill came from using a radius of 450 km for precipitation and 500 km for temperature. Selecting smaller search radii might improve local detail, but at the expense of providing fewer stations for variogram modeling and interpolation.

[33] Variogram modeling allows for the possibility that spatial correlation may depend on orientation, for example in mid latitudes one might expect that observing stations might be more highly correlated in an east-west direction than north-south because of the prevailing weather patterns. Creating direction-dependent variograms is known as anisotropic variogram modeling [Kitandis, 1997]. We examined the potential added value of this approach by calculating variograms separately for directions 45° either side of the north-south axis and comparing this with directions 45° either side of the east-west axis. We found no significant difference that would warrant anisotropic modeling at the expense of increased variogram uncertainty due to fewer data.

[34] There have been many extensions to kriging, such as indicator kriging to model binary distributions (such as precipitation occurrence) and kriging with an external drift,

which uses information from a covariate (such as elevation) to assist interpolation. For temperature we incorporate elevation dependencies by using external drift kriging [Goovaerts, 2000]. Block kriging is a further extension that enables the estimation of areal means rather than point values [Grimes *et al.*, 1999; Journal and Huijbregts, 1978]. We chose not to adopt this, instead interpolating to a fine resolution master grid and averaging this to different resolutions. Still, block kriging would be worth considering as a better estimation of areal means in future updates.

[35] Kriging of precipitation anomalies is more complicated than for temperature due to the binary nature of precipitation occurrence. Indicator kriging is an established method to model such spatially discrete variables and developed from the application of interpolating concentrations in ore reserves, which often occur in discrete bodies. Indicator kriging has been successfully applied to precipitation [Atkinson and Lloyd, 1998; Barancourt and Creutin, 1992; Teo and Grimes, 2007]. We adopted a similar method to Barancourt and Creutin [1992], whereby the rainfall is transformed to a binary distribution dependent on being above or below a threshold. We selected a threshold of 0.5 mm to define a rainy day. Adopting rainy day thresholds lower than this has been shown to be sensitive to data quality, such as under reporting of small rainfall amounts due to bad observer practice [Hennessy *et al.*, 1999]. Variogram modeling and ordinary kriging was then performed on the binary variable to produce interpolated values of the probability of observing a rainfall event above 0.5 mm. Those grid boxes with a probability higher than a certain probability threshold are then assigned a rainfall anomaly using ordinary kriging. A probability threshold of 0.4 was used, selected by testing various thresholds with cross validation. Barancourt and Creutin [1992] show that an unbiased estimate of the probability threshold is just the proportion of observations that are above 0.5 mm.

[36] All variogram modeling and kriging was implemented in custom FORTRAN code based partly on GSLIB code [Deutsch and Journel, 1998]. The GSLIB kriging code was altered to use the more stable singular value decomposition method to solve the kriging linear equations rather than Gaussian elimination [Press *et al.*, 1986].

[37] The modeled “theoretical” variograms (Figure 3) are all spherical, apart from precipitation amount which is exponential. The ranges of the variograms range from 470 km, for rainy day probability, up to 1262 km for precipitation amount.

3.4. Uncertainty Estimates

[38] Obtaining estimates of uncertainty was a high priority for this data set. Uncertainty arises from many sources, including all stages of the observation and analysis, from measurement and recording errors to data quality, homogeneity and interpolation. Incorporating all these sources would be ideal, but without significantly more resources this was not possible. We therefore focused on a best assessment of the interpolation uncertainty. Our cross validation exercise to select the best interpolation methodology [Hofstra *et al.*, 2008] showed typical interpolation error to be much larger than the expected magnitude of other sources of uncertainty. Folland *et al.* [2001] adopt temperature uncertainty in the order of tenths of Kelvin for

inhomogeneities, thermometer exposure and urbanization compared with an interpolation error of several Kelvin [Hofstra *et al.*, 2008].

[39] The final uncertainty estimate is dependent on the uncertainty in the monthly means/totals and the uncertainty in the daily anomalies. Both thin-plate splines, used for monthly interpolation, and kriging, used for daily interpolation, are stochastic methods that allow an estimate of interpolation uncertainty. ANUSPLIN uses the methods described by Hutchinson [1993] and Hutchinson [1995], a detailed description of which is beyond the scope of this paper. For the monthly uncertainty we would have ideally liked to calculate uncertainty for each month separately, however computational constraints prohibited this. We therefore calculated the monthly uncertainty by using the uncertainty determined by interpolating the monthly climatology (calculated from all available years of data) and applied this to all years. Therefore the monthly uncertainty for January 1960 is the same as for January 1961 and all other Januarys in the period.

[40] Daily uncertainty was determined from the kriging method. Kriging provides a measure of the expected mean at an interpolated point as well as the variance. Several climate studies have used the kriging variance as a proxy for uncertainty. Lebel and Amani [1999] analyzed single event accumulated rainfall in a station-dense region of the Sahel to reveal the behavior of the estimation variance with varying station density. Similarly, Grimes *et al.* [1999] merged gauge and satellite data based on their relative uncertainties determined using the kriging variance.

[41] However previous work on kriging showed that the kriging variance is not a true estimate of uncertainty [Journel and Rossi, 1989; Monteiro da Rocha and Yamamoto, 2000], but rather just a product of the variogram. Kriging variance is independent of local variation and dependent only on station separation. For example, the kriging variance of precipitation is the same at an interpolated point regardless of whether the neighboring stations have all recorded no precipitation or have all recorded widely varying extreme amounts.

[42] The best solution to quantify kriging uncertainty is to perform an ensemble of stochastic simulations. This approach produces a set of interpolated realizations, all of which 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 [Deutsch and Journel, 1998]. This methodology has been adopted in the past with precipitation observations [Ekstrom *et al.*, 2007; Kyriakidis *et al.*, 2004; Teo and Grimes, 2007]. Unfortunately the computational requirements increase linearly with the ensemble size. Therefore to have a reasonable ensemble to determine uncertainty (at least 30 members) would have required many months of computational power compared to the two weeks required for the single realization of 57 years of daily data at high spatial resolution for four variables. Shortcuts have been proposed to speed up simulations by reducing the degree of randomness of the simulations. One method is to execute the simulations in the same grid box order for each simulation, thereby removing the need to calculate kriging weights for each simulation separately [Bellerby and Sun, 2005; Teo and Grimes, 2007]. We did not consider attempt-

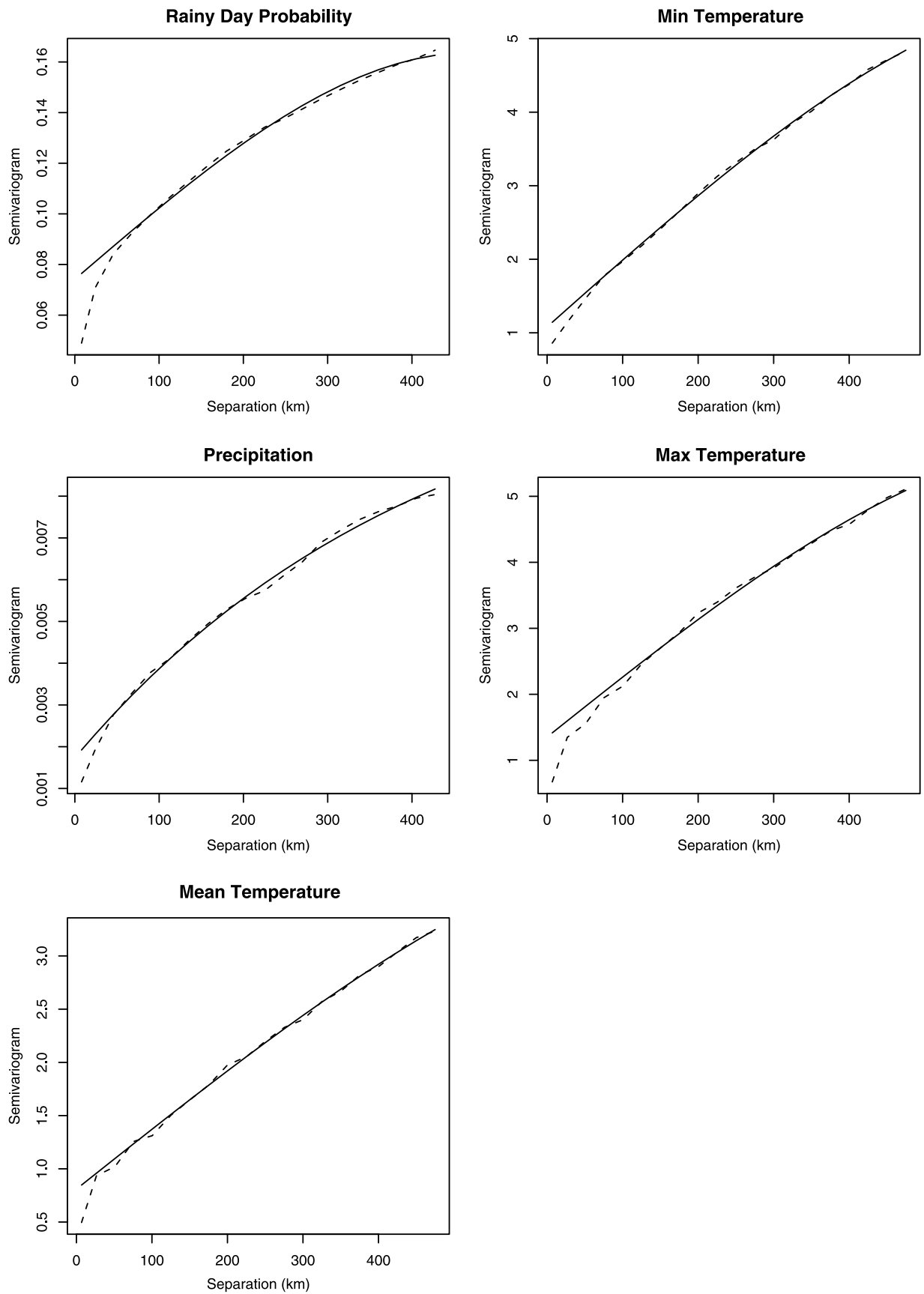


Figure 3. Experimental (dashed lines) and theoretical (solid lines) semivariograms for rainy day probability, precipitation amount, and mean, minimum, and maximum temperature.

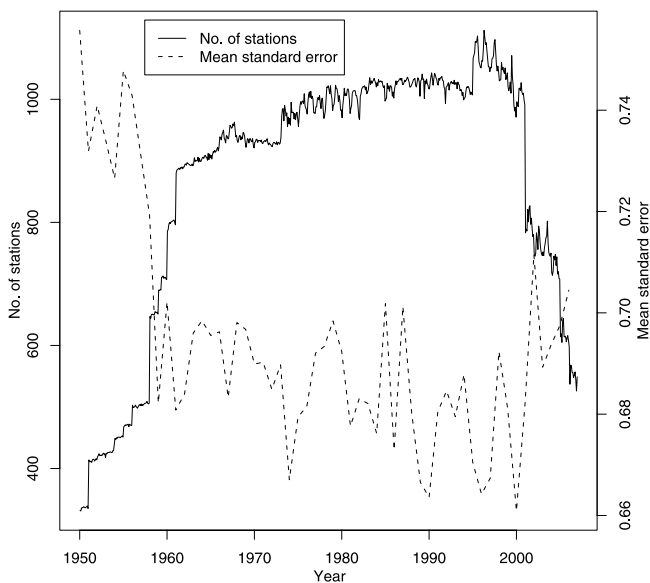


Figure 4. Annual average standard error for 0.1° master grid averaged across the entire region for mean temperature, along with the monthly number of stations with less than 20% missing observations.

ing such optimizations, but they would be well worth investigating for future updates to the data set.

[43] We adopted a solution provided by *Yamamoto* [2000], who gives an alternative method of assessing kriging uncertainty using just the data provided by the single realization. Kriging interpolates to a point by calculating a weighted sum of neighboring observations, with the weights determined by the variogram model and the separation distances. The method of addressing uncertainty is based on the premise that we would expect higher uncertainty at an interpolated point when the neighbors are more variable. When neighbors are similar, one would expect less uncertainty. Through cross validation, *Yamamoto* [2000] shows that their “interpolation variance” shows much closer correspondence to the true error than the kriging variance. We applied this method to every grid point for every day to arrive at the standard error for the daily anomaly. For precipitation, the standard error (in units of proportion of monthly total) is converted to mm by multiplying by the interpolated monthly total. For temperature the kriging standard error is in Kelvin.

[44] As discussed in section 3.3, for precipitation we use separate interpolations for occurrence and magnitude. The uncertainty estimate is calculated from the magnitude interpolation, regardless of whether the occurrence model designated a wet or dry day. Unfortunately we cannot easily incorporate uncertainty from the occurrence model [*Barancourt and Creutin*, 1992], which would best be addressed through simulations.

[45] To calculate the final uncertainty at a grid square we combined the uncertainties from the monthly climatology and the daily anomaly in quadrature, i.e., the square root of the sum of the squares of the two uncertainties. The final

uncertainty for the 0.1° master grid is given for every grid square on every day as a standard error.

[46] The average uncertainty over the entire domain is largely dependent on the number of stations (Figure 4). There is a tendency for greater uncertainty at the start and end of the period due to the reduced number of stations. For temperature, the uncertainty also shows a strong annual cycle, with higher uncertainty in spring and lower uncertainty in autumn (Figure 5). There is also a marked reduction in temperature uncertainty in December compared with the adjacent months. An examination of the monthly maps of uncertainty (not shown) shows this to be due to generally lower uncertainty in December in the central and southern latitudes. Uncertainty is generally higher in snow covered northern latitudes in winter due to decreased spatial consistency.

[47] Standard errors for the 25 and 50 km gridded data sets are calculated from the uncertainty in the high-resolution master by allowing for spatial autocorrelation. Error propagation theory [*Goodman*, 1960] gives that for a linear function of correlated variables (equation (1)), the standard error is dependent on the variance-covariance matrix of the dependent variables (equation (2)). The variances in \mathbf{M} are the square of the standard error of the master grid and the covariances are calculated directly from the variogram.

$$f = \mathbf{a}^T \mathbf{x} \quad (1)$$

$$\sigma_f^2 = \mathbf{a}^T \mathbf{M} \mathbf{a}, \quad (2)$$

where $\mathbf{M} = \begin{bmatrix} \sigma_1^2 & COV_{12} & COV_{13} & \dots \\ COV_{12} & \sigma_2^2 & COV_{23} & \dots \\ COV_{13} & COV_{23} & \sigma_3^2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$ is the

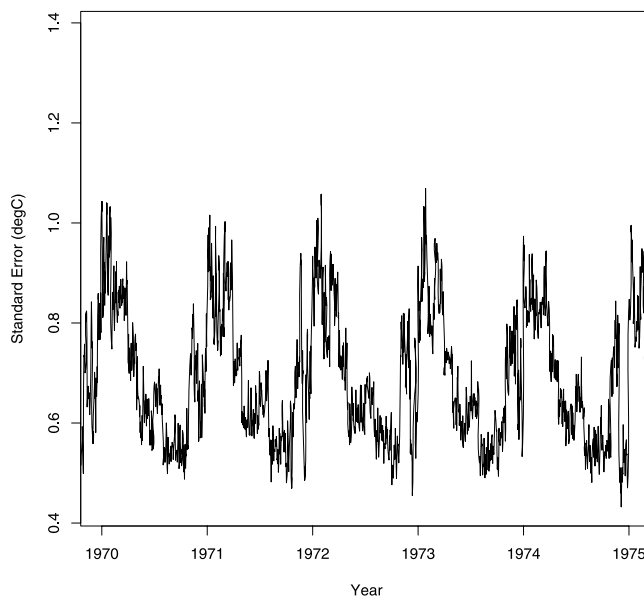


Figure 5. Daily mean standard error for 0.1° master grid for mean temperature averaged across the entire region for the years 1970–1974.

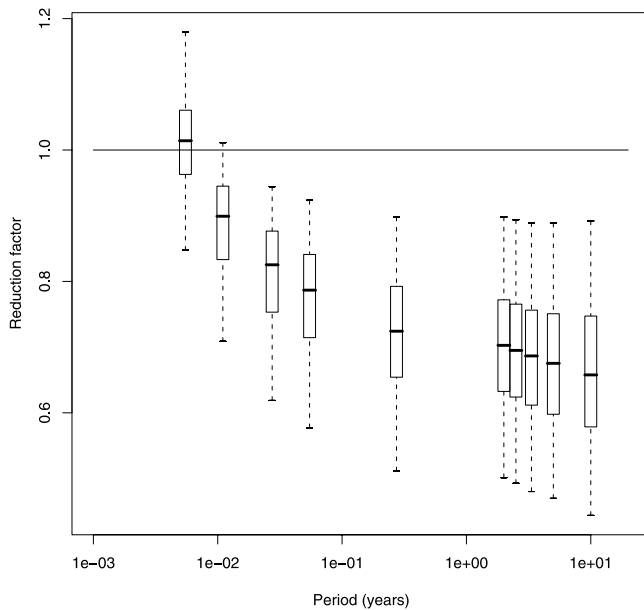


Figure 6. Areal reduction factor for daily quantiles of precipitation from the median (50% quantile) up to the 10-year return level. Bars show the variation across the stations, marking the median, 25% and 75% range (box) and the 5% and 95% range (dashes).

variance-covariance variance of \mathbf{x} . When f is the mean of \mathbf{x} , equation (2) reduces to

$$\sigma_f^2 = \frac{1}{n^2} \sum_i \sum_j M_{ij}, \quad (3)$$

where n is the number of dependent variables contributing to the mean.

4. Extremes of Observed and Interpolated Data

[48] With increasing attention to analysis of extremes, the gridded data set will likely be used for validation of extremes in climate models as well as looking at changing extremes. We have therefore briefly examined the extremes in the gridded data compared with the station observations.

[49] There are two elements in the interpolation methodology that affect the behavior of extremes. The first is the smoothing introduced by the spline and kriging interpolation, as these nonexact interpolators will smooth out peaks and troughs in a surface. The second is the methodology of creating grid square averages by interpolating to a high-resolution grid then averaging this to coarser resolution.

[50] To examine the effect of the kriging smoothing we compared the extremes in the station data against those in a cross validation data set, by selectively removing each station then using its neighbors to interpolate to the missing station's location. This latter cross-validated data set was used in an earlier study for determining the best interpolation method [Hofstra et al., 2008]. We calculated the magnitude of the extremes in both data sets for various higher annual quantiles and return periods and compared the two by calculating the reduction factor as either a propor-

tional decrease in the return level for precipitation or the anomaly for maximum temperature.

[51] Figures 6 and 7 show the reduction factor for precipitation and the reduction anomaly for maximum temperature. Quantile amounts were calculated as empirical quantiles using the empirical distribution function, while higher return period levels were calculated using L moments with generalized extreme value code provided by Hosking [1990]. The reduction factors show a clear reduction in all extremes higher than the annual 75th percentile for precipitation and 90th percentile for maximum temperature. The median reduction for the 10-year return level is a factor of 0.66 for precipitation or an anomaly of -1.1°C for maximum temperature. However, for some stations the precipitation 10-year return level could be reduced by more than half, or the maximum temperature intensity by more than 3°C .

[52] The reduction of extremes can also be seen in a map of the precipitation 10-year return level for the stations and cross-validated data (Figures 8a and 8b). Note that Figures 8a and 8b are only for the central part of our gridding domain. We have plotted the values by interpolating to a regular grid using an exact interpolator, natural neighbor interpolation [Sibson, 1981], so as not to smooth the values further. The map of extremes at the station scale (Figure 8b) is generally of higher magnitude and with higher peaks than the interpolated data (Figure 8a). Also shown is the precipitation 10-year level for the 0.1° master gridded data set (Figure 8c). This shows similar magnitudes to the cross-validated data (Figure 8a). Since the cross-validated station data has retained the true climatology at each station (we interpolated only the anomaly then added it back to the true climatology at the omitted station), we can conclude that most of the smoothing in extremes is from the

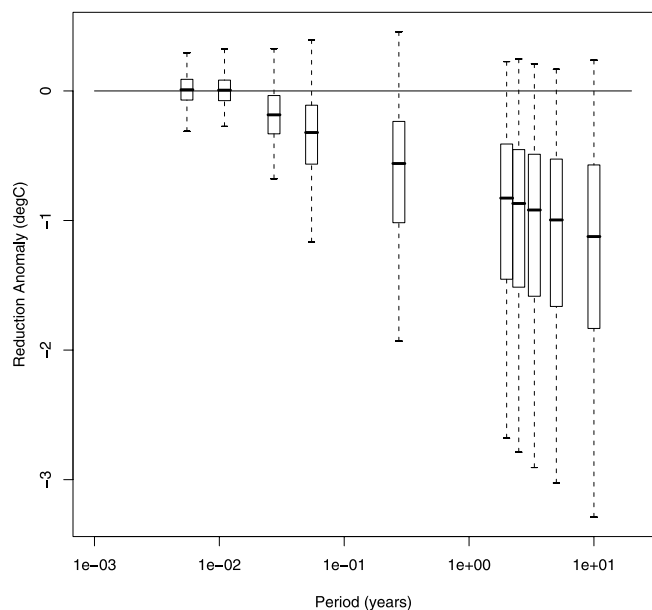


Figure 7. Areal reduction anomaly for daily quantiles of maximum temperature from the median (50% quantile) up to the 10-year return level. Bars show the variation across the stations, marking the median, 25% and 75% range (box) and the 5% and 95% range (dashes).

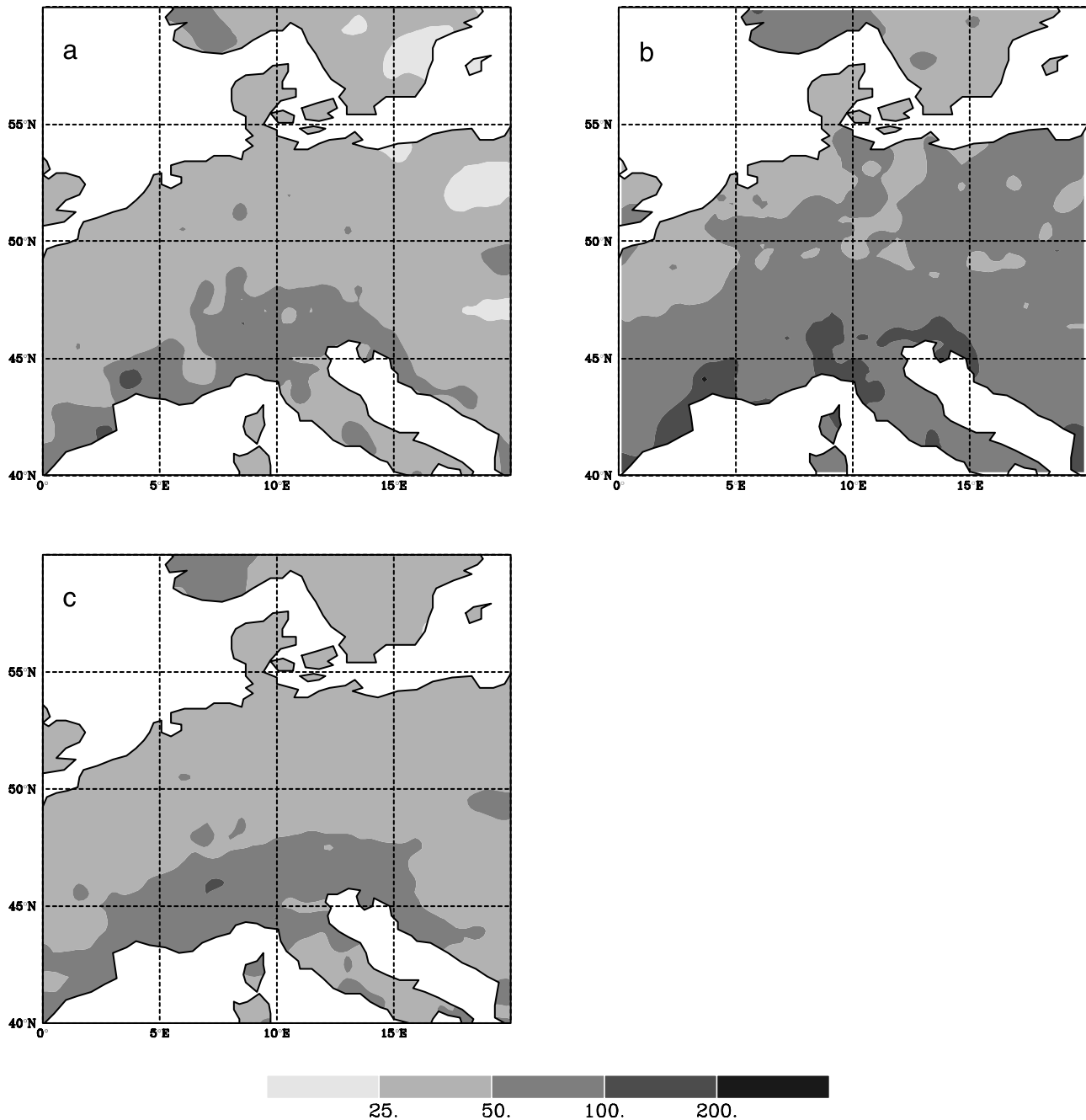


Figure 8. Ten-year return level for daily precipitation for (a) cross-validated station data, (b) station data, and (c) the final gridded data.

kriging of the daily anomalies, rather than from the interpolation of the climatology.

[53] The effect of averaging the 0.1° master grid to create 25 and 50 km data sets would be expected to reduce the extremes further (a desired effect), but since the interpolated fields are already smoothed, we would expect the effect to be less than the kriging smoothing.

[54] In conclusion, the interpolation methodology has reduced the intensities of the extremes, which is what would be expected for grid square average data. We should therefore be able to compare the extremes in this data set

directly with regional climate models at the same spatial scale.

5. Conclusion

[55] We have created a high-resolution European land-only daily gridded data set for precipitation and mean, minimum and maximum temperature for the period 1950–2006. This data set is unique in its spatial extent, resolution and the use of many more European observing stations than in other European or global sets.

[56] An important part of the data set is daily estimates of interpolation uncertainty, provided as standard errors. While interpolation uncertainty is the largest source of uncertainty in spatially interpolated data, there are other sources of uncertainty that would be worthwhile to include in future updates to this data set, including uncertainty related to measurement, homogeneity, and urbanization. A simple approach to model measurement error would be to assume a Gaussian distributed random error. For temperature, Folland et al. [2001] suggest this approach with a standard deviation of 0.2°C , an approach also adopted by Brohan et al. [2006]. A similar approach could be taken for precipitation. Homogeneity of records can be assessed using probabilistic methods [Peterson et al., 1998] to quantify the probability and magnitude of jumps in the record. Bias in temperature records can contribute to uncertainty, arising from sources such as thermometer exposure and urbanization. Folland et al. [2001] addressed both these issues and decided for thermometer exposure on a random error with standard deviation of 0.1°C pre-1900 reducing to zero by 1930. Urbanization, more relevant to the period of our data set, was handled similarly by Folland et al. [2001] with a Gaussian distributed random error with standard deviation increase of $0.0055^{\circ}\text{C}/\text{decade}$ since 1900. With all these uncertainty sources interacting in a complex manner, the most appropriate means to quantify these probabilistic errors in the final interpolated result would be by using stochastic simulations [Deutsch and Journel, 1998; Webster and Oliver, 2001].

[57] Spatial interpolation has a large impact on the magnitudes of extremes. We showed that the largest smoothing of the extremes occurs in the interpolation of daily anomalies. Using a data set of cross-validated station observations, we showed that the median reduction for the 10-year return level is a factor of 0.66 for precipitation or an anomaly of -1.1°C for maximum temperature.

[58] **Acknowledgments.** ENSEMBLES is a research project (contract GOCE-CT-2003-505539) supported by the European Commission under the 6th Framework Programme 2002–2006.

References

- Ali, A., T. Lebel, and A. Amani (2005), Rainfall estimation in the Sahel. Part I: Error function, *J. Appl. Meteorol.*, *44*, 1691–1706, doi:10.1175/JAM2304.1.
- Atkinson, P. M., and C. D. Lloyd (1998), Mapping precipitation in Switzerland with ordinary and indicator kriging, *J. Geogr. Inf. Decis. Anal.*, *2*, 65–76.
- Barancourt, C., and J. D. Creutin (1992), A method for delineating and estimating rainfall fields, *Water Resour. Res.*, *28*, 1133–1144, doi:10.1029/91WR02896.
- 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.
- Bougeault, P., P. Binder, A. Buzzi, R. Dirks, R. Houze, J. Kuettner, R. B. Smith, R. Steinacker, and H. Volkert (2001), The MAP special observing period, *Bull. Am. Meteorol. Soc.*, *82*, 433–462, doi:10.1175/1520-0477(2001)082<0433:TMSOP>2.3.CO;2.
- 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 dataset from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- 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, doi:10.1029/2005JD006280.
- Craven, P., and G. Wahba (1978), Smoothing noisy data with spline functions, *Numer. Math.*, *31*, 377–403, doi:10.1007/BF01404567.
- Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 369 pp., Oxford Univ. Press, Oxford, UK.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski (2007), Validation of satellite rainfall products over East Africa's complex topography, *Int. J. Remote Sens.*, *28*, 1503–1526, doi:10.1080/01431160600954688.
- Ekstrom, M., P. C. Kyriakidis, A. Chappell, and P. D. Jones (2007), Spatiotemporal stochastic simulation of monthly rainfall patterns in the United Kingdom (1980–87), *J. Clim.*, *20*, 4194–4210, doi:10.1175/JCLI4233.1.
- Folland, C. K., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, *28*, 2621–2624, doi:10.1029/2001GL012877.
- Frei, C., and C. Shär (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.
- Gleason, B. E. (2002), *Data Documentation for Data Set 9101, Global Daily Climatology Network*, V1.0, 26 pp., Natl. Clim. Data Cent., Asheville, Tenn.
- Goodman, L. (1960), On the exact variance of products, *J. Am. Stat. Assoc.*, *55*, 708–713, doi:10.2307/2281592.
- Goovaerts, P. (2000), Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol. Amsterdam*, *228*, 113–129, doi:10.1016/S0022-1694(00)00144-X.
- Grimes, D. I. F., E. Pardo-Igúzquiza, and R. Bonifacio (1999), Optimal areal rainfall estimation using raingauges and satellite data, *J. Hydrol. Amsterdam*, *222*, 93–108, doi:10.1016/S0022-1694(99)00092-X.
- Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. Goodess (2006), Downscaling heavy precipitation over the UK: A comparison of dynamical and statistical methods and their future scenarios, *Int. J. Climatol.*, *26*, 1397–1415, doi:10.1002/joc.1318.
- Hennessy, K. J., R. Suppiah, and C. M. Page (1999), Australian rainfall changes, *Aust. Meteorol. Mag.*, *48*, 1–13.
- Hofstra, N., M. R. Haylock, M. New, P. D. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily European climate data, *J. Geophys. Res.*, doi:10.1029/2008JD010100, in press.
- Hosking, J. R. M. (1990), L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc. Ser. B.*, *52*, 105–124.
- Hutchinson, M. F. (1993), On thin plate splines and kriging, in *Computing in Science and Statistics*, edited by M. E. Tarter and M. D. Lock, pp. 55–62, Univ. of Calif., Berkeley, Calif.
- Hutchinson, M. F. (1995), Interpolating mean rainfall using thin plate smoothing splines, *Int. J. Geogr. Inf. Sci.*, *9*, 385–403.
- Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, 600 pp., Academic, London.
- 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.
- Kitandis, P. K. (1997), *Introduction to Geostatistics*, Univ. of Cambridge Press, Cambridge, UK.
- Klok, E. J., and A. M. G. Klein Tank (2008), Updated and extended European dataset of daily climate observations, *Int. J. Climatol.*, in press.
- Kolmogorov, A. N. (1941), Interpolirovanie i ekstrapolirovanie stacionarnykh sluchainykh posledovatel' nostei, *Izv. Akad. Nauk SSSR*, *5*, 3–14.
- Krige, D. G. (1951), A statistical approach to some basic mine valuation problems on the Witwatersrand, *J. Chem. Metall. Min. Soc. S. Afr.*, *52*, 119–139.
- Kyriakidis, P. C., P. C. Miller, and J. Kim (2004), A spatial time series framework for simulating daily precipitation at regional scales, *J. Hydrol. Amsterdam*, *297*, 236–255, doi:10.1016/j.jhydrol.2004.04.022.
- Lebel, T., and A. Amani (1999), Rainfall estimation in the Sahel: What is the ground truth?, *J. Appl. Meteorol.*, *38*, 555–568, doi:10.1175/1520-0450(1999)038<0555:REITSW>2.0.CO;2.
- Lebel, T., G. Bastin, C. Obléd, and J. D. Creutin (1987), On the accuracy of areal rainfall estimation: A case study, *Water Resour. Res.*, *23*, 2123–2134, doi:10.1029/WR023i011p02123.
- Marquardt, D. W. (1963), An algorithm for least squares estimation of non-linear parameters, *J. Soc. Ind. Appl. Math.*, *11*, 431–441, doi:10.1137/0111030.
- Matheron, G. (1963), Principles of geostatistics, *Econ. Geol.*, *58*, 1246–1266.
- 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.
- Osborn, T. J., and M. Hulme (1998), Evaluation of the European daily precipitation characteristics from the Atmospheric Model Intercomparison Project, *Int. J. Climatol.*, *18*, 505–522, doi:10.1002/(SICI)1097-0088(199804)18:5<505::AID-JOC263>3.0.CO;2-7.

- Parker, D. E., and E. B. Horton (2005), Uncertainties in central England temperature 1878–2003 and some improvements to the maximum and minimum series, *Int. J. Climatol.*, 25, 1173–1188, doi:10.1002/joc.1190.
- Peterson, T. C., H. Daan, and P. D. Jones (1997), Initial selection of a GCOS surface network, *Bull. Am. Meteorol. Soc.*, 78, 2145–2152, doi:10.1175/1520-0477(1997)078<2145:ISOAGS>2.0.CO;2.
- 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.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1986), *Numerical Recipes: The Art of Scientific Computing*, 818 pp., Cambridge Univ. Press, Cambridge, UK.
- Sibson, R. (1981), A brief description of natural neighbor interpolation, in *Interpreting Multivariate Data*, edited by V. Barnett, pp. 21–36, John Wiley, New York.
- Teo, C.-K., and D. I. F. Grimes (2007), Stochastic modelling of rainfall from satellite data, *J. Hydrol. Amsterdam*, 346, 33–50, doi:10.1016/j.jhydrol.2007.08.014.
- Webster, R., and M. Oliver (2001), *Geostatistics for Environmental Scientists*, 271 pp., John Wiley, Chichester, UK.
- Yamamoto, J. K. (2000), An alternative measure of the reliability of ordinary kriging estimates, *Math. Geol.*, 32, 489–509, doi:10.1023/A:1007577916868.
-
- M. Haylock, PartnerRe, Bellerivestrasse 36, CH-8034 Zurich, Switzerland. (smaug@kagi.com)
- N. Hofstra and M. New, Climate Research Lab, Oxford University Centre for the Environment, South Parks Road, Oxford, OX1 3QY, UK.
- A. M. G. Klein Tank, Royal Netherlands Meteorological Institute, P.O. Box 201, NL-3730 AE De Bilt, Netherlands.
- E. J. Klok, TNO Built Environment and Geosciences, Princenlaan 6, NL-3584 CB Utrecht, Netherlands.
- P. D. Jones, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.