# Creating surface temperature datasets to meet 21st Century challenges

## Met Office Hadley Centre, Exeter, UK

## 7th-9th September 2010

## White papers background

Each white paper has been prepared in a matter of a few weeks by a small set of experts who were pre-defined by the International Organising Committee to represent a broad range of expert backgrounds and perspectives. We are very grateful to these authors for giving their time so willingly to this task at such short notice. They are not intended to constitute publication quality pieces – a process that would naturally take somewhat longer to achieve.

The white papers have been written to raise the big ticket items that require further consideration for the successful implementation of a holistic project that encompasses all aspects from data recovery through analysis and delivery to end users. They provide a framework for undertaking the breakout and plenary discussions at the workshop. The IOC felt strongly that starting from a blank sheet of paper would not be conducive to agreement in a relatively short meeting.

It is important to stress that the white papers are very definitely not meant to be interpreted as providing a definitive plan. There are two stages of review that will inform the finally agreed meeting outcome:

1. The white papers have been made publicly available for a comment period through a moderated blog.
2. At the meeting the approx. 75 experts in attendance will discuss and finesse plans both in breakout groups and in plenary. Stringent efforts will be made to ensure that public comments are taken into account to the extent possible.

# Spatial and temporal interpolation of environmental data

*Draft white paper for discussion at the international workshop: "Creating surface temperature datasets to meet 21st Century challenges", Met Office Hadley Centre, Exeter, UK, 7th-9th September 2010.*

Tom Smith, Phil Jones, Elizabeth Kent, Maurice Cox, Noel Cressie, Dick Dee, Richard Smith

## 1. Introduction

Environmental data analyzed to a regular spatial and temporal grid is often desired for monitoring and climate studies. For example, monitoring of regional to global temperature change and changes in the daily temperature range and extremes may use analyzed temperatures. We use the term 'analyses' in the broadest sense to encompass any form of transformation to a regular grid (so from simple gridding through to dynamical reanalyses). Resolution depends on the period and region of the analysis: typically coarser analysis grids correspond to longer periods and larger areas. Some analyses are updated in near-real time.

Land near-surface temperature analyses produced by UEA/CRU/MOHC, NOAA/NCDC, and NASA/GISS have all been used for climate monitoring and studies of historical variations. Each of these studies employs different quality control, and different amounts of smoothing, filtering, and interpolation to produce gridded fields. How well the mean and other features of the temperature are resolved in analyses depends critically on the analysis methods used. Here we discuss interpolation analyses and methods, paying regard to the inevitable uncertainty associated with environmental data, in an attempt to guide the development of improved analyses.

## 2. Characterization of input data uncertainties

Uncertainties associated with the input observations can be a major cause of uncertainty in the analysis grid values and must be quantified before choosing the interpolation method. Input uncertainties, reflecting both systematic (bias) and random effects are required for the implementation of all interpolation techniques. Establishing measuring instrument traceability is vital as a first step in combining observations from different sources. Further uncertainties arise from sampling. Systematic effects, correlated across observations, are usually considered the most problematic. Examples include temporally and spatially varying biases due to changing thermometer exposures, urbanization, evaporation from uninsulated buckets used to sample seawater, and under-catch by rain gauges. Every effort must be made to quantify and adjust for bias in the analysis input, the adjustment process itself being a further source of uncertainty (Joint Committee for Guides in Metrology JCGM 100:2008, p5). Further, the contribution to variability from unbiased random effects requires quantification.

Metadata describing observational instrumentation and methods are invaluable, but may be unavailable, particularly for historical observations. Where adjustments are applied, the relationship between the observed and analysis input data must be fully documented and the unadjusted data retained or recoverable through a databank. Evaluation of the residual bias is particularly challenging and may be the largest component of the uncertainty associated with large-area averages.

Random errors without bias, by definition, average to zero over many observations. Sources of random error include inaccuracies in the measurement, transmission and transcription errors, and lack of precision in an observation, its location or time. For monthly averages over regions containing a number of stations, there may be enough data to average out most random error (Brohan et al., 2006). However, analyses on shorter time and space scales may be much more contaminated by random instrument errors. Estimation of the random error of individual observations can be difficult. That is especially true for historical observations since information about instruments and methods is often unavailable. In some cases the distinction between random errors and bias is blurred. For marine data a bias in data from an individual ship can be considered as a random error if there are sufficient observations from other ships with different biases providing observations nearby. It is therefore important to account for both the number of observations and the number of different platforms in such cases to allow properly for error characterization. It should be noted that some random errors might not average to zero following data transformations or for derived variables such as surface fluxes that combine several variables in non-linear parameterizations.

Uncertainty due to inadequate sampling becomes more important as smaller regions or shorter periods are analyzed. Data sufficient to sample a 5° spatial and one month temporal region may badly under-sample scales of less than 1° spatial and daily. Some interpolation techniques fill unsampled regions with values inferred from statistical or dynamical relationships with values in regions that are more adequately sampled.

Statistical methods to quantify the uncertainty in observations are described by Smith and Cressie (2010). Typically the uncertainty and covariance structure are modeled using either a marginal statistical model or a hierarchical statistical model. Other techniques to evaluate uncertainty include comparisons with high quality observations, comparisons of observations made using different measurement methods, or the use of comparisons with model output such as feedback from assimilation into reanalysis.

## 3. Interpolation techniques

Analyses to a regular grid require interpolation, averaging and filtering of irregularly spaced and often sparse point measurements. Such interpolation may be carried out in a number of ways, and the analyst must make choices about how to derive the best product for the purpose, given the characteristics of the input data and the field to be constructed. Not all methods incorporate uncertainty in a direct manner. A summary of methods, focusing on kriging, can be found in Smith and Cressie (2010). Kriging is optimal linear spatial interpolation and is commonly used to construct gridded environmental analyses, although there are non-linear versions based on the hierarchical statistical model (e.g., Cressie and Wikle, 2011, Ch. 4). In meteorological and oceanographic applications kriging is often referred to as optimal interpolation. The underlying assumption of Gaussian linear models is expected to be acceptable for temperature and many other environmental variables. Precipitation is one exception where the assumption of Gaussian models may not hold and alternative techniques may be needed (Haylock et al. 2008, Hofstra et al. 2008). For some variables, it may be possible to transform the data prior to analysis to produce a new variable with a Gaussian distribution. Examples where data transformation is desirable include the analysis of wind speed, rainfall on large space and time scales, or of extreme values of many parameters. Temporal interpolation methods have developed largely independently of spatial methods. Spatio-temporal interpolation methods are discussed in considerable detail in Cressie and Wikle (2011).

Where sampling is sufficient, the analysis may begin by averaging values within the defined grid cells. Different averaging methods may be employed, and the analyst will usually try to choose a method that limits the variance of the average. These averaged values, which are assumed to be representative of their grid cells, can then be interpolated to propagate information to surrounding grid cells containing insufficient data to produce averages. For greatest accuracy, spatial interpolation may be limited to regions near grid cells with measurements. However, sometimes more complete analyses are required, and spatial covariance estimates may be used to produce interpolation to more distant regions. In addition, temporal covariance may be used to aid interpolation of regions that are not consistently sampled (e.g., Wikle and Cressie, 1999).

An alternative to a direct high-resolution analysis is producing analyses in stages. The basic analysis would have a coarse scale, perhaps monthly and 5º spatially. Such an analysis could be supported by the available data at most locations, beginning in 1900 or earlier. The next-stage analysis would be higher-resolution corrections to the first analysis. The higher-resolution corrections would be computed only in regions where data were sufficient to support it. In addition, the higher-resolution corrections could be forced to average to zero over the coarse grid, to keep the lower- and higher-resolution analyses consistent. Since the corrections do not involve large-scale variations, simpler statistics could then be used to produce them compared to a direct high-resolution analysis. A two-stage analysis of sea surface temperature (SST) similar to that outlined here is being developed and tested by R. Reynolds (personal communication), and Haylock et al. (2008) present a three-stage analysis for land temperatures. Johannesson et al. (2007) describe a statistical approach of this idea applied to globally extensive total-column-ozone data.

The analysis method used should allow grid-value uncertainties to be evaluated. These uncertainties are a consequence of random and systematic data errors, as well as analysis sampling errors. For a multi-stage analysis, the uncertainties at each stage of the analysis need to be evaluated, and methods need to be developed for combining them. The hierarchical statistical models are particularly adept at this. The appropriate errors to consider are a function of scale. For large-scale variations, errors of fine-resolution adjustments are not important. At larger scales, bias in errors may be appreciable, while at fine scales the effects of sampling may cause most uncertainty. Where no fine-resolution correction may be produced due to insufficient sampling, an uncertainty given by the variance of the correction may be assigned. However, it should be made clear what the errors represent and the limits of the analysis due to data errors or insufficient sampling.

The use of basic information about covariance at temporal and spatial scales can be extended to extremely data-sparse regions and periods by the use of multivariate analyses and dataset reconstruction methods. Typically a well-sampled period will be analyzed to determine the important modes of variability and the available data for a data-sparse period projected onto those modes. An example would be the use of sparse anomalously warm observations in the tropical eastern Pacific to construct the large-scale anomalies associated with El Niño. Such techniques are widely used in the construction of SST datasets. Relationships among variables may be used to generate fields of sparsely or unobserved quantities. An example is the use of relationships among SST, pressure and marine precipitation diagnosed from satellite observations to estimate fields of marine precipitation using SST and pressure observations for the pre-satellite era (Smith et al. 2009).

## 4. Reanalysis

A different approach to generating global fields, known as reanalysis (Trenberth et al. 2010), is through the synthesis of observations in the context of a physical model,. Reanalysis uses tools and techniques developed for numerical weather prediction (NWP) to assimilate meteorological observations into multidecadal global datasets. These datasets provide an estimate of the atmosphere's past evolution that encompasses both observed and unobserved (model-derived) physical parameters. A wide variety of space-based and ground-based observations can be combined in this manner.

Data assimilation techniques used for reanalysis are essentially statistical procedures, in which all available prior information about data uncertainties (e.g. biases, error covariances) is used to estimate the most likely state of the atmosphere, given the observations and the laws of physics as approximated by the model. The role of the model is to impose dynamical and physical constraints on the estimates and to infer information about unobserved parameters and data voids from the available observations. The equations of motion are used to interpolate observational information in space, time, and across parameters. Such interpolated fields provide the ability, for example, to extract wind information from surface-pressure observations, and to improve rainfall estimates based on satellite measurements of temperature and humidity.

Feedback from the assimilation of observations into reanalyses has proved valuable for quality control and data homogenization. Since reanalysis uses and compares observations from different sources in a single physical framework, it can help to expose data-quality issues. It has been demonstrated that the information overlap among different instruments can be effectively used in reanalysis to identify and correct biases in many of the data used (Dee and Uppala 2009).

Reanalysis also has the potential to guide the design of the observing system by providing information to help ensure that measurements are made in the right places with the right frequency (Trenberth et al. 2002). Reanalysis has proven to be an important tool for climate research; however, it should be remembered that errors in reanalysis interpolated fields due to model bias or due to changes in the observing system (which may not necessarily involve the variable of interest) may make them unsuitable for some applications.

## 5. Choice of interpolation technique

Each step of an analysis requires making choices to deal with data and physical modeling problems, and each choice needs to be carefully considered. For forming analyses within grid cells with observations, potential problems include random and systematic errors in observations and in models, the irregular distribution of observations and their density within analysis grid cells. For interpolation to larger regions, potential problems include the irregular and sometimes sparse distribution of stations over continents, which can cause large sampling errors in the analysis. All of these problems contribute to analysis uncertainty, which can change from place to place and time to time, and which is often incompletely understood by climate researchers who use the analyzed products.

Typically, anomalies from the annual cycle are interpolated, since anomalies tend to have larger scales and be less affected by topography compared to full temperatures. Forming anomalies is a type of data transformation that requires a base-period average (often referred to as a climatology). The base period may be a well sampled modern period of *in situ* data (such as 1961–90) that may be supplemented with satellite-based data. A separate interpolation should be performed for the absolute temperatures, incorporating elevation and other factors such as distance from coasts or other bodies of water. Absolute interpolated temperatures can be developed by adding the absolute to the anomaly-interpolated values.

Besides forming anomalies, it may be desirable to perform other data transformations to analyze temperature extremes better (particularly important when daily data are considered). Such transformations might be helpful for analyzing finer-resolution adjustments. For example, daily temperature extremes are often used as measures of climatic variation and their accurate representation in an analysis could be critical in some applications. A study would need to evaluate possible transformations and their influence on analysis of extremes. Various transformations have been tried for daily data (see, e.g., discussion in Haylock et al. 2008). Different climates in different parts of the world mean that it is unlikely that there is a single best transformation that could be universally applied. For daily temperature data, Haylock et al. (2008) found that the daily anomaly from the monthly mean worked very well. This approach has the advantage of forcing the daily average of the interpolated data to the monthly average, while still allowing different networks of daily and monthly data to be used.

The analyses themselves would likely be performed using a statistical model that incorporates covariance information to interpolate incomplete fields of data. If a coarse analysis is first performed followed by a finer-resolution analysis, it may be desirable to use different types of analysis for each stage. A reduced-space analysis using spatial empirical orthogonal functions or similar functions to define large-scale covariance may be best for a large-scale analysis. For a finer-scale analysis, exponential or similar covariance functions may be better for defining covariances for small-scale corrections. Although theory may be used to determine the best method for the analysis of ideal data, the actual available data are far from ideal. Therefore testing and evaluation of methods is required.

With all interpolation techniques (for temperature and pressure data) it is important to recognize that there will be a hierarchy of interpolations: anomaly and absolute at the monthly timescale and daily anomalies from the monthly average at the daily scale. For precipitation, the occurrence/non-occurrence nature of the variable means that other hierarchical combinations must be made. Simple anomalies do not work as well for precipitation and many have used percentage anomalies (as the variance is strongly related to the amount), but other transformations could be used. Moving to the daily scale involves other considerations. Haylock et al. (2008) used percentages of the monthly totals (ensuring conformity between the daily and monthly timescales), but in dry climates/seasons it is necessary not to forget the occurrence aspect. Over-smoothed interpolated fields will result if this issue is not addressed. The effect is most noticeable with extremes (see the next section).

The interpolation technique selected should have certain desirable statistical properties (unbiased, efficient, etc.). In addition to producing the analyzed grid values, the technique should provide output uncertainties (uncertainties associated with the grid values). Because each grid value depends on common information, the grid values have themselves covariances associated with them. These output uncertainties and covariances would be obtained by propagating the input uncertainties and covariances through the interpolation "model". When a multi-stage analysis is used, uncertainties would be propagated through each stage in turn.

The interpolation technique should be validated to ensure its acceptability in terms of such properties as fidelity (faithfulness to the raw data) and smoothness (not possessing spurious behavior). Whether or not an interpolation technique fully employs principles of approximation theory such as filtering, smoothing, and regularization, validation is important to test the technique

## 6. Application and examples

Besides near-surface land temperatures, historical analyses of other important climate variables have been developed, including SST, surface pressure, and precipitation. Many of these analyses are facilitated by satellite-based data that can be used to form statistics needed for the analysis of historical periods. Methods used for these analyses are often similar, and the knowledge and experience gained from their development should assist analysis improvements.

Some analyses of climate variables are over both land and ocean using consistent methods. As noted above, R. Reynolds is developing a high-resolution SST analysis by producing high-resolution (4 km daily) corrections for a lower-resolution analysis (25 km daily). The SST data are not sufficient for analyses of sub-daily variations. For land temperatures, a similar analysis could be developed, which could then be merged with the SST to provide a global high-resolution analysis. It is not clear whether data are sufficient for analyzing sub-daily land temperatures except in a few well-sampled regions. The highest resolution to be analyzed should be evaluated as part of analysis development.

Potentially, atmospheric reanalyses can be used to provide information about sub-daily variations in SST, by providing estimates of ocean surface winds and solar insulation via cloud, both of which affect the diurnal cycle in the SST. A more modest application of the same idea would use atmospheric information from reanalyses to improve estimates of daily SST variability in the pre-satellite era.

Applications for improved temperature analyses include studies for monitoring of changes of the mean and daily extremes. To perform these studies adequately, it is important that the extremes be well represented in the analyses. Some potential problems in representation of extremes are discussed in Haylock et al. (2008), who show that analyses may obscure some information on extremes that is present in raw data. High-resolution analyses or adjustments to lower-resolution analyses should be designed to minimize such problems. Figures 1 and 2 (from Haylock et al. 2008, for daily maximum temperature and precipitation data) illustrate some of the potential problems with interpolation of daily data. The figures show the reduction in the estimate of extreme values. This reduction is illustrated by calculating values of various extremes from the interpolated datasets compared to estimating the same extremes from the original station series and then interpolating these estimates. Across Europe, there is a reduction of ~1 °C for the 10-year return period extreme and about 75 % for a similar extreme daily precipitation estimate. For both variables, rare extreme estimates are reduced the most.

With combination of analyses of anomaly datasets from the land and the marine realms, there are decisions to be made at the boundaries (coasts and islands). The estimated accuracy of monthly averages depends on the number of samples, but the marked differences in the temporal correlation decay between land and SST values need to be carefully considered. It is expected that in the future more consistent approaches to analysis of land and ocean data will produce global datasets of higher quality than those presently available. Over the oceans, SST anomaly analyses have been produced using interpolation methods similar to those that can be applied to near-surface land temperatures. For example, Smith et al. (2008) discuss a merged SST and land temperature anomaly analysis, where SST and land analyses were separately produced using similar statistical analysis methods. However, the resolution of that analysis is coarse: monthly and 5º spatially. To improve the resolution of such an analysis would require higher-density base data for forming analysis statistics. Those statistics would need to be analyzed to ensure that they are stable at higher resolutions. In addition, the data to be analyzed would need to be sufficiently dense to be used with the higher-resolution statistics. Berliner et al. (2000) developed a spatio-temporal statistical 7-month-ahead forecast, with full uncertainty measures given for the forecast.

## 7. Presentation of interpolated data

Interpolated datasets must be properly documented and preferably presented in a self-describing data format. Each dataset should be uniquely identifiable through version control. Documentation should detail data sources, quality assurance, the interpolation methodology and parameters used, and how the associated (combined) uncertainties were calculated. The scales of variability resolved should be indicated and also when and where the scales change due to changes in the input data. Documentation should also explain how the uncertainties should be used to indicate where there might be problems with the raw data or the model. Besides the combined uncertainties, the analyses should include different uncertainty components (associated with random errors, bias, and sampling error) and documentation should explain how to use each to determine potential problems at different scales and for different applications. It may be desirable to include additional information alongside the interpolated data and the associated uncertainties, such as the covariances, the number of samples and stations or platforms, and data flags.

## 8. Summary and concluding remarks

The method used to construct interpolated datasets should be chosen based on characteristics of the input data and the field to be constructed. Any bias adjustments should be applied before analysis and the uncertainty due to the bias adjustment evaluated. The quality of the choice of method will impact on the resulting fields. All aspects of uncertainty should be quantified and estimates of data quality provided alongside the analyzed field. All sources of uncertainty should be taken into account as far as possible because of their influence on the reliability of conclusions inferred from the analysis.

It should be recognized that there would never be a single analysis for all uses. The best interpolation method depends on the question being asked; for example, kriging does a poor job for determining temperature extremes. Thus, links to and comparisons with other analyses should also be available. Such comparisons are now carried out for a number of climate variables, such as SST and precipitation, and many researchers find them useful. Communications between analysis groups, statisticians, and the greater climate-study community also should be encouraged, so that the analyst may more clearly know what is needed to serve that community.

## 9. Recommendations

- The choice of interpolation technique for a particular application should be guided by a full characterization of the input observations and the field to be analyzed. No single technique can be universally applied. It is likely that different techniques will work best for different variables, and it is likely that these techniques will differ on different time scales.
- Data transformations should be used where appropriate to enhance interpolation skill. In many cases, the simple transformation of the input data by calculating anomalies from a common base period will produce improved analyses. In many climate studies, it has been found that separate interpolations of anomaly and absolute fields (for both temperature and precipitation) work best.
- With all interpolation techniques, it is imperative to derive uncertainties in the analyzed gridded fields, and it is important to realize that these should additionally take into account components from observation errors, homogeneity adjustments, biases, and variations in spatial sampling.
- Where fields on different scales are required, interpolation techniques should incorporate a hierarchy of analysis fields, where the daily interpolated fields should average or sum to monthly interpolated fields.
- Research to develop and implement improved interpolation techniques, including full spatio-temporal treatments is required to improve analyses. Developers of interpolated datasets should collaborate with statisticians to ensure that the best methods are used.
- The methods and data used to produce interpolated fields should be fully documented and guidance on the suitability of the dataset for particular applications provided.
- Interpolated fields and their associated uncertainties should be validated.
- The development, comparison and assessment of multiple estimates of environmental fields, using different input data and construction techniques, are essential to understanding and improving analyses.

## 9. References


Berliner, L.M., Wikle, C.K., and Cressie, N., 2000: Long-lead prediction of Pacific SSTs via Bayesian dynamic models. *J. Climate*, **13**, 3953-3968.

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

Cressie, N. and Wikle, C.K., 2011: *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.

Dee, D.P., and Uppala, S., 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. R. Meteorol. Soc*., **135**, 1830-1841.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M. 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.

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

JCGM 100:2008. Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM). Joint Committee for Guides in Metrology. http://www.iso.org/sites/JCGM/JCGM-introduction.htm

Johannesson, G., Cressie, N., and Huang, H.-C., 2007: Dynamic multi-resolution spatial models. *Environm. And Ecol. Statist.*, **14**, 5-25.

Smith, R.L. and Cressie, N.A., 2010: Statistical interpolation methods, unpublished document, http://www.surfacetemperatures.org/.

Smith, T.M., Arkin, P.A. and Sapiano, M.R.P., 2009: Reconstruction of near-global annual precipitation using correlations with sea surface temperature and sea level pressure, *J. Geophys. Res*., **114**, D12107, doi:10.1029/2008JD011580.

Smith, T.M., Reynolds, R.W., Peterson, T.C. and Lawrimore, J., 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283-2296.

Trenberth, K.E., Dole, R., Xue, Y., Onogi, K., Dee, D., Balmaseda, M., Bosilovich, M., Schubert, S., Large, W., 2010, Atmospheric Reanalyses: A major resource for ocean product development and modeling, " in Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. & Stammer, D., Eds., ESA Publication WPP-306.

Trenberth, K.E., Karl, T.R. and Spence, T.W., 2002: The need for a systems approach to climate observations. Bulletin of the American Meteorological Society, 83, 1593–1602.

Wikle, C.K. and Cressie, N., 1999: A dimension-reduction approach to space-time Kalman filtering. *Biometrika*, **86**, 815-829.

## 10. Figures

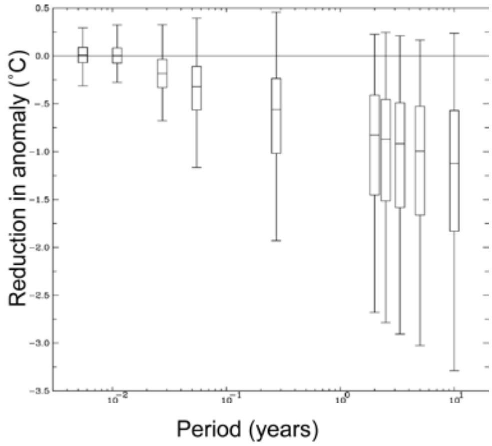


Figure 1: Areal reduction anomaly (y-axis in ºC) for daily quantiles of maximum temperature from the median (50 % quantile) up to the 10-year return level. Bars show the variation across all European stations, marking the median, 25 % and 75 % range (box) and the 5 % and 95 % range (dashes). (Figure 7 from Haylock et al., 2008.) The x-axis gives extremes from the median (on the left) through to the 10-year return period on the right.

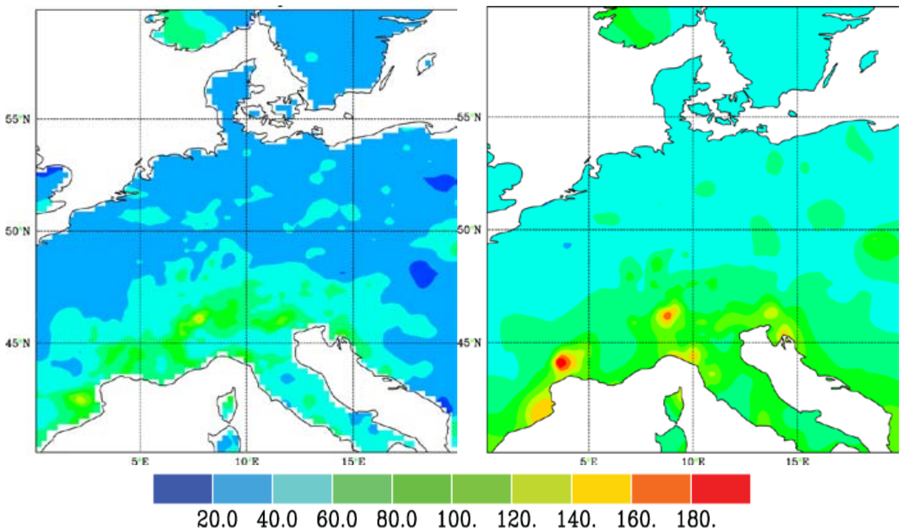


Figure 2: 10-year return period of daily rainfall extremes (mm, based on the period 1961–2006). The left panel is based on estimates of this extreme from the gridded database (E-OBS, Haylock et al., 2008) with the right panel gridded interpolation of the same extreme from the original station precipitation series.