Seventh Framework Programme
Theme 6

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D1.10
Gridding improvements

WP no:	1
WP leader:	URV
Lead beneficiary for deliverable :	UEA
Name of author/contributors:	Richard Cornes and Phil Jones
Nature:	Other
Dissemination level:	PU
Deliverable month:	24
Submission date: May 6, 2016	Version nr: 1

Report for Deliverable 1.10 (D1.10): Development of improvements to the gridding to enhance procedures during periods of extreme weather

By Richard Cornes[1 and 2] and Phil Jones[2]

[1] KNMI, Netherlands
[2] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

1. Introduction

The E-OBS gridded dataset is widely used for applications that require a complete gridded data field across European land. The dataset is updated twice yearly often with the addition of newly acquired data series, as well as on a rolling monthly basis. However, aside from bug corrections very little work has been undertaken to develop E-OBS since its initial development as part of the Ensembles project (Haylock et al. 2008). In this report we describe the improvements that have been made to E-OBS under the UERRA project, and which were conducted in order to fulfil the requirements of Deliverable D1.10 "Development of improvements to the gridding to enhance procedures during periods of extreme weather". In accordance with the Description of Work for UERRA (Annex I - Part B Description of Work, §2.2.2) we focus on improvements to the gridding of precipitation data using a gamma-transformation of the data. However, through the development of that method we discovered some further limitations of the E-OBS gridding-process that we have rectified through the development of a new gridding technique. This technique is applicable to all variables and exceeds the requirement of the Description of Work but since the technique marks a significant improvement to the E-OBS gridding procedure we also describe it in this report. This new technique also paves the way for the multiple-realization approach to gridding uncertainty that will be reported in a future Deliverable (D 1.14). The techniques described in this paper are currently being written into a paper for submission to a peer-reviewed journal.

2. Developing the Existing Gridding Process

The current method (Haylock et al., 2008) used in gridding the ECA&D precipitation data under the E-OBS scheme consists of five main stages:

1. At each station daily proportions are calculated relative to the station monthly rainfall total
2. The monthly totals are gridded to a high-resolution grid (0.1° rotated grid) using a trivariate thin-plate spline
3. The daily values (expressed as proportions of the monthly total) are gridded to the same high-resolution using indicator kriging
4. The gridded daily proportions are multiplied by the respective gridded monthly totals

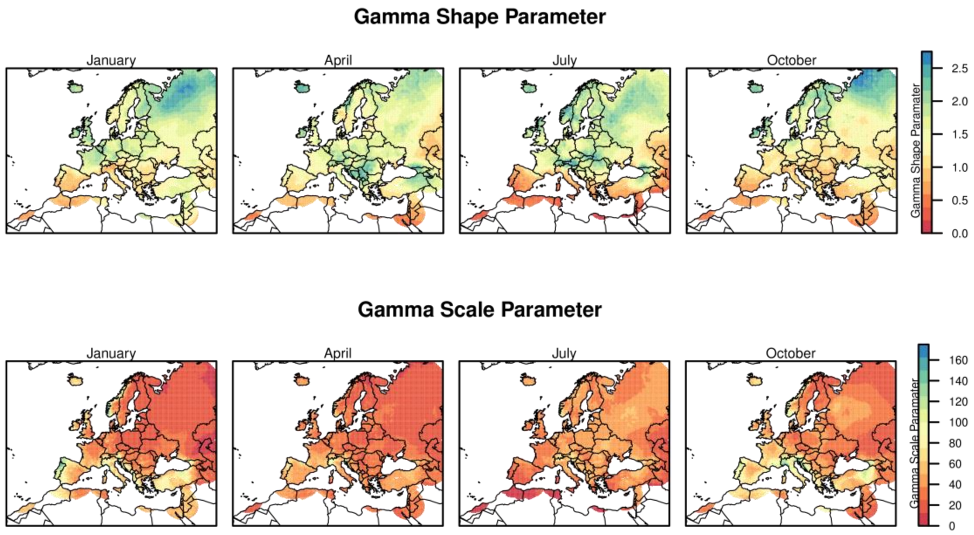


Figure 1: The gridded shape and scale parameters for four months of the year, calculated over the base period 1961-90.

5. The high resolution gridded rainfall data are averaged to a coarser grid resolution to provide spatial averages of rainfall

For the other variables gridded in E-OBS the same steps are taken except that the daily anomalies are relative to the monthly mean and the daily anomalies are gridded using simple kriging. The techniques described in this report relates to **Stage 2** of the gridding procedure described above. Since the daily anomalies are constrained by the monthly values, improvements to the monthly gridded data will have a significant effect on the final gridded daily data and should help reduce some of the over-smoothing of extreme values: a problem that was initially recognized by Haylock et al. (2008). In the following sections we describe two improvements that have been made to the E-OBS dataset: the use of a gamma-distribution transformation for rainfall data and the application of regression kriging to the gridding of all variables.

3. The Gamma-Transform Procedure

At each station the monthly rainfall totals are converted to probability estimates from a gamma distribution that is fitted to data over the 1961-90 base-period, with the shape and scale parameters at each station being retained. The two parameters were estimated using Maximum Likelihood Estimation (MLE) with the starting values for the optimization derived from L-moments (Hosking, 1990), following the example of Stagge et al. (2015). Both the monthly probabilities (between 0 and 1) and the shape/scale parameters are gridded (Figure 1) to the high-resolution “master” grid that is used throughout the E-OBS gridding procedure. To form monthly rainfall totals at the high-resolution grid, the probabilities are then converted back to absolute units using the gridded shape and scale parameters. These values were computed using Newton-Raphson iteration. To ensure that the gridded probabilities lie in the range [0,1] the probabilities were transformed using a logit transformation with an inverse logit calculated from the gridded data. Similarly, negative shape/scale parameters were prevented by gridding square-root transformed values. The gridding process then proceeds using these gridded monthly totals in stage 3 of the gridding process outlined above.

The reasoning behind this gamma-transformation follows that described by Bradley et al. (1987), in that small-scale spatial variations that typically occur in the rainfall field can prevent successful gridding of rainfall totals: the probabilities have a much smoother field that can be gridded more easily. As a reflection of this, a bi-variate spline (see Section 4) is used for gridding the probabilities,

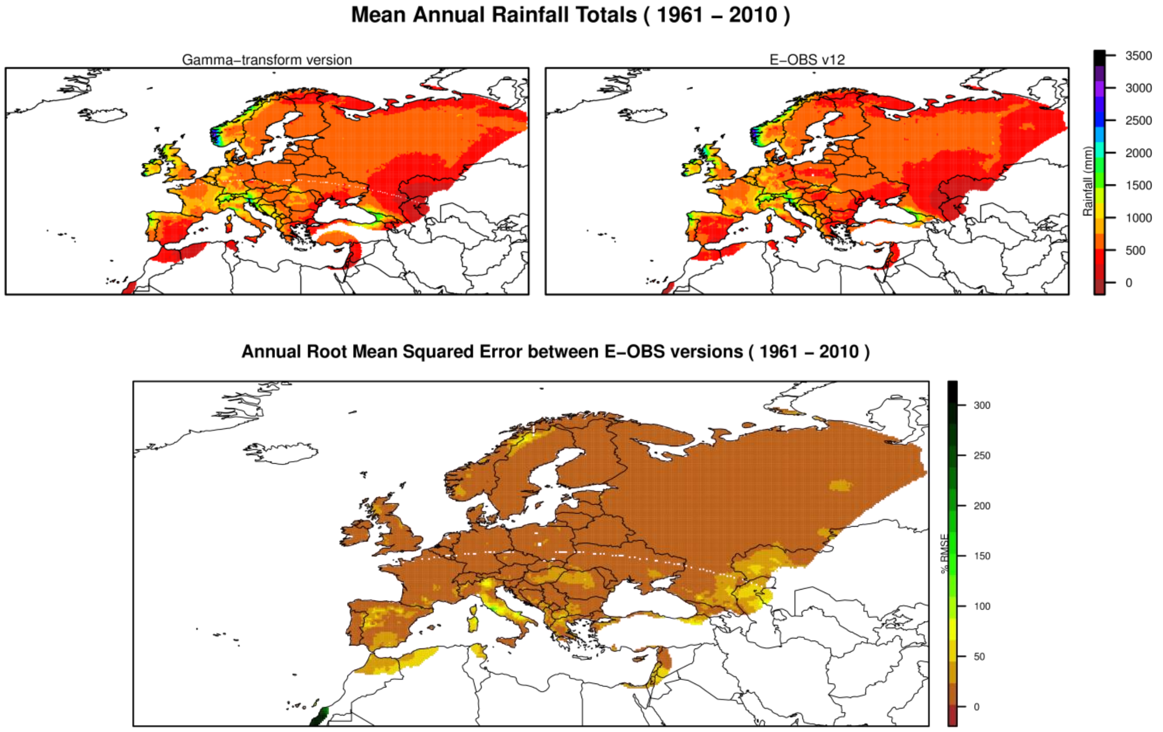


Figure 2: The average over 1961-2010 of the annual rainfall totals using the new Gamma-transformation procedure and the existing technique. Also shown in the lower plot is the root mean square error (RMSE) between annual totals from the two datasets again over the 1961-2010 period and expressed as a proportion of the annual rainfall total.

whereas a tri-variate spline is more appropriate for rainfall totals. The gamma distribution is chosen since the positive skewness that often occurs in rainfall data can be adequately captured by this flexible probability distribution but also because this distribution provides a better representation of very high rainfall extremes.

4. Evaluating the Gamma-transform technique

We have carried out a number of comparisons of the E-OBS data gridded using the new gamma-transformation technique and the existing gridding method. In Figure 2 we plot the mean annual rainfall totals in the two datasets along with the RMSE of the annual totals over the period 1961-2010, expressed as a proportion of the mean annual total in the original E-OBS dataset. In general the gamma-transform has the greatest effect across the south of the gridding domain, and across mountainous areas. The differences are not constant over time however, with the largest differences occurring after ca. 2000 (Figure 3). This is likely a result of the changing station density over time, and is likely affected by the lack of stations across France since 2005.

We have also compared the two gridded datasets by comparing the data against the high-resolution datasets produced by various National Meteorological Services (NMS) across Europe, in a similar manner to the evaluation procedures described in the UERRA deliverable D3.2. Since the NMS datasets contain many more station data than are available to E-OBS they should give a better measure of the true conditions. As a demonstration of this evaluation, in Figure 4 we have plotted the RMSE values of the two E-OBS versions against the high-resolution Alpine Precipitation Gridded Dataset (APGD, Isotta et al. 2014). From this figure it can be seen that the gamma-transformation technique removes some of the large differences over the high mountain areas that occur in the current E-OBS data.

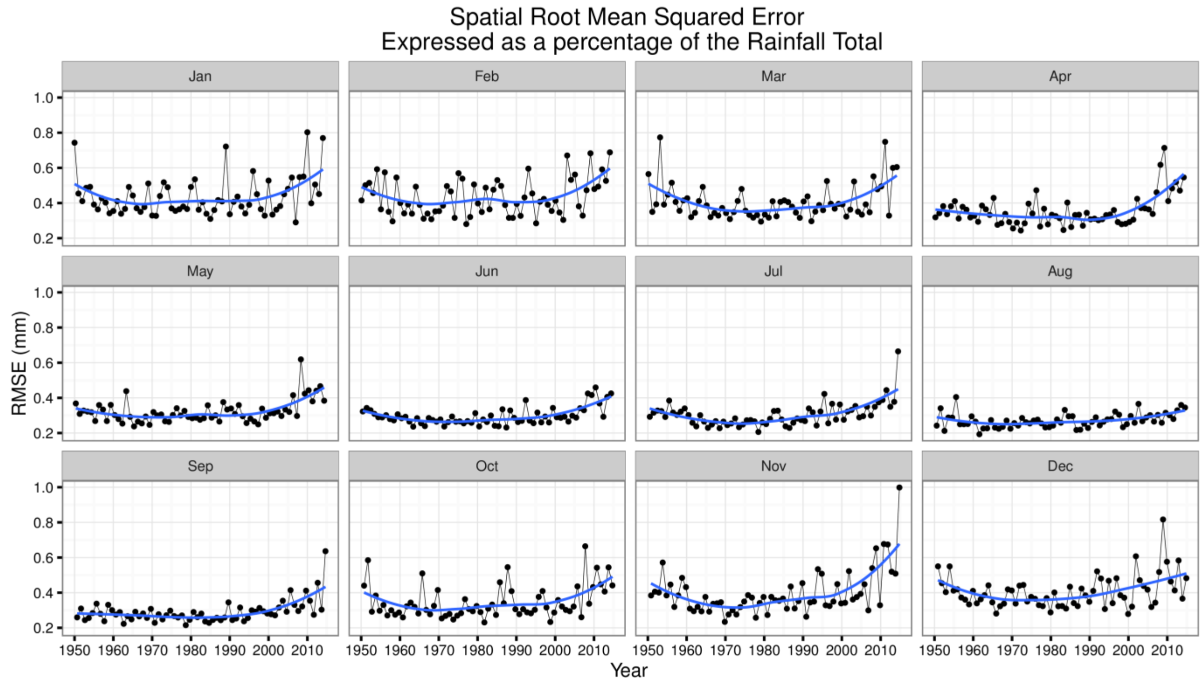


Figure 3: The RMSE across the gridding domain for each year, expressed as a percentage of the average total rainfall each year.

3. Regression Kriging

The original E-OBS version grids monthly station values by using a trivariate thin-plate spline in a manner such that

$$y_i = g(x_1 i, x_2 i, x_3 i) + \varepsilon_i \qquad i = 1, \dots, n$$

where the variable y_i a function of three predictors (latitude, longitude and altitude) for each of n observations for a given time period (monthly). g is a smoothing function that is estimated using

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 + \lambda J\,_m^d(\mathrm{f})$$

where the fitted values f_i are evaluated with respect to a smoothing parameter λ and by a roughness penality $\lambda J\,_m^d(\mathrm{f})$, which depends on the number of independent variables ($d = 3$) and an mth order differentiation that is chosen such that $2m > d + 1$. This expression represents a compromise between a close fitting of the data and achieving an appropriate degree of smoothness in f. Minimization is achieved using Generalized Cross Validation (GCV) to obtain an optimal value of λ which takes a positive real value. In the case of $\lambda = 0$ an exact interpolation is achieved.

A common problem with the use of GCV in the selection of the smoothing parameter is that a value of λ may be selected that is very close to zero, indicating over-fitting of the data and results in a spline that approaches an exact interpolator. This was found to occur in the fitting of a thin-plate spline to the E-OBS data. The result of such over-fitting is that a spline is produced that performs well for the areas of high-station density, but which produces an unrealistic result in the areas of poor station coverage. Furthermore, any errant station values can produce a large deviation in the smoothed surface, even in areas with good station coverage.

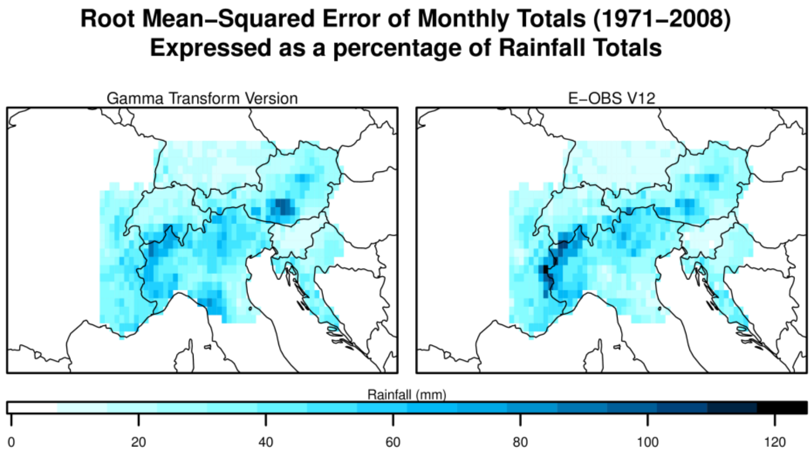


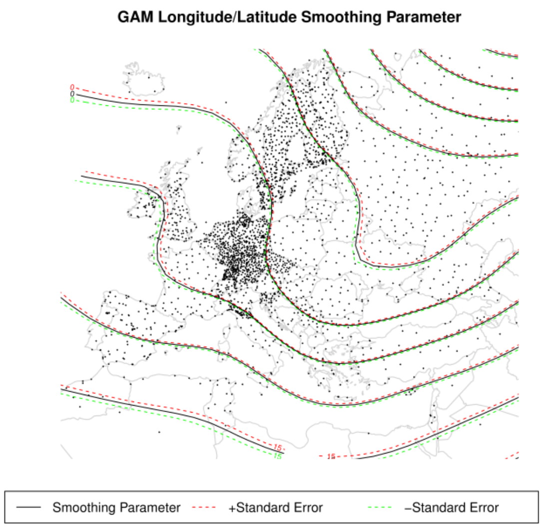
Figure 4: The RMSE of the two E-OBS version compared to the Alpine Precipitation Gridded Dataset (APGD).

To solve these problems we have developed a new method for the gridding of the monthly data, which uses a two-stage process to grid the monthly values. First a Generalized Additive Model (GAM) is fitted to the data of the form

$$y_i = g_1(x_{1i}, x_{2i}) + g_2(x_{3i}) \quad i = 1, \ldots, n$$

where the smoothing function g_1 is a thin-plate regression spline (Wood, 2003), and g_2 is a cubic smoothing spline. This model captures the spatial trend of the variable being gridded and to ensure that over-smoothing does not occur in the interpolation the effective degrees of freedom are restricted to a low number, although it should be stressed that the smoothing parameters are still chosen by GCV within that restriction. Overall this model is a robust fit in the sense that it is much less vulnerable to the changing station density of stations over space and time that is an inherent problem with E-OBS, and which will be explored in more detail in Deliverable D1.13. However, while this model performs well overall it will miss a great deal of the local detail which is captured in those areas of Europe with a high station density. To capture this detail the residuals from the GAM are interpolated to a regular grid using simple kriging. This regression kriging approach is somewhat akin to the method currently used in E-OBS in the separation of the monthly interpolation (absolute values) and daily interpolation (anomalies from the monthly value), except that the residuals in this case are anomalous relative to the spatially smoothed trend. However, in a similar manner to the daily kriging, a single variogram is used for the residuals kriging which is derived as the average across all months in the gridding period (currently 1950-2015).

Regression kriging has been widely used in other disciplines (notably ecology) but has received relatively little attention for the gridding of climate data. With the problem of spatially and temporally variable station density in the ECA&D database the technique is particularly suited to the gridding of data under the E-OBS scheme. We have developed the idea of regression kriging further, however, by the use of a the Generalized Additive Model (GAM) in the first stage of the process: a generalized linear model is typically used for this (Hengl et al. 2007). The GAM offers more flexibility by allowing a non-linear relationship between the predictor and predictand. As a demonstration of this we plot in Figure 5 the two smoothing terms for one month (January 2000) for maximum daily temperature. Despite the spatially varying station density a smooth spatial spline is produced, with the altitude component representing a value close to the Dry Adiabatic Lapse Rate of -5° km^{-1} but with a degree of non-linearity in the rate of change.



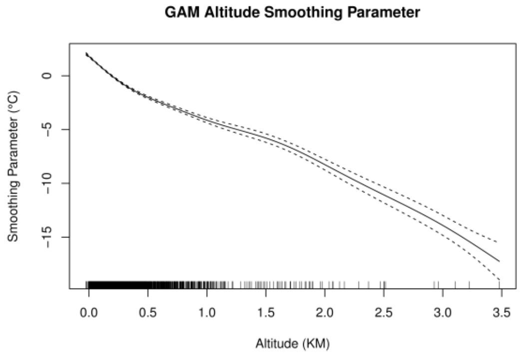


Figure 5: The longitude/latitude and altitude smoothing parameters for a GAM model fitted to maximum daily temperature values for January 2000. The black dots in the left figure indicate the station locations.

The use of this regression kriging technique also allows for the generation of an ensemble of gridding realizations with the uncertainty in the residual space being generated through Gaussian Conditional Simulation. This will described in a forthcoming deliverable (D1.14).

References

Bradley, R.S. et al. 1987: Precipitation fluctuations over Northern Hemisphere Land Areas since the Mid-19th Century, Science, 237, 171-175.

Haylock, M. R. et al., 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res., 113, D20119.

Hengl, T. et al. 2007: About regression-kriging: from equations to case studies, Computers & Geosciences, 33, 1301-1315.

Isotta, F.A. et al., 2014: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. Int. J. Climatol. 34, 1657-1675.

Stagge, J. H. et al. 2015: Candidate Distributions for Climatological Drought Indices (SPI and SPEI). Int. J. Climatol. 35, 4027–4040.

Wood, S.N. 2003: Thin plate regression splines, J.R.Statist.Soc.B 65, 95-114.