Seventh Framework Programme
Theme 6

Project: 607193 UERRA

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

Deliverable D1.11
Gridding resolution assessment

WP no:	1
WP leader:	URV
Lead beneficiary for deliverable :	KNMI
Name of author/contributors:	Richard Cornes
Nature:	Other
Dissemination level:	PU
Deliverable month:	30
Submission date: September 26, 2016	Version nr: 1

Report for Deliverable 1.11 (D1.11): Assessment of the potential for enhancing the gridding resolution in parts of Europe, together with more comprehensive comparisons with NMHS derived gridded products

By Richard Cornes[1 and 2]

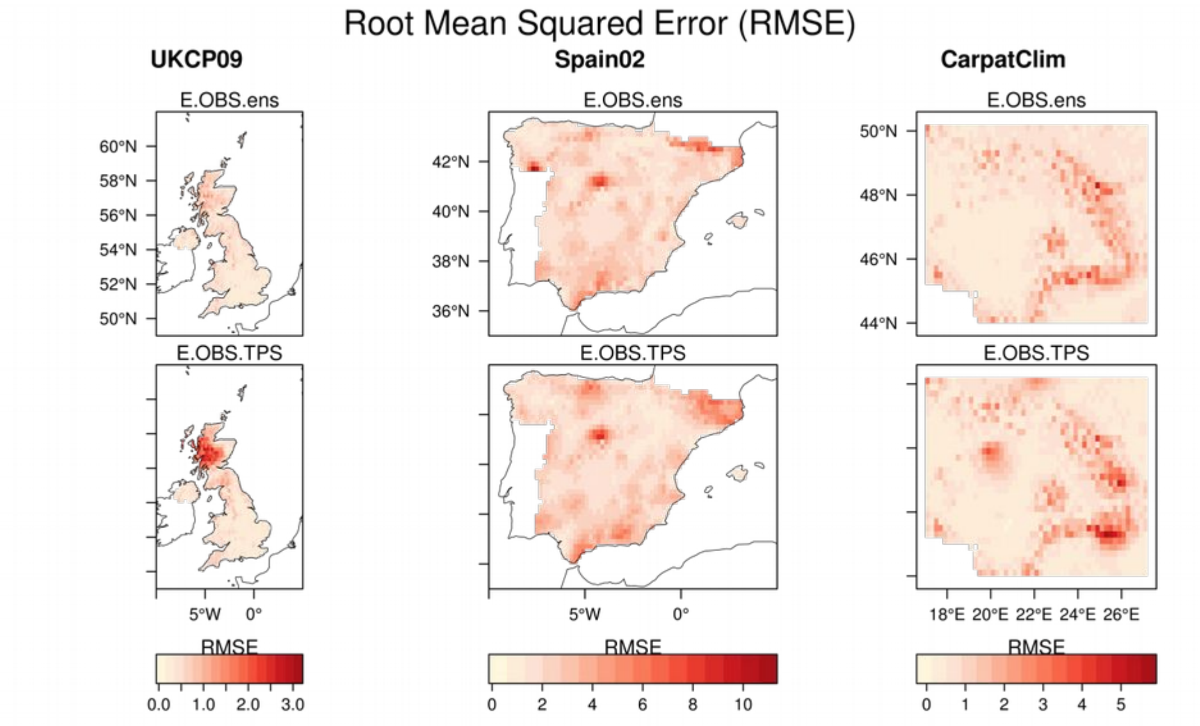
[1] KNMI, Netherlands
[2] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK

Summary

The regression-kriging approach (see Deliverable D1.10) is assessed through the comparison against selected high-resolution gridded datasets produced by National Meteorological Service centres across Europe. Reduced error is apparent in this comparison for the new gridding techniques in E-OBS indicating improvements to E-OBS that are a sole result of changes to the gridding technique; the input station data remain the same. We also highlight a recently discovered artefact in E-OBS that arises from the use of a rotated grid, and propose a solution to resolve this problem. We also assess the potential for increasing the grid-spacing (resolution) of E-OBS through the incorporation of additional environmental parameters into the gridding model.

1. Comparison against NMS high-resolution gridded datasets

In Deliverable report D1.10 we described an improvement to the gridding technique used in E-OBS (regression kriging), which provides a more stable spline interpolation and improves the assessment of uncertainty in E-OBS through the production of an ensemble of equally probable grid realizations. To assess the degree to which this new technique improves E-OBS, we have compared the gridded data produced by this new method against gridded datasets produced by most National Meteorological Services (NMS) across Europe. These datasets generally incorporate many more station series than are available to E-OBS and would therefore be expected to produce gridded datasets that are closer to the 'true' climate field. Indeed, it is the relative sparsity of stations used in E-OBS – and the spatial/temporal variability of the station density – that necessitates the three-stage gridding procedure used in E-OBS (see Haylock et al., 2008 and Deliverable D1.10); where station density is high most gridding procedures will produce similarly acceptable results (Ekström et al. 2007). In this report we focus on gridded datasets produced by three NMS across Europe: the UKCP09 data across the UK (Perry & Hollis, 2005), the SPAIN02 dataset across Spain (Herrera et al, 2012, 2016) and CARPATCLIM dataset, which covers the Carpathian Basin. Across the regions covered by these datasets the density of stations used in E-OBS is much less than is available to the NMS gridded datasets and the improvements offered by the new gridding technique in E-OBS can be assessed with respect to this reduced density. The different climate regimes of these three regions also provides a useful test of the new E-OBS gridding techniques. In this comparison the new version of E-OBS will be referred to as E-OBS$_{new}$, while E-OBS$_{old}$ is used to refer to the original dataset.



1.1 An ensemble-mean comparison

Figure 1: The Root-Mean Square error (RMSE) of maximum daily temperature between the new (E.OBS.ens) and old (E.OBS.TPS) E-OBS versions relative to three NMS datasets. The RMSE is in the unit of degrees Celsius, and have been calculated for each grid-cell for all months over the period 1971-2009. The new E-OBS dataset is the average across 100 ensemble members, which was calculated before the RMSE value was calculated.

In comparison to the UKCP09 dataset the new regression-kriging technique used in E-OBS for maximum daily temperature removes a significant artefact across western Scotland (Figure 1), where the traditional trivariate thin-plate spline used in E-OBS leads to a large error relative to the NMS gridded data. Across that region there are few stations available for use in E-OBS, and while the density in UKCP09 is lower than across more southerly regions of the UK, there are still more than are available to E-OBS. The Root Mean Squared Error (RMSE) statistic[i] used in that plot is calculated across all months in the period 1971-2009. Across other areas of the UK the error in both E-OBS versions is similar, relative to UKCP09. Across the other two regions evaluated in Figure 1 (Spain and the Carpathian Basin) the RMSE results are generally similar in E-OBS$_{new}$ and E-OBS$_{old}$, with the exception of certain restricted grid-cells. To highlight the differences between E-OBS$_{new}$ and E-OBS$_{old}$ we have plotted (Figure 2) the difference in RMSE between the E-OBS and NMS datasets. This highlights further the regions were E-OBS$_{new}$ is closer to the NMS data: these are regions in blue. Areas of comparable error are cream, and worse grid cells are shown in red. Clearly most grid cells in the new E-OBS versions are the same or closer to the NMS data than E-OBS$_{old}$.

The improvement in the new E-OBS gridding technique highlighted across Scotland in the UKCP09 comparison of Figures 1 & 2 likely arises from the use of the Generalized Additive Model (GAM, see D1.10). In the old E-OBS version the thin-plate spline almost always is an exact interpolator, which means that the spline fits through each of the station data points with little or no smoothing. As a result the spline produces abrupt changes in regions of rapidly changing topology and/or in data-sparse regions. The GAM is not so vulnerable to such features. Similar features are apparent across other mountainous areas of Europe (not shown).

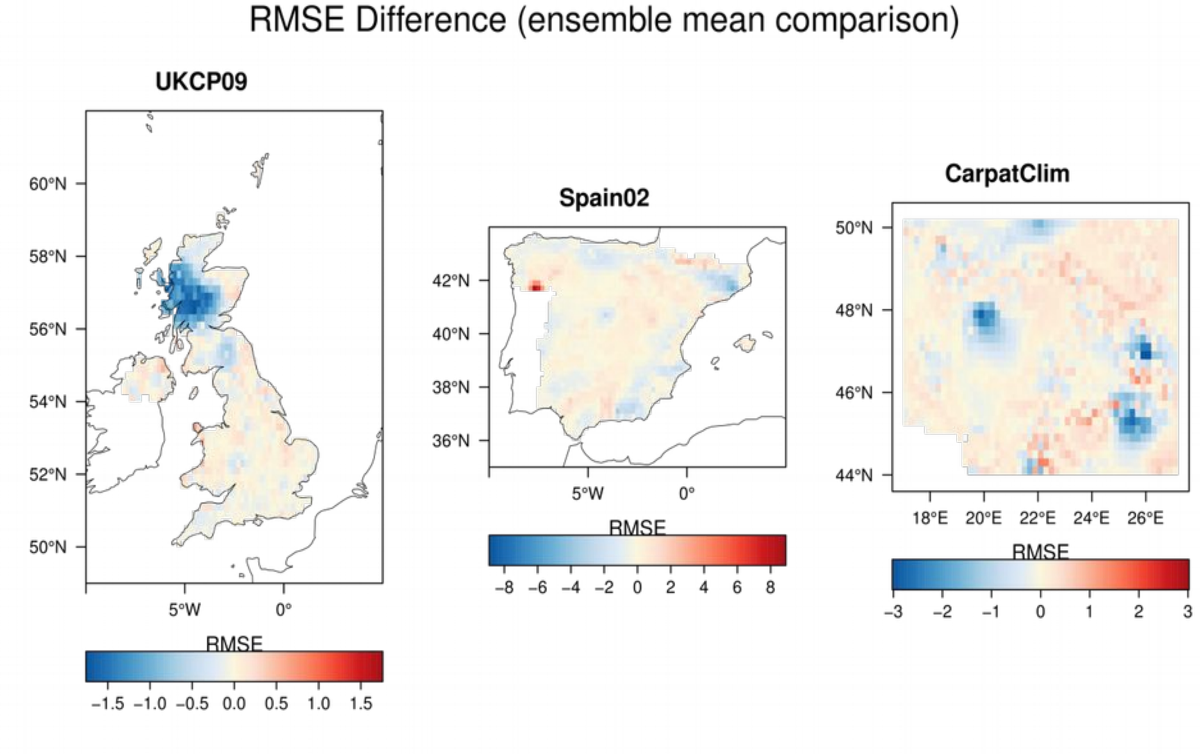


Figure 2: The difference between the RMSE of the E-OBSnew ensemble mean and E-OBSold relative to the NMS data.

1.2 An Ensemble-Range Comparison

The regression kriging method developed to improve E-OBS allows for the production of an ensemble of grids through the conditional simulation of the residuals of the GAM model (see D1.10). As such we are able to assess the E-OBS dataset relative to the NMS gridded data with respect to the uncertainty in the interpolation. We have calculated the RMSE of the E-OBS versions against the NMS data but contrary to the discussion above (§1.1) this has been calculated in this section for each ensemble realization of the new E-OBS dataset. At each grid-cell we have calculated the minimum and maximum RMSE against the NMS data. It must be noted that it is likely that different ensemble members produce these RMSE values, i.e. the grid-cells are considered in insolation across the ensemble. Furthermore, it must be stressed that the ensemble provided in this regression kriging approach is quite different to a model ensemble, which uses different initial conditions; in the E-OBS ensemble it is the station density that determines the ensemble spread. This will be discussed further in the forthcoming Deliverable D1.14.

The difference of these RMSE values against the RMSE from the old E-OBS version are plotted in Figure 3, again for maximum daily temperature over all months 1971-2009. As with Figure 2, blue values in this plot indicate that E-OBS$_{new}$ is better than E-OBS$_{old}$, in the sense that the RMSE against the respective NMS data is lower. The minimum values on the whole show a reduced error of E-OBS$_{new}$ compared to E-OBS$_{old.}$ Again the improvement across western Scotland in the UKCP09 data is a distinctive feature of these results. The maximum values in these plots tend to be red, which indicates that the largest RMSE value in the E-OBS$_{new}$ ensemble is higher than is found for E-OBS$_{old.}$ This is to be expected as this shows the maximum value away from the spline-interpolated value. Interestingly, however are the blue values in these maximum RMSE-difference values, which show where at *worse* E-OBS$_{new}$ is an *improvement* over E-OBS$_{old.}$ Areas where this is the case tend to be the more mountainous regions, which again highlights the deficiencies of E-OBS$_{old}$ in interpolating the ECA&D station data in areas of sharp topographic contrast.

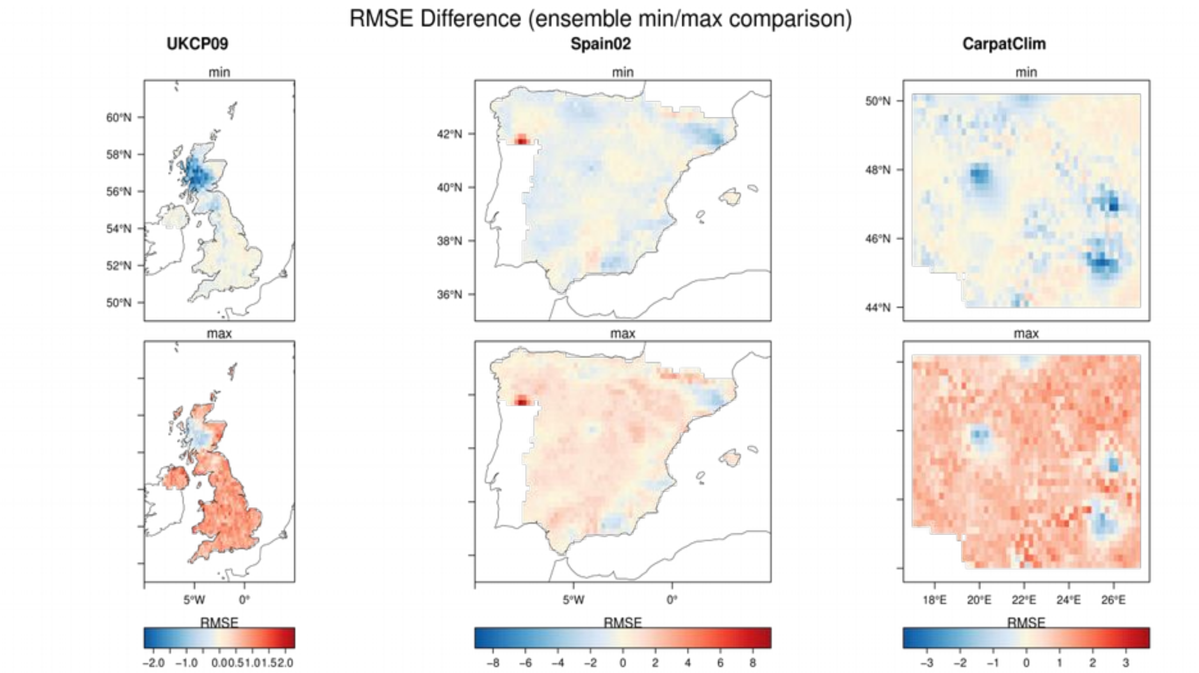


Figure 3: The minimum and maximum RMSE between the NMS data and E-OBS, across the E-OBS ensemble. These values are calculated on a per grid-cell basis

2. The box-averaging artefact

The E-OBS gridding scheme follows a three-stage process, which is described in Deliverable D1.10. The gridding process initially interpolates the input station data to a 0.1° grid-spacing; these values are then averaged to various coarser resolutions. This aims to construct a dataset of box-average values that replicate a Regional Climate Model (RCM) simulation in terms of spatial variance. The 0.1° "master grid" has a rotated pole in order to reproduce the grid used in many RCMs. The final E-OBS dataset is produced on a variety of coarser grids and on both the rotated grid and on a grid with regular latitude-longitude coordinates. In order to produce the regular grids the master-grid is first rotated back to the regular coordinate grid and then the values are aggregated to the coarser grid through box-averaging. However, a result of this rotation is that moving northwards across the gridding domain gradually fewer master-grid cells are used in the average for each coarse grid-cell. This is demonstrated in Figure 4, which plots the number of component master-grid points in each cell at the 0.25 resolution. As well as the reduction in component points moving north over the domain, there is also a more complex artefact of rotation evident. This artefact became particularly noticeable from the production of the ensemble grid in E-OBS$_{new}$ as part of the improvements to E-OBS as part of this project, and results in a final dataset that has a spatially uneven variance on account of the rotation of the master-grid points.

To resolve this we have used a master grid that has been projected onto the Lambert Equal Area grid. This ensures that the semi-variogram for the kriging interpolation is positive definite on a sphere, but also allows the master-grid coordinates to be converted back to a regular grid without distortion. These values can then be aggregated to a coarser regular grid at various grid-spacings, but with an equal number of component points per coarse grid-cell across the gridding domain.

A complication arises, however, in that the coarse grids are dependent on the starting points and grid-spacing of the master grid. Hence using the 0.1° grid-spacing in the master grid we are not able to produce the 0.25° regular grid, which is the most widely used E-OBS product. This could be achieved if a higher resolution grid (0.05°) was used and this is being investigated, but even in that case the grid would be offset from the usual starting coordinates of the current E-OBS version. We therefore propose that box-averaging is conducted to a 0.2° resolution, and then these values are interpolated to the required grid-spacing, i.e. at a grid-spacing greater than 0.2°. This would also provide, in principle,

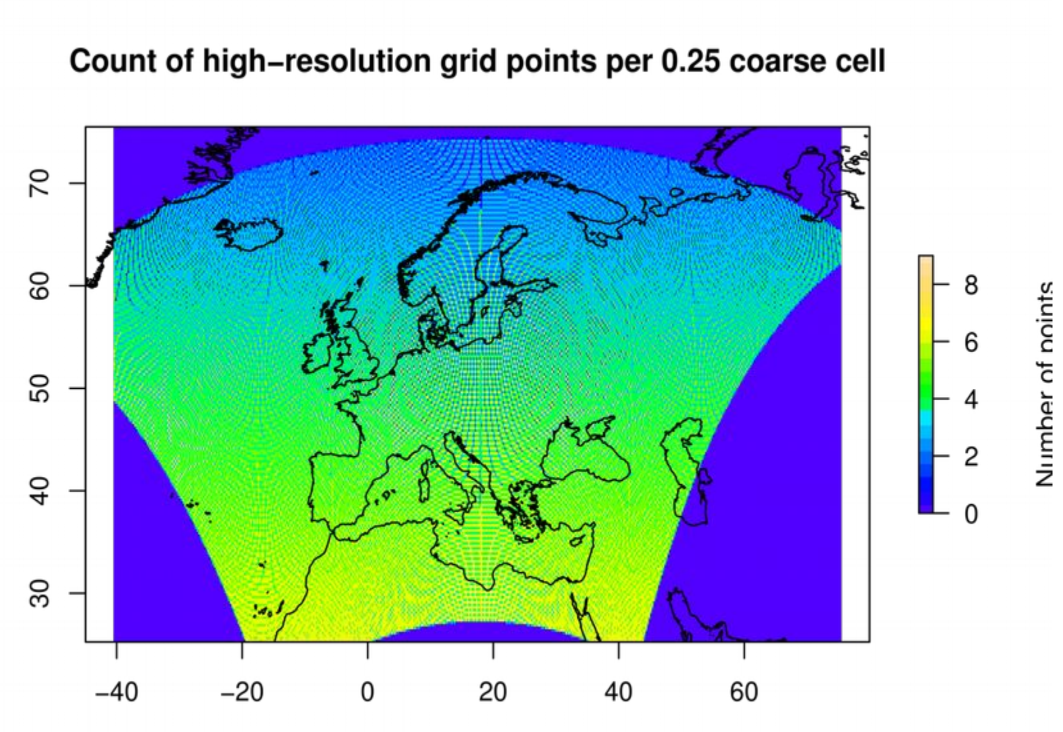


Figure 4: The number of high-resolution 0.1° master-grid points incorporated into each 0.25° grid cell in the E-OBS gridding scheme.

more flexibility for data users in choosing the target grid-spacing that is required. It should also be noted that the comparisons in Figures 1-3 of this report are conducted using E-OBS at the 0.2° resolution, using a Lambert Equal Area-projected master grid.

3. Incorporating additional environmental parameters in the gridding

A frequent request made to the ECA&D group at KNMI is for the provision of E-OBS on a higher resolution than the current highest resolution of 0.22°. The choice of resolutions of E-OBS is a legacy of the original aim of E-OBS, which was to provide a comparative dataset for the Ensembles RCM simulations (Haylock et al., 2008). With the latest suite of RCM simulations produced as part of the CORDEX initiative, the E-OBS dataset now lags behind the resolution of these RCMs (0.11° at the highest resolution). There is a clear requirement, therefore, for the provision of a European-wide gridded dataset at a grid-spacing finer than 0.22°.

When moving to the production of a version of E-OBS with a higher grid-spacing it becomes necessary to consider the scale of the climate phenomena that are being interpolated (Hutchinson & Bischof, 1983; Hutchinson, 1983). Sharples et al. (2005) indicate that for the interpolation of rainfall, a grid-spacing of 0.1° is optimal; this matches the grid-spacing currently used for the E-OBS master grid. When moving to a higher resolution E-OBS grid, the addition of additional environmental parameters – to supplement latitude, longitude and altitude that are currently used - should be considered (Daly et al., 2008). We have tested a range of such parameters derived from the GTOPO30 Digital Elevation Model (DEM, Figure 5). With the use of the Generalized Additive Model in the interpolation of E-OBS$_{new}$ such parameters are incorporated in an additive manner, and this flexibility is a further advantage of the use of these models in the interpolation of the ECA&D station data in E-OBS .

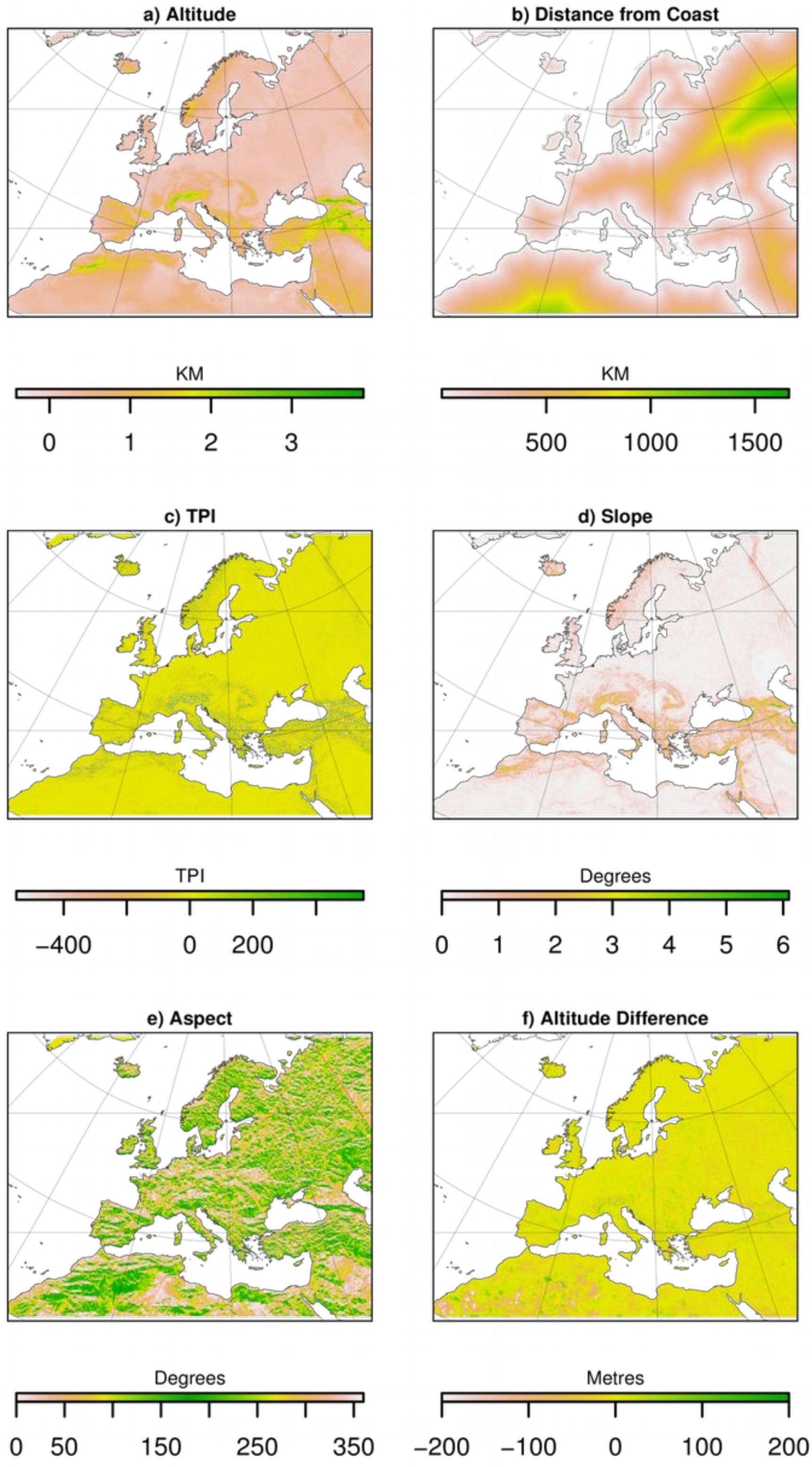


Figure 5: Maps showing the predictors used for interpolation at 0.1 degree resolution. a) The altitude in the GMTED2010 DEM, b) The distance to the coast, c) The topographic position index (TPI) which is he value of each altitude value relative to surrounding values, d) the slope angle, e) the aspect angle and f) the difference between the GMTED2010 and GTOPO30 DEM datasets. In b) the continental outline is at the scale 1:110m as used to define the coastal distance.

Further testing is required on the use of the parameters to E-OBS is currently being investigated, and will be reported as part of the forthcoming Deliverable D1.14.

Testing has also been carried out using the GMTED1010 DEM. These data supersede GTOPO30, and GMTED2010 is now the preferred DEM for continental-scale applications (Danielson & Gesch, 2010). The altitude values in GMTED2010 differ from GTOPO30 by as much as 100m across Europe (Figure 5f), although it should be noted that GMTED2010 uses GTOPO30above latitude 84°N. Using GMTED2010 for the E-OBS interpolation rather than GTOPO30 produces a very different result for both temperature and rainfall. However, when compared against the NMS gridded the interpolated data perform worse than when GTOPO30 is used, in terms of mean error statistics. This arises because most NMS in the gridding of their climate data use GTOPO30 or one of its derivatives. Hence, it is not that using the different DEM data produced a worse interpolation *per se,* but rather that a discrepancy results from the different elevation values, and because the gridding of the station data is extremely sensitive to altitude as a predictor. In the production of the new E-OBS version as part of the UERRA project the GTOPO30 DEM data will continue be used, but future assessments should be made to assess the influence of the DEM data on comparisons of gridded station data but also in the comparison of reanalysis and climate model simulations.

References

Danielson, J.J. and Gesch, D.B. 2010. Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). U.S. Geological Survey, U.S. Department of the Interior.

Ekström, M. Phaedon C. Kyriakidis, A. C., and Jones, P.D. 2007: Spatiotemporal Stochastic Simulation of Monthly Rainfall Patterns in the United Kingdom (1980–87). J. Climate, 20, 4194–4210, doi: 10.1175/JCLI4233.1.

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.

Herrera et. al. (2012) Development and Analysis of a 50 year high-resolution daily gridded precipitation dataset over Spain (Spain02). Int J Climatol, 32:74-85 DOI: 10.1002/joc.2256.

Herrera et. al. (2016) Update of the Spain02 Gridded Observational Dataset for Euro-CORDEX evaluation: Assessing the Effect of the Interpolation Methodology. Int J Climatol, 36:900–908. DOI: 10.1002/joc.4391.

Hutchinson MF (1998). Interpolation of rainfall data with thin plate smoothing splines, Part II: Analysis of topographic dependence, Journal of Geographic Information and Decision Analysis, 2: 168-185.

Hutchinson, M F; Bischof, R J (1983): A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. Australian Meteorological Magazine, 31(3), 179-184,

Perry, M.C., Hollis, D.M., 2005: The generation of monthly gridded datasets for a range of climatic variables over the UK. Int J Climatol, 25, 1041-1054.

Sharples, J.J., Hutchinson, M. F. and Damian R. Jellett, 2005: On the Horizontal Scale of Elevation Dependence of Australian Monthly Precipitation. J. Appl. Meteor., 44, 1850–1865, doi: 10.1175/JAM2289.1.

i The Root Mean Square Error (RMSE) is derived from variables x and y as $\sqrt{\frac{\sum_{i=1} n\left(x-y\right)^{2}}{n}}$.