Horizon 2020

H2020-EO-2014 New ideas for Earth-relevant Space Applications

EUSTACE

(Grant Agreement 640171)

EU Surface Temperature for All Corners of Earth

Deliverable D*1.6*

Report on Gridded data product for Europe, based on the homogenised LSAT

Deliverable Title	*Gridded data product for Europe, based on the homogenised LSAT*	
Brief Description	*Long temperature records are subject to several sources of uncertainty and may contain inhomogeneities that affect their long term stability. In D1.4, a homogenized station-based dataset for maximum and minimum daily temperature is constructed for Europe. The current deliverable is the gridded data product for Europe, based on these homogenised EUSTACE surface temperature records.*	
WP number	1	
Lead Beneficiary	*KNMI*	
Contributors	*Antonello Squintu, KNMI* *Gerard van der Schrier, KNMI*	
Creation Date Version Number Version Date	16/08/2017 2.0 -	
Deliverable Due Date Actual Delivery Date	M24 09/03/2018	
Nature of the Deliverable	R	*R - Report*
		DEM – Demonstrator, Pilot, Prototype
		DEC – Dissemination, Exploitation or Communication
		O - Other
Dissemination Level/ Audience	PU	*PU - Public*
		CO - Confidential, only for members of the consortium, including the Commission services

Version	Date	Modified by	Comments
2.0	20170817	Gerard van der Schrier	Full draft, submitted to Science Coordinator for comments
	09/03/2018	Nick Rayner	Submitted version. Minor text changes and clarification of data dissemination plan.

Table of Contents

1. Executive Summary

This document describes the procedure to process the daily homogenised land surface air temperatures (daily maximum and minimum temperature) based on the pan-European ECA&D dataset into a gridded product. The approach followed will be similar to the one developed by the FP7 UERRA project and should be seen as a further development of the approach introduced by Haylock et al. (2008). Since a newly developed gridding procedure has been used for this deliverable, grids based on the non-homogenized data are provided as well for comparison purposes. A more detailed documentation of the followed gridding procedure is provided by Cornes et al. (2018)

Once the methods have been fully published in the peer-reviewed literature, the resulting ensemble of 100 realisation of daily fields for Europe will be made publicly available.

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, Section B1.1):

No.	Objective	Yes	No
1	Intensively develop the hitherto immature use of Earth Observation estimates of Earth's surface **skin** temperature to enable new Climate Data Records of the surface **air** temperature Essential Climate Variable (ECV) to be created, for all locations over all surfaces of Earth (i.e. land, ocean, ice and lakes), for every day since 1850. EUSTACE will achieve this by: combining information estimated from multiple satellites with surface air temperature measurements made *in situ* and **creating complete analyses** of surface air temperature, through the application of novel statistical in-filling methods.	X	
2	Integrate these new daily surface air temperature Climate Data Records into a range of applications in Earth System Science and Climate Services and research, amongst others. EUSTACE will achieve this via the active and continuous engagement of trail-blazer users, and the provision of products through already-existing user community data portals and service mechanisms, in standard formats.		X
3	Undertake and report detailed research into the relationships between surface skin temperature estimated from Earth Observation satellite measurements and surface air temperature observed *in situ* by conventional measurements, over all surfaces of the Earth, including the polar regions. This is likely to provide information useful for refining coupling in Earth system models.		X

4	Create a sustainable, automated system at an appropriate level of maturity for the potential production of the products beyond the lifetime of the project. To enable this, EUSTACE will also identify Earth Observation and conventional data streams that could be used to update the surface air temperature Climate Data Records in the future, including those from Sentinel missions.	X	
5	Extensively validate the new surface air temperature Climate Data Records against independent, surface-based reference data, sourced by the project for this purpose.		X
6	Develop and report new, consistent, validated estimates of uncertainty both in already-existing Earth Observation surface skin temperature estimates and in the new surface air temperature Climate Data Records, at all locations and times across the Earth's surface.	X	
7	Develop links with related activities within Europe and beyond to help to ensure the execution of a joined-up work programme, the Copernicus Services and to enable the provision of requirements for the future surface skin temperature and surface air temperature observing system.	X	
8	Other – not directly linked to one of the above objectives		

3. Detailed Report

3.1 Gridding technique

The gridded dataset is developed using a modified approach from the original E-OBS gridding calculations documented by Haylock et al. (2008). The modified approach is documented by Cornes et al. (2018) and funded through the FP7 UERRA project.

In this novel version of E-OBS we use Generalized Additive Models (GAMs) for the interpolation. GAMs are an extension of generalized linear models, which themselves extend simple linear regression by allowing the independent variable to take a distribution other than Gaussian. GAMs extend this further by allowing the independent variable to be dependent upon one or more unknown smooth predictor variables. In this interpolation, models are formed that combine thin-plate and cubic spline bases: the former with two or more dimensions and the latter for one dimensional covariates. Thin-plate splines applied to large datasets are notoriously time consuming. Computational efficiency is achieved in the models used here by relying on the fact that the effective degrees of freedom (EDF) of a thin-plate spline model --- in this case the spatial EDF --- is often much less than (n) – with (n) the number of stations used for the interpolation. Therefore a low-rank approximation can be used which approximates a full thin-plate spline smooth but with a much lower computational burden. This is achieved by truncating the space of the roughness components of the spline using an eigen-decomposition based on Lanczos iteration. This reduces the magnitude of the operation from $O(n^3)$ to $O(n^2k)$, with (k) an estimate of the EDF, and such models are referred to as Thin-Plate Regression Splines (TPRS) (Wood, 2003).

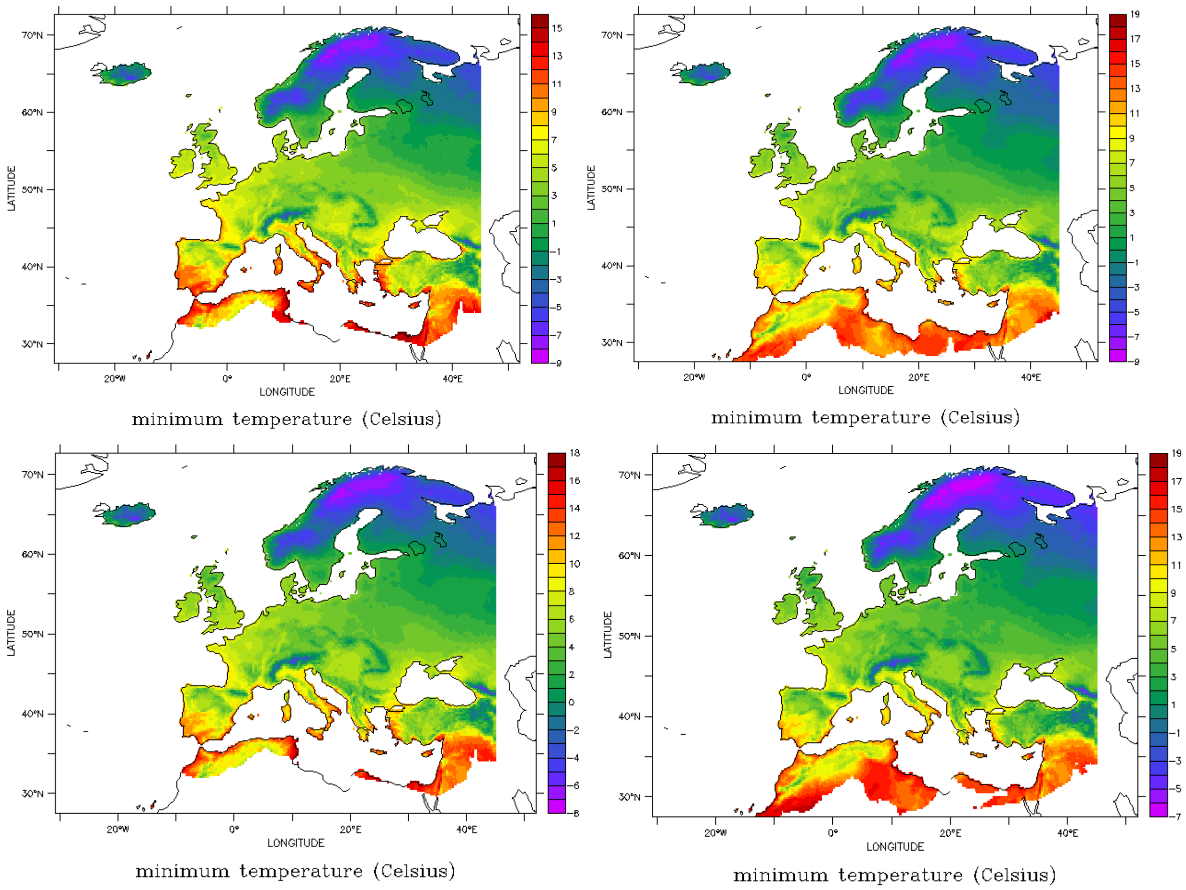

Fig.3.1 Comparison between averaged daily minimum temperature over the period 1951/01/01 to 1980/12/31 (upper row) and 1981/01/01-2010/12/31 (bottom row) for the gridded data set based on the homogenized data (left column) and non-homogenized data (right column).Plots based on the 100-member ensemble mean.

In the model used, spatial autocorrelation is explicitly accounted for. As discussed by Hefley et al. (2016) this captures first-order spatial autocorrelation of the variable. Second-order functions capture autocorrelation in the covariance of the distribution and include the geoadditive models of Kammann and Wand (2003), whereby spatial autocorrelation is modelled using a pre-defined covariance function. The low-rank approximations used in the splines used here relies on the selection of a basis dimension (k) prior to the fitting of the model. This is the key to the efficiency of the method (Wood, 2006). In the case of the interpolation of the data in E-OBS both of these methods produced very similar results, when the basis dimension (k), in the case of the first-order model, was set to a relatively large value relative to (n). In the case of temperature data the EDF is often of the order of (n/100) (Jones et al., 1997). In this study various k-values were tested for both the climatology and daily interpolations. For temperature k= 500 was used for the climatological interpolations.

Cornes et al. (2018) have also tested the potential for improved model fitting through the incorporation of additional co-variates to the climatological models. In the case of the temperature variables they have included distance from the nearest coast and the topographic position index (TPI). Distance from the coast was calculated using a relatively coarse 1:110m resolution coastline so as to provide a measure to the nearest large body of water rather than minor coastal features (Daly et al., 2008). TPI is calculated as the difference between a cell and the mean of the nearest eight cells. It is intended to provide a measure of small scale features such as frost hollows.

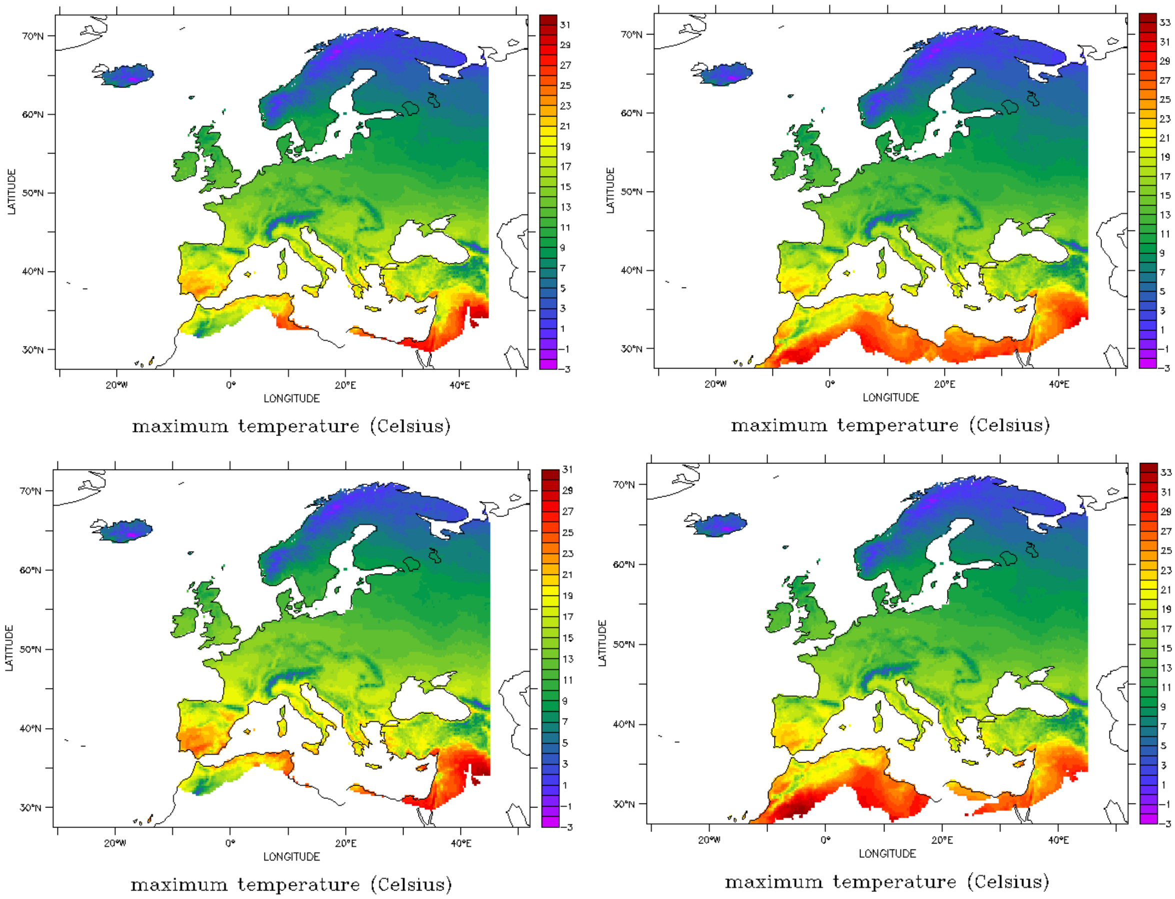

Fig.3.2 Comparison between averaged daily maximum temperature over the period 1951/01/01 to 1980/12/31 (upper row) and 1981/01/01-2010/12/31 (bottom row) for the gridded data set based on the homogenized data (left column) and non-homogenized data (right column). Plots based on the 100-member ensemble mean.

These topographic indices, excluding distance from the coast, were calculated using the 30 arc-second (approx. 1km) GMTED2010 Digital Elevation Model (DEM). For model fitting the grid cell nearest to the stations were used. GMTED2010 supersedes the GTOPO30 DEM used in the original version of E-OBS and is the preferred DEM for continental-scale analyses (Danielson and Gesch, 2011). The altitude values used for interpolation were also taken from GMTED2010, although for model fitting the altitude values provided in the station metadata were used.

3.2 Quantifying uncertainty in E-OBS

In the initial construction of the E-OBS dataset the possibility of generating an ensemble of grids for each day was suggested, with the spread across the ensemble providing a measure of uncertainty in the gridded values (Haylock et al. 2008). Due to the significant computational burden that this procedure entails that approach was rejected in favour of a single uncertainty value for each day, which was derived from a combination of the monthly-climatological uncertainty estimates (following the method described by Hutchinson 1998a, 1998b) and daily kriging uncertainty, derived using the approach described by Yamamoto (2000).

In this version of E-OBS we generate an ensemble of 100 equally probable grid realizations of the gridded daily fields by drawing simulations from the posterior distribution of the fitted model.

Fig. 3.3. gives a flavour of the uncertainty information by showing the mean squared error of this ensemble for daily minimum temperature, averaged over the 1981-2010 period. The mean squared error is calculated using the ensemble mean as 'observation' and relates to the spread in the ensemble, where high values relate to a large spread in the ensemble (and thus to a high uncertainty). Values are calculated using the homogenized dataset for daily minimum temperature. The pattern of the uncertainty information, especially when averaged over such long periods, reflects the density of the station network used in the gridding, with the lowest spread among the ensemble members in areas like Germany, the Czech Republic, Slovenia, south Sweden and south Finland which have high-density networks. In areas where the topography is complex, like over Norway, northern Scandinavia or the Alpine region, the uncertainty is relatively high, despite the high-density networks which are available.

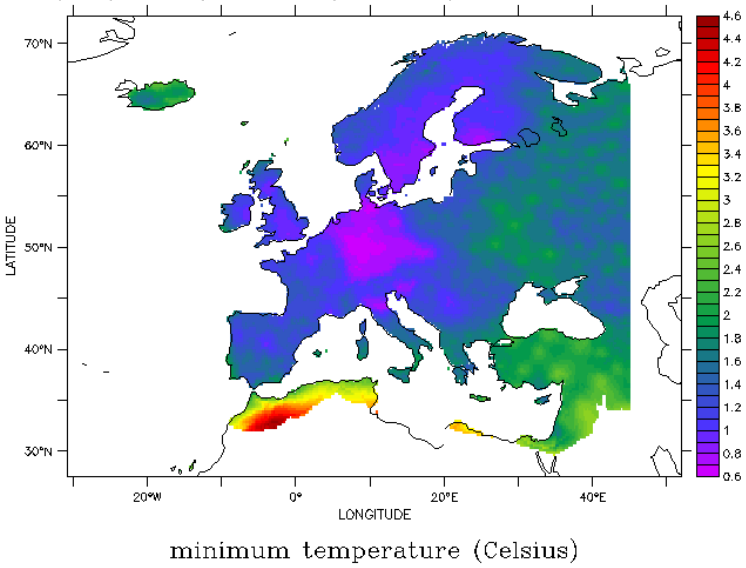

Fig.3.3 Mean standard error of the ensemble for daily minimum temperature, averaged over the period 1981/01/01-2010/12/31for the gridded data set based on the homogenized data Plots based on the 100-member ensemble with respect to the ensemble mean.

3.3 Homogenization of data

The station data for daily maximum and minimum temperature are homogenized using the quantile matching approach (Squintu et al., 2018) and are the result of a blending process between homogenized original series form nearby stations followed by a quantile-based homogenization approach for blended series meant to remove the breaks introduced with the blending. A EUSTACE report on the homogenization is available as Deliverable 1.4.

A first version of a pan-European homogenized station dataset has been made available to EUSTACE. Further research has brought to light that the homogenization process for daily maximum and minimum temperature data can be further refined; new breakpoints in the temperature series can be found, which indicate inhomogeneities in the data, if a second iteration of the homogenization procedure is applied. This further improvement is possible because an initial homogenization provides a larger set of reference series against which break points can be detected. Because a large set of references need to be used in this approach, a larger share of homogeneous references makes the signal-to-noise ratio to detect breaks points higher.

Homogenized data from this second iteration have been used as the basis for the current deliverable.

3.4 Data dissemination

Once the methods used have been published (see below), these data will become publicly available.

3.5 Planned publications

The process to produce pan-European homogenized temperature data and the changes in terms of the quality in our ability to monitor climate variability and change in Europe, will be documented by Squintu et al. (2018). Documentation on the newly developed gridding algorithm for E-OBS can be found is provided by Cornes et al. (2018). Both publications are currently being prepared.

A separate publication to document the calculation of the gridded data sets based on homogenized temperature data is not foreseen.

3.6 References

Cornes, R. et al. 2018. "An Ensemble Version of the E-OBS dataset" (in preparation)

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol., 28 (15), 2031–2064, doi: 10.1002/joc.1688, URL http://doi.wiley.com/10.1002/joc.1688, joc.1492.

Danielson, J. J., and D. Gesch, 2011: Global multi-resolution terrain elevation data 2010 (GMTED2010). Tech. rep., U.S. Geological Survey Open-File Report 2011-1073, 26p

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 (D20), D20 119, doi:10.1029/2008JD010201, URL http://doi.wiley.com/10.1029/2008JD010201.

Hefley, T. J., and Coauthors, 2017: The basis function approach for modeling autocorrelation in ecological data. Ecology, 98 (3), 632–646, doi:10.1002/ecy.1674, URL https://arxiv.org/pdf/1606.05658.pdfhttp://doi.wiley.com/10.1002/ecy.1674, Xiv:1606.05658v1

Hutchinson, M. F., 1998a: Interpolation of Rainfall Data with Thin Plate Smoothing Splines Part I: Two Dimensional Smoothing of Data with Short Range Correlation. J. Geogr. Inf. Decis. Anal., 2, 168–185.

Hutchinson, M. F., 1998b: Interpolation of rainfall data with thin plate smoothing splines - Part II: Analyis of topographic dependence. J. Geogr. Inf. Decis. Anal., 2 (2), 152–167.

Jones, P. D., T. J. Osborn, and K. R. Briffa, 1997: Estimating sampling errors in large-scale temperature averages. J. Clim., 10 (10), 2548–2568, URL https://ueaeprints.uea.ac.uk/33947/.

Kammann, E. E., andM. P.Wand, 2003: Geoadditive models. J. R. Stat. Soc. Ser. C (Applied Stat.,103 52 (1), 1–18, doi:10.1111/1467-9876.00385, URL http://doi.wiley.com/10.1111/1467-9876.

Squintu, A.A. et al. 2018., "Homogenization of ECA&D temperature data-set" (in preparation)

Wood, S. N., 2003: Thin plate regression splines. J. R. Stat. Soc. Ser. B (Statistical Methodol.,110 65 (1), 95–114, doi:10.1111/1467-9868.00374, URL http://dx.doi.org/10.1111/1467-9868.

Wood, S. N., 2006: Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.

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

4. Cross Project Links

The links with the FP7-UERRA project (http://www.uerra.eu/) are strong and have been essential is producing this deliverable.