# 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 D1.4**

***Report on Homogenised daily Land Surface Air Temperature***

| | | |
|---|---|---|
| Deliverable Title | Homogenised daily Land Surface Air Temperature, with report | |
| Brief Description | *The report briefly describes the homogenisation of ECA&D station-based surface air temperature data.* | |
| WP number | | 1 |
| Lead Beneficiary | *University of Bern* | |
| Contributors | *Yuri Brugnara, University of Bern*<br>*Antonello Squintu, KNMI*<br>*Gerard van der Schrier, KNMI* | |
| | | |
| Creation Date | 03/11/16 | |
| Version Number | 1.0 | |
| Version Date | 03/11/16 | |
| Deliverable Due Date | M16 | |
| Actual Delivery Date | 07/11/16 | |
| 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** |
|---|---|---|---|
| 1 | 03/11/16 | | First full draft, submitted to Science Coordinator for comment |
| 2 | 08/12/16 | | Full draft after comments |
| | | | |
| | | | |
| | | | |

## Table of Contents

# 1. Executive Summary

This document describes the procedure to produce daily homogenised land surface air temperatures (daily maximum and minimum temperature) based on the pan-European ECA&D dataset, as well as some results. Within EUSTACE, the existing ECA&D procedure to construct long and continuous temperature series, by filling-in gaps and extending records by combining original and continuation series after a station relocation, is modified to make this 'blending' step homogeneous as well. This is described in this report.

The homogenised data will serve as the basis for further work, such as the production of a homogeneous pan-European gridded dataset for land surface air temperature.

# 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 Constructing homogeneous series by blending series from two separate sources.

Long temperature records in Europe (and elsewhere) usually suffer from station relocations. There are many reasons why meteorological stations need to relocate; expanding cities (making the measurement site unsuitable for a meteorological station satisfying the WMO regulations) is only one of these reasons. A common reason for stations to be relocated all over Europe in the 1950s was the construction of airstrips or airports (military and civilian). These airports required meteorological variables to be measured at the site of the airport, which motivated the move of meteorological stations from nearby cities or villages to these airports. In some cases, the change was very dramatic. One example is the move of the meteorological station in the city of Groningen (the Netherlands) to the nearby airport Eelde in the countryside. Although the series from Eelde are considered to be a continuation of the Groningen data, the two datasets have a vastly different character (fig. 1).

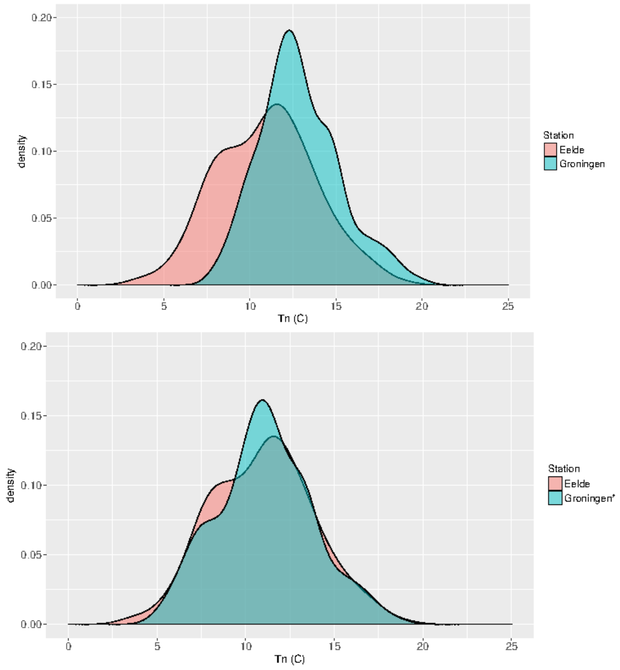



*Figure 1: Distribution for daily minimum temperatures observed at the meteorological stations Groningen (red) and Eelde (blue) for the period in which these stations overlap (top panel). The bottom panel shows the same data, after adjustment of the (older) Groningen data. Courtesy of Theo Brandsma.*

In the ECA&D dataset, the data from these two stations are stored as two separate series and in a 'blending' step, these series are joined to provide temperature series which are as long and complete as possible (ECA&D Project Team, 2012). However, simply joining these series would result in a severe inhomogeneity at the transition. Here, routines are made which adjust the earlier of the two series to the characteristics of the more recent series, before joining these to make one continuous record.

The blending procedure chooses, for every station, a base series (B) as the latest ending series and a set of donating series $D_i$ from nearby stations. For every gap in the base series, this procedure takes data from the closest donating series (within 12.5km and 50m elevation difference). The donating data can be used to fill in long periods of missing data or single days (fig. 2).

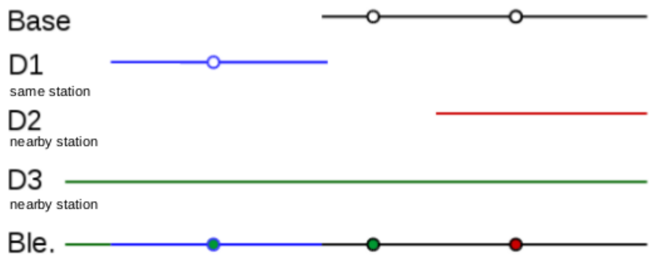



*Figure 2: Illustration of the blending procedure developed in EUSTACE. The recent data from the 'Base' series is joined by older data from the 'Donating' series (D1). Any gaps in these records are filled by other Donating series (D1 and D2). The resulting 'Blended' series consists of data from several sources.*

The method used to homogeneously construct the blended series is based on quantile matching (Trewin, 2012; Toreti et al. 2012). This is one of the two methods selected earlier in the EUSTACE project to trial for this purpose and takes into account that more extreme values in temperature need to be adjusted differently than more moderate temperatures.

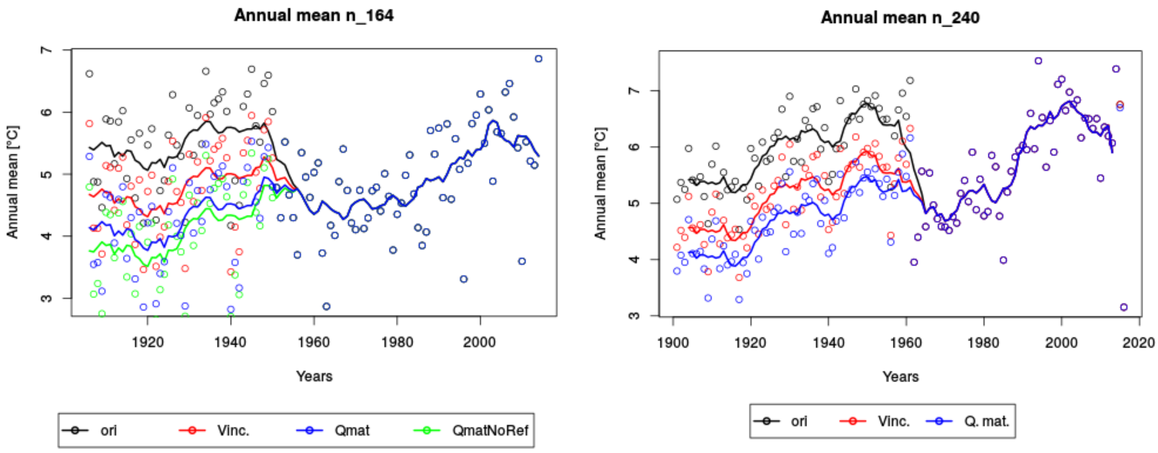



*Figure 3: Results of the modified blending step for Groningen/Eelde (left) and Geneva (right).The black lines show a simple joining of the two parts of the series (before and after relocation) without the homogenisation step. The blue lines show the result using the quantile matching technique, while the red line shows the results using the less sophisticated method of Vincent. The green line in the left panel represents a quantile matching homogenisation performed without reference series thanks to the presence of an overlapping period between the base and donating series (which is a situation that seldom occurs).*

Results of the modified blending procedure, for the Groningen/Eelde case discussed above and for station Geneva (Switzerland), are shown in fig. 3. The breaks in these series, due to the relocation of these stations to nearby airports, is clearly recognizable. A second method trialled for homogenisation is from Vincent et al. (2002), which is based on the comparison of mean monthly values before and after the break. This is a much simpler (and less capable) method. In an analysis using a small dataset, including benchmark series, the Vincent method was compared against the quantile matching and we observed that it was less efficient in producing homogenised series. Series from Geneva are blended as well, but here no overlapping period exists between the earlier and the more recent series (Fig. 3). The method developed here is capable of handling this situation as well.

### 3.2 Break detection in series measured at the same site

Temperature series measured at the same site can still be affected by important inhomogeneities (for example due to changes in the instrumentation), whose position in time is usually unknown because information on the station history is not readily available. Several statistical tests exist to detect these inhomogeneities, although only a few are completely automatic and can therefore be applied to large datasets. The automatic break detection implemented for EUSTACE is adapted from Kuglitsch et al. (2012) and can be divided into four phases:

1. Selection of the reference series

2. Break detection using three different methods with annual and semi-annual resolution

3. Extraction of the significant breakpoints

4. Refinement of the position of each breakpoint with monthly resolution

These Phases in the process are now detailed in turn, followed by an assessment of the effectiveness of the break detection method.

**Phase 1**

The selection of the reference series depends on some parameters that can be changed by the user. For application to ECA&D, a reference series was used only if it met each of the following criteria:

- contains at least *240* months of ‘valid’ data in common with the candidate
- its Pearson correlation coefficient with the candidate is at least *0.6* (calculated from first differences of yearly means)
- its geographical distance from the candidate is not larger than *1000* km

A month is considered ‘valid’ if no more than *5* days are missing. The first criterion implies that the candidate series must have at least *240* months of valid data (if not, break detection is aborted). A year (or a season) cannot have missing months. A maximum number of *8* reference series is used. These are selected giving priority to the amount of simultaneous data with the candidate and subsequently to the correlation.

The search for reference series starts within a radius of 100 km and increases progressively until eight long reference series are found or the maximum allowed distance (1000 km) is reached. If the number of reference series meeting the criteria is less than three (including short ones, i.e. <80% of the candidate’s length, selected only if not enough long series are found), the break detection is considered not possible and the successive phases are skipped. However, station density in Europe is high enough to allow a sufficient number of reference series everywhere and this situation is seldom reached.

**Phase 2**

The break detection is performed separately on annual means, 'winter' means (ONDJFM) and 'summer' means (AMJJAS), combining 3 methods:

- CAUME (Caussinus and Mestre, 2004)
- RHtests (Wang et al., 2007, Wang, 2008)
- GAHMDI (Toreti et al., 2012)

Each method delivers to the next phase those breakpoints that are detected by comparison to at least 3 of the reference series (using a tolerance window of 1 yea. For example, if two reference series assign a breakpoint to 1969, while a third reference series assigns it to 1970, then 1969 is delivered to the next phase).

**Phase 3**

A breakpoint is considered significant only if it was delivered by at least two of the detection methods used (a tolerance of 2 years is applied).

**Phase 4**

A window of 7 years around each breakpoint (narrower if near the boundaries or near another breakpoint) is extracted from the monthly candidate and reference anomaly series. A monthly series of differences between the candidate and the weighted average of the references (weights are the squared correlations) is calculated for this window. Then a Student's t-test is performed on each possible pair of sub-periods in the difference series (i.e., the 7-year window is divided into two, shifting the boundary between the sub-periods along the 7-year window by one month each time). The pair of sub-periods that gives the highest probability of rejecting the null-hypothesis (that the means of the two subperiods are identical) is selected: the breakpoint is assigned to the last month of the subperiod representing the first (i.e., earlier) part of the window for that pair. If a breakpoint was found significant only in a certain season, then only the months belonging to that season are used. The size of the breakpoints is estimated by the same R function that performs the t-test using the same window.

Figure 4 shows an example for the series of Trieste.

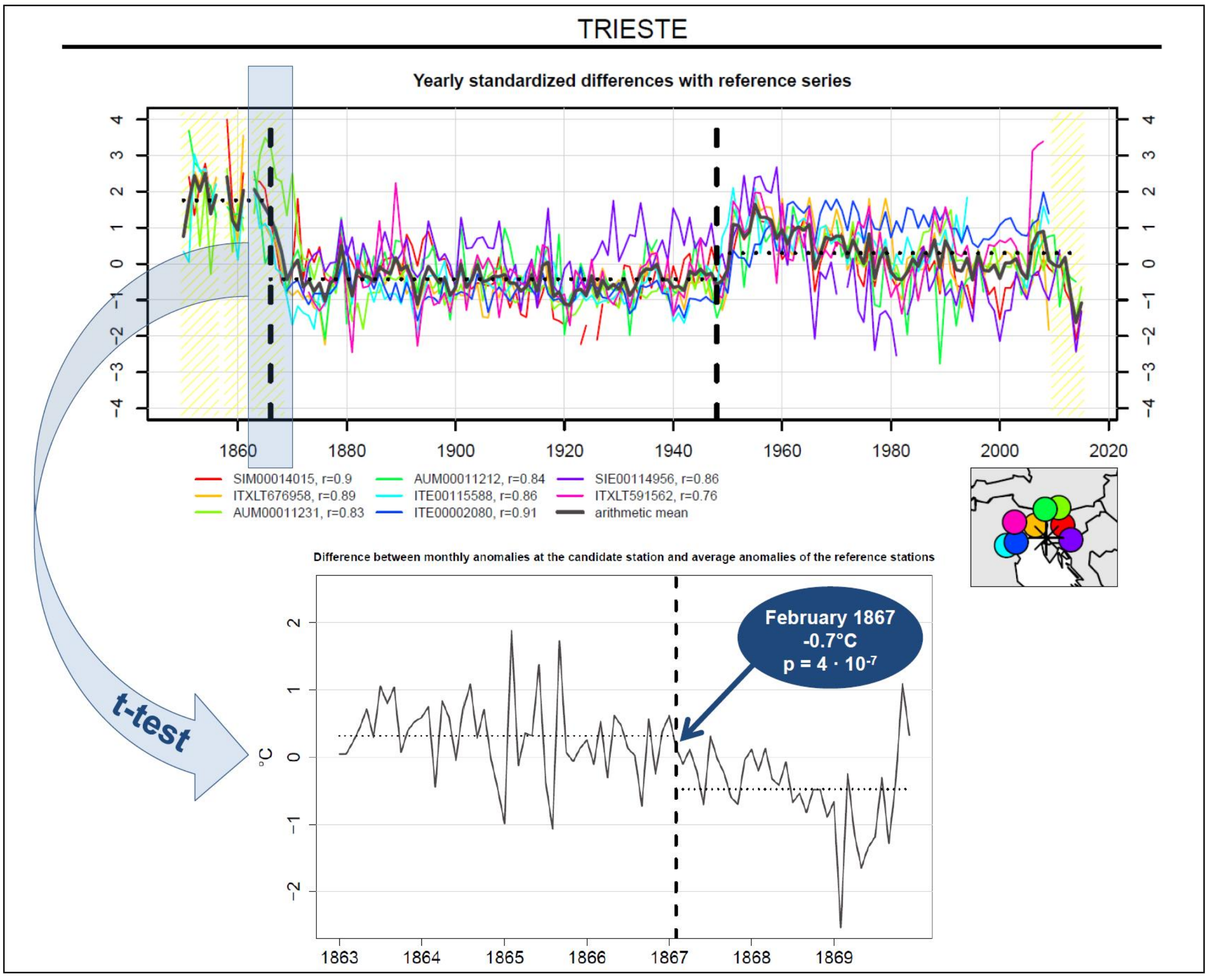



*Figure 4: Break-detection algorithm applied to the mean temperature series of Trieste.*

**Performance**

The probability of detection

$$POD = \frac{hits}{hits + misses}$$

where a hit is when a detected breakpoint is within two years of the real one. The mean absolute error (MAE) of the detection (for hits) was evaluated using the benchmark described in Venema et al. (2012). In total, 66 unique surrogate series from this benchmark data set have been analysed. For each surrogate series the break detection was performed six times with a different number of reference series (from 3 to 8; note that using more than 8 reference series is in general not recommended due to the rapid increase of the probability of false detections). Figure 5 shows the relationships of the POD and mean absolute error to the score of the break detection (defined as the sum of the correlation coefficients of the reference series: a high score means many well-correlated reference series).

As expected, the POD increases with the score. The MAE also increases with the score: this can be explained by the fact that with a low score only the largest inhomogeneities are detected, for which the uncertainty of the breakpoint's position is lower. This is shown in Fig. 6, where the absolute error for each detected breakpoint in the benchmark data set is plotted as a function of the estimated absolute size of the breakpoints (a 'proxy' for the signal-to-noise ratio of the inhomogeneities). Large errors disappear when the size is larger than 2 K, and are rare for a size between 1 K and 2 K.

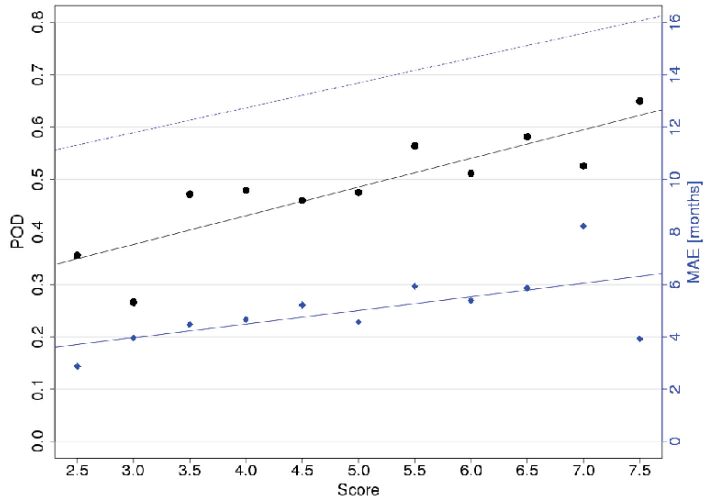



*Figure 5: Probability of detection (POD; black points) and mean absolute error (MAE; blue points) as a function of the detection score. The upper dashed blue line represents a regression of the average absolute errors obtained by selecting the month randomly within +/-1 year of the breakpoint detected with annual resolution.*

The error reduction of the t-test with respect to a random detection (upper blue dashed line in the plots) is on average ca. 65%. Overall, in more than 40% of the hits the date of the breakpoint (month and year) is detected with no error, while the error is less than 6 months in about 75% of the cases. It is important to say that the results can be significantly different when using other benchmarks. A multi-benchmark approach will be used in the scientific paper that will describe this method.

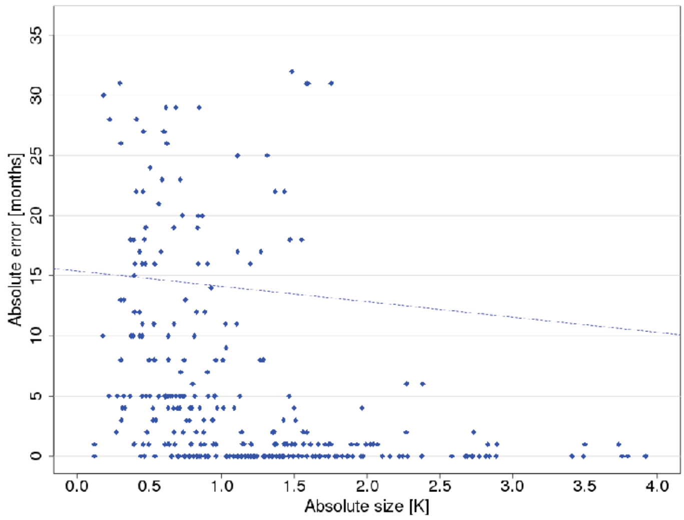



*Figure 6: Scatter plot of absolute errors in the position of the breakpoints versus the absolute size of the breakpoints. Dashed line as in Fig. 5, representing the average of random errors.*

### 3.3 Homogenisation of detected breaks

Based on the detected breaks for daily maximum and daily minimum temperature, the series are homogenised following the quantile matching technique. The approach is similar to what is described in Sect. 3.1 for the blending, with the difference that the location of the break is now provided by the statistical tool described in Sect. 3.2 rather than the known relocation date of the station.

Figure 7 shows an example of the effects of homogenisation for the station Altdorf (Switzerland). Several breaks are detected in this station and the homogenisation adjusts the original series to higher *and* lower values depending on the time window.

Figure 8 shows the effects of this final homogenisation step on the European data, in terms of the difference in trends in daily maximum temperature over the 1951-2010 period. Since breaks in the series strongly affect the estimate of the trends (confusing e.g. relocations or changes in the measurement equipment with climatic change), this metric can be used to assess the effects of the homogenisation. Figure 8 shows the value of trends prior to homogenisation and after homogenisation, and the difference between the two. It shows (looking at the size and the colour of the circles) that the trends based on homogenised data are more spatially homogenous and most series with negative trends (which are not consistent with trends of neighbour stations) have disappeared. The plot showing the difference in trends shows that there is no general sign of correction in trends due to the homogenisation.

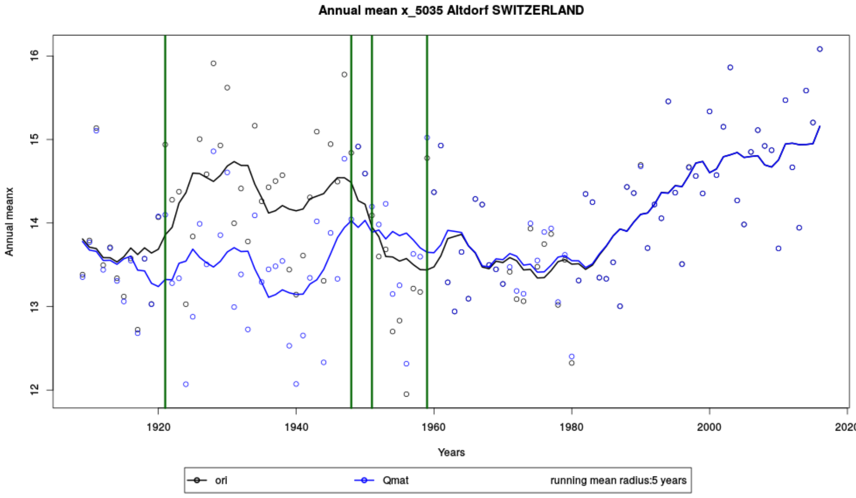



*Figure 7: Annual mean daily maximum temperature for Altdorf. The original data is in black, the homogenised data in blue and the vertical green lines denote the detected breaks in the series.*

Homogenising daily minimum temperatures proved to be a little more challenging than daily maximum temperatures (probably because of the fact that daily minimum temperature can be much less spatially homogeneous than daily maximum temperature). After homogenisation some stations showed negative trends, despite their neighbours exhibiting positive trends. An analysis of why these stations failed to produce trends with similar values to those of surrounding stations revealed undetected outliers and undetected breaks that affected the adjustment process. Such issues (affecting only 5 series out of more than 2200) have been solved manually by removing the outliers and adding the positions of these clear breaks. Figure 9 shows the effects of the homogenisation step on the European data for daily minimum temperature, in terms of the difference in trends over the 1951-2010 period. Here, as for maximum temperature homogenization, it's possible to observe an improved spatial homogeneity of the trends in the series and the disappearance of almost all the negative trends. The adjustments shown do not have a uniform sign or amplitude, similar to what was seen in the adjustment of the daily maximum temperature; this underlines that our method is not biased in any one direction.

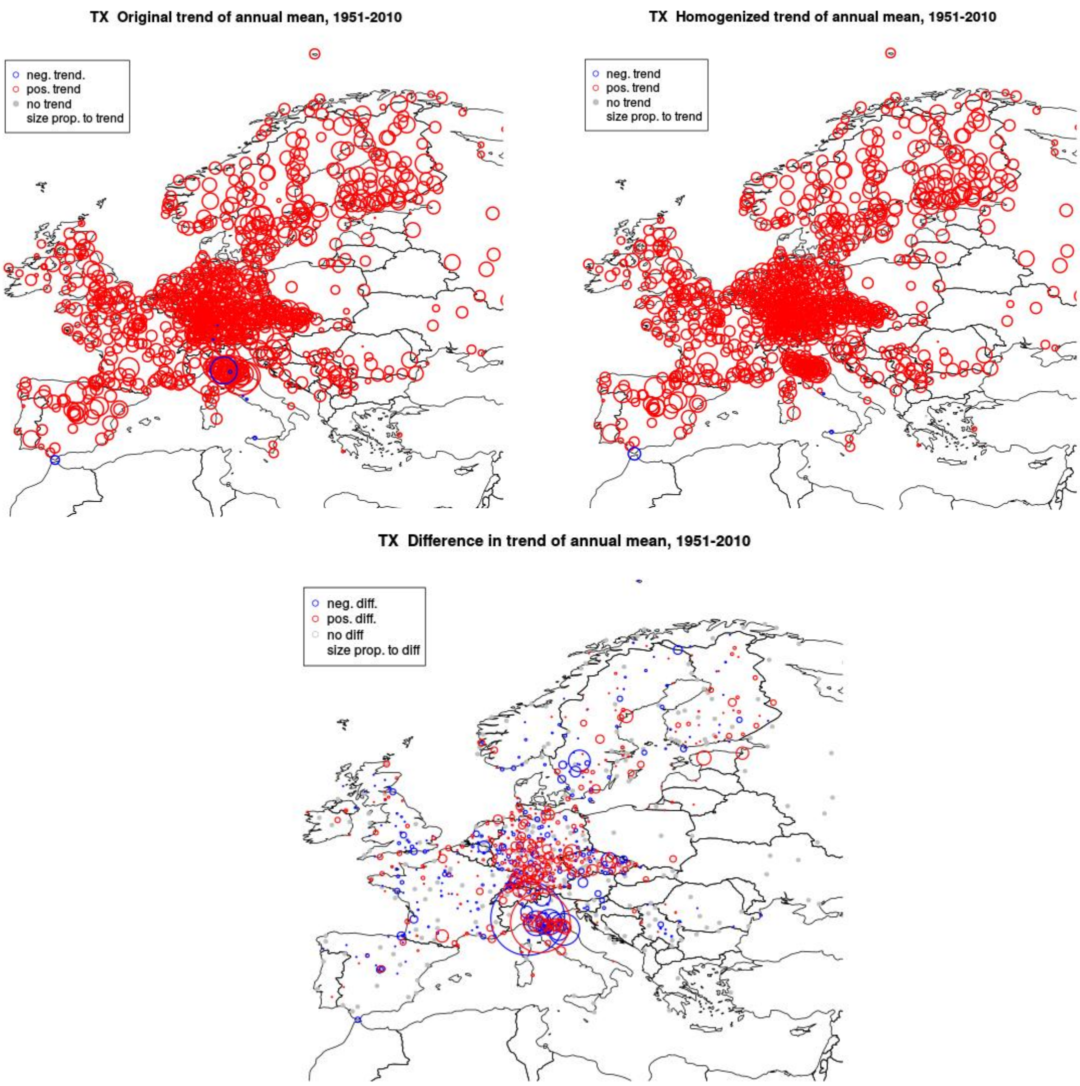
## 3.4 Summary

The daily minimum and daily maximum surface air temperatures in the ECA&D dataset have been homogenised. This is done in two steps. One is by modifying the existing ECA&D blending procedure, which aims to make series as long and as complete as possible, so that it can join, in a homogeneous way, temperature series of two nearby stations. This is relevant since virtually all long temperature series in Europe have seen one (or more) relocations. The second step is by first determining, using statistical tools, breaks in the series followed by a homogenisation step. The methods used to homogenise the data are the two methods identified earlier in EUSTACE.

The homogenised station data will become available through the ECA&D webportal as part of a comprehensive project to further improve the quality of the station data in ECA&D.

Availability of data is subject to the data policy restrictions of the National Meteorological Services who own the data.

The homogenised data will form the basis for the construction of a gridded European data set of daily maximum and minimum temperature and will contribute to the production and/or




validation of global data sets later in the EUSTACE project.

*Figure 8: Trends in annual means of daily maximum temperature over the 1951-2010 period, for the original (non-homogeneous) data (upper left), for the homogenised data (upper right) and the difference between the two (bottom). Red(blue) circles indicate positive(negative) trends in the top maps and positive(negative) difference in trends before and after homogenisation in the bottom map.*

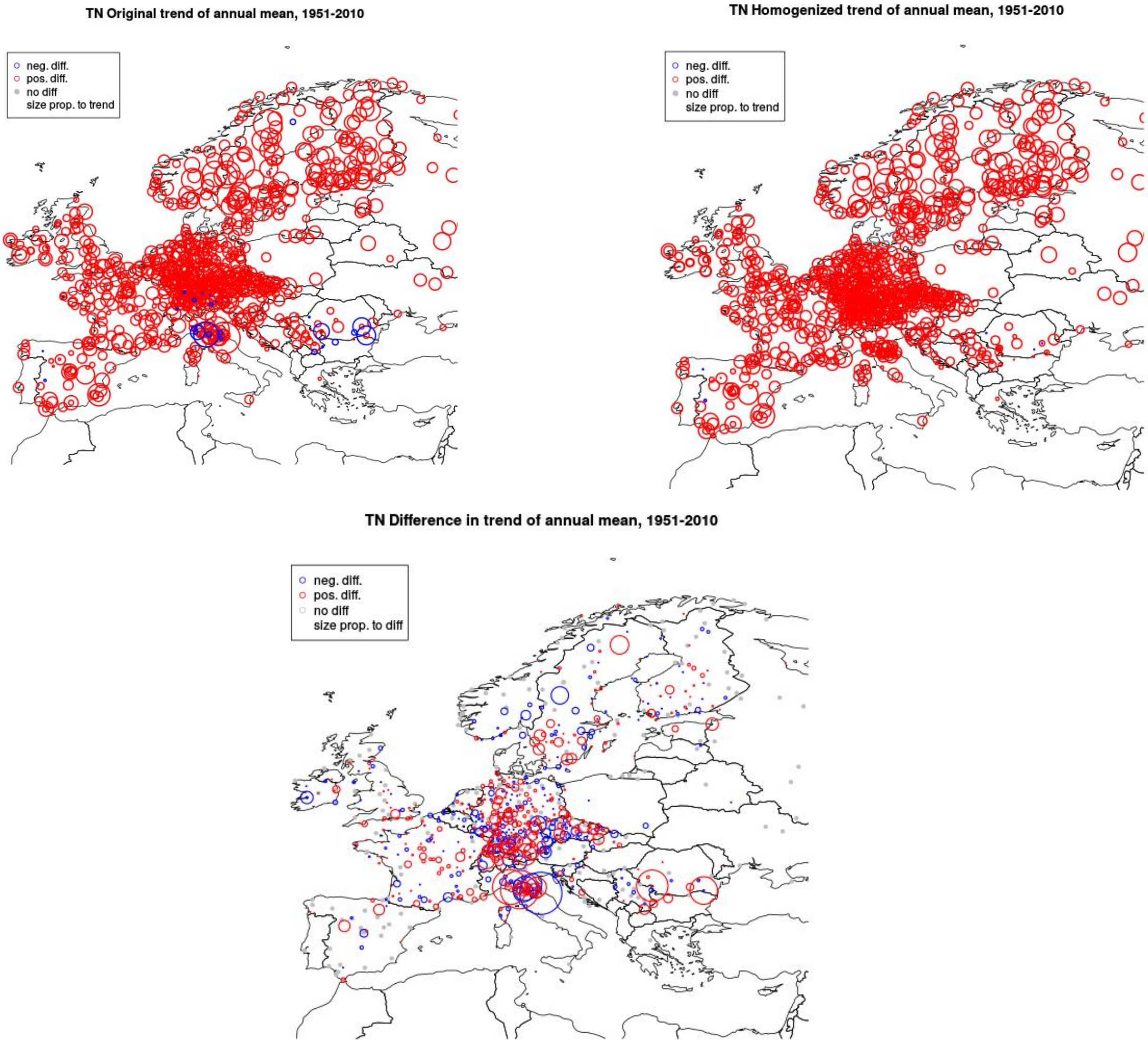



*Figure 9: Trends in annual means of daily minimum temperature over the 1951-2010 period, for the original (non-homogeneous) data (upper left), for the homogenised data (upper right) and the difference between the two (bottom).*

## 4. References

H Caussinus and O Mestre. 2004. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society*, 53(3):405-425.

ECA&D Project Team. 2012.European Climate Assessment & Dataset Algorithm Theoretical Basis Document (ATBD), Royal Netherlands Meteorological Institute KNMI,De Bilt, NL., version 10.5

FG Kuglitsch, R Auchmann, R Bleisch, S Brönnimann, O Martius, and M Stewart. 2012. Break detection of annual swiss temperature series. *Journal of Geophysical Research*, 117(D13).

A Toreti, F G Kuglitsch, E Xoplaki, and J Luterbacher.2012. A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology*, 51(2):317-326.

VKC Venema, O Mestre, E Aguilar, I Auer, JA Guijarro, P Domonkos, G Vertacnik, T Szentimrey, P Stepanek, P Zahradnicek, et al. 2012. Benchmarking homogenisation algorithms for monthly data. *Climate of the Past*, 8(1):89-115.

LA Vincent, X Zhang, BR Bonsal and WD Hogg. 2002. Homogenisation of daily temperatures over Canada. J. Climate 15:1322-1334.

X L Wang. 2008. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or f test. Journal of Applied Meteorology and Climatology, 47(9):2423-2444.

X L Wang, Q H Wen, and Y Wu. 2007. Penalized maximal t test for detecting undocumented mean change in climate data series. Journal of Applied Meteorology and Climatology, 46(6):916-931.