



Homogenization of daily maximum temperature series in the Mediterranean

Franz G. Kuglitsch,^{1,2} Andrea Toreti,^{1,2,3} Elena Xoplaki,^{1,2,4} Paul M. Della-Marta,⁵ Jürg Luterbacher,⁶ and Heinz Wanner^{1,2}

Received 9 December 2008; revised 29 May 2009; accepted 10 June 2009; published 15 August 2009.

[1] Homogenization of atmospheric variables to detect and attribute past and present climate trends and to predict scenarios of future meteorological extreme events is a crucial issue for the reliability of analysis results. Here we present a quality control and new homogenization method (PENHOM) based on a penalized log likelihood procedure and a nonlinear model applied to 174 daily summer maximum temperature series in the Greater Mediterranean Region covering the last 50–100 years. The break detection method does not rely on homogeneous reference stations and was chosen owing to the lack of metadata available. The correction procedure allows the higher-order moments of the candidate distribution to be corrected, which is important if the homogenized series are to be used to quantify temperature extremes. Both procedures require a set of highly correlated neighboring stations to correct climate series reliably. After carrying out the homogeneity procedure, 84% of all time series were found to contain at least one artificial breakpoint. Time series of the eastern Mediterranean (one breakpoint in 24 years on average) show significantly more breakpoints than do series of the Western Basin (one breakpoint in 36 years on average). The mean adjustment (standard error) of all daily summer maximum temperatures is $+0.03^{\circ}\text{C}$ ($\pm 0.38^{\circ}\text{C}$) for the western Mediterranean, $+0.16^{\circ}\text{C}$ ($\pm 0.52^{\circ}\text{C}$) for the central Mediterranean, and $+0.19^{\circ}\text{C}$ ($\pm 0.30^{\circ}\text{C}$) for the eastern Mediterranean, indicating a reduced increase in mean summer daytime temperature compared to that detected by analyzing raw data. The adjustments for higher-order moments were not uniform. Most significant mean changes due to homogenization were detected for both: the hottest ($+0.15^{\circ}\text{C} \pm 0.66^{\circ}\text{C}$) and coldest decile ($-0.83^{\circ}\text{C} \pm 1.28^{\circ}\text{C}$) compared to the raw data in the central Mediterranean. This study demonstrates that homogenization of daily temperature data is necessary before any analysis of temperature-related extreme events such as heat waves, cold spells, and their impacts on human health, agriculture, and ecosystems can be studied.

Citation: Kuglitsch, F. G., A. Toreti, E. Xoplaki, P. M. Della-Marta, J. Luterbacher, and H. Wanner (2009), Homogenization of daily maximum temperature series in the Mediterranean, *J. Geophys. Res.*, 114, D15108, doi:10.1029/2008JD011606.

1. Introduction

[2] The current climate change debate requires a detailed analysis of long instrumental series and a quantification of natural and anthropogenic impacts in climate data [e.g.,

Alexandersson and Moberg, 1997; Vincent et al., 2002; Caussinus and Mestre, 2004; Intergovernmental Panel on Climate Change (IPCC), 2007]. Instrumental data sets are of major interest since they are the basis for analyzing past and present climate changes and a fundamental aspect of modeling future climate scenarios.

[3] However, many studies have shown that most climate series are characterized by artificial breakpoints that have to be corrected [e.g., Peterson et al., 1998; Szentimrey, 1999; González-Rouco et al., 2001; Cocheo and Camuffo, 2002; Maugeri et al., 2002; Brandsma and Können, 2006; Della-Marta and Wanner, 2006]. Usually, these breakpoints are introduced by changes in measurement conditions, relocation of weather stations, land-use changes, new instrumentation or changes in observational hours among others [Peterson et al., 1998; Aguilar et al., 2003]. In general, these changes are not documented by metadata or the metadata are difficult to access in archives. Statistical homogenization procedures have been developed for

¹Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

²Climatology and Meteorology Group, Institute of Geography, University of Bern, Bern, Switzerland.

³Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, Italy.

⁴Energy, Environment, and Water Research Center, The Cyprus Institute, Nicosia, Cyprus.

⁵Federal Office for Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland.

⁶Climatology, Climate Dynamics, and Climate Change Section, Department of Geography, Justus-Liebig University of Giessen, Giessen, Germany.

detecting and adjusting such inhomogeneities on various temporal scales. Most approaches have been developed for homogenizing monthly and annual mean series [e.g., *Caussinus and Lyazrhi*, 1997; *Vincent*, 1998; *Szentimrey*, 1999; *Wang*, 2003; *Caussinus and Mestre*, 2004] and applied to regional studies in Europe [e.g., *Caussinus and Mestre*, 2004; *Auer et al.*, 2005; *Begert et al.*, 2005; *Brunetti et al.*, 2006; *Auer et al.*, 2007; *Toreti and Desiato*, 2008], North America [e.g., *Zhang et al.*, 2000] and Australia [e.g., *Della-Marta et al.*, 2004]. The extensive use of monthly data can be explained mainly by their accessibility in digital form [e.g., *Aguilar et al.*, 2003; *Della-Marta et al.*, 2004].

[4] While the methods used to homogenize monthly and annual data are well established, only a limited number of publications address the challenge of daily temperature data correction and of other spatially highly correlated climate parameters [e.g., *Demarée et al.*, 2002; *Maugeri et al.*, 2002; *Vincent et al.*, 2002; *Brandsma and Können*, 2006; *Brunet et al.*, 2006]. In general these techniques use highly correlated neighboring station series to correct candidate station values and are therefore applicable only in areas with high station density. However, all of these methods are only able to homogenize inhomogeneities which affect the first-order moment of a distribution function, the mean, and do not correct higher-order moments such as the variance and skewness properties of a distribution function. The only reliable method to homogenize all moments of daily extremes and averages was proposed by *Trewin and Trevitt* [1996] and improved by *Della-Marta and Wanner* [2006] to use any reference station rather than relying on simultaneous measurement data. In an application of their method, *Della-Marta et al.* [2007a] homogenized 25 long-term daily maximum temperature series over Europe. All above mentioned studies conclude that the required adjustments often have a similar magnitude as the climate signal, long-term climate variations, trends or cycles. Hence, climate data homogenization is necessary for all climate studies based on instrumental data.

[5] Climate change detection has moved away from documentation of changes in monthly and annual averages to be focused now on the analysis of extreme events (e.g., heat waves, droughts, floods, and storms) [*Frich et al.*, 2002; *Stott et al.*, 2004; *Alexander et al.*, 2006; *Hegerl et al.*, 2006; *Della-Marta et al.*, 2007b; *Founda and Giannakopoulos*, 2009]. Therefore a continuous development of daily data homogenization methods and its large-scale application is required. To meet these demands, this study (1) proposes an improved homogenization method based on well-established models [*Caussinus and Mestre*, 2004; *Della-Marta and Wanner*, 2006] for the detection and the correction of artificial breakpoints in long daily climate series and (2) applies the new method to 174 daily maximum temperature series in the Greater Mediterranean Region (GMR).

[6] The Mediterranean area is regarded as a “Hot Spot” of climate change [*Giorgi*, 2006; *Diffenbaugh et al.*, 2007] that suffers from more extreme temperature events, an increase of summer heat wave frequency and duration in the western parts [e.g., *Della-Marta et al.*, 2007b] and rising summer temperature variability [*Xoplaki et al.*, 2003, 2006; *Jones et al.*, 2008] combined with a decrease in water resources having an unprecedented impact on ecology, economy and society [*IPCC*, 2007].

[7] The strength of this homogenization approach is in its ability (1) to detect an unknown number of nondocumented breakpoints in a candidate station by using a dynamic programming algorithm based on mean annual maximum temperatures of its 10 highest correlated neighboring stations and (2) to correct daily extremes and averages using highly correlated and piecewise-homogeneous neighboring stations. Since the spatial correlation is highest for daily summer (JJAS) maximum temperature series a homogenization of these series is most reliable. The investigation of daily summer maximum temperature extremes, the appearance of heat waves and its stressful impacts on society and economy, agriculture and the environment is a major concern of future research activities and needs further attention [e.g., *Beniston et al.*, 2007; *Diffenbaugh et al.*, 2007; *Brown et al.*, 2008; *Fang et al.*, 2008; *Founda and Giannakopoulos*, 2009].

[8] Unfortunately, there are only a couple of studies dealing with temperature and precipitation extremes in parts of the GMR based on nonhomogenized data [e.g., *Alpert et al.*, 2002; *Kostopoulou and Jones*, 2005; *Moberg and Jones*, 2005; *Moberg et al.*, 2006]. *Brunet et al.* [2007] and *Della-Marta et al.* [2007b] were the first who analyzed homogenized summer daily temperature data in Western Europe including the Iberian Peninsula. *Della-Marta et al.* [2007b] found that over the period 1880–2005 the length of summer heat waves over Western Europe has doubled and the frequency of hot days has almost tripled and was accompanied by an increase in the variance of daily maximum temperature. Despite the very long and rich history in monitoring climatic parameters in most parts of the Mediterranean, no climate studies based on homogenized daily data have been performed for the detection of extreme events in the entire Mediterranean area.

[9] In order to answer vital questions regarding climate change, extreme events and their impacts in the Mediterranean, National Meteorological and Hydrological Services (NMHSs), the European Climate Assessment & Data set (ECA&D; eca.knmi.nl) [*Klein Tank et al.*, 2002], privately funded projects (e.g., IEDRO, <http://www.iedro.com>) and collaborative projects of the WMO and the European Union have put great effort in building up a homogenized climate database in the GMR. Ongoing international projects dealing with (1) climate data rescue, (2) homogenization issues, (3) the attribution of present climate trends, (4) the analysis of extreme events, and (5) their impacts on ecology, economy, and the societies of the Mediterranean environment are the WMO–Initiative on Mediterranean Climate Data Rescue (WMO-MEDARE initiative [*Brunet and Kuglitsch*, 2008], www.omm.urv.cat/MEDARE/index.html) and the integrated EU-IP CIRCE project on Climate Change and Impact Research: The Mediterranean Environment (www.circeproject.eu), which was initiated in 2007. A strong collaboration with these projects and the development of a new homogenization technique enabled us to build up a comprehensive quality controlled and homogenized daily maximum temperature (TX) database including 174 sites from 16 countries along the GMR covering the last decades.

2. Data and Methods

[10] The daily data used in this study consist of daily summer (JJAS) maximum temperature time series (TX)

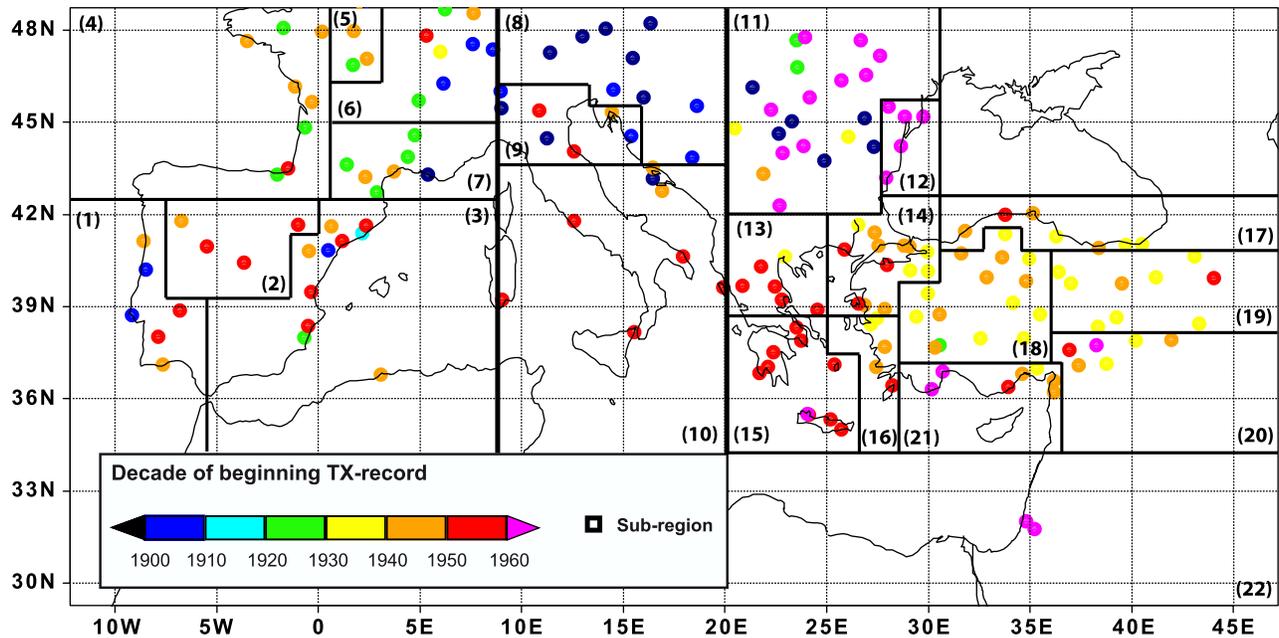


Figure 1. Location of all 174 stations used to homogenize daily maximum temperature series. The start period for each series is indicated by the colors. The subregions defined by the black boundaries (see also Table 1) aggregate station series that were highly correlated and homogenized by each other.

from 174 stations in 16 countries across the Greater Mediterranean Region. Many more series exist but were excluded owing to (1) too short measuring periods and/or (2) too many missing or unreliable data. Figure 1 shows the geographical distribution and start period of all 174 sites included. The data stem from the ECA&D [Klein Tank *et al.*, 2002], the WMO-MEDARE [Brunet and Kuglitsch, 2008], and several NMHSs and consider stations in Algeria, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, France, Greece, Israel, Italy, Portugal, Romania, Serbia, Slovenia, Spain, Switzerland, and Turkey.

[11] Ninety-five percent of all stations cover the period 1961–2005. Eighty well-distributed station records are available for the period 1930–2005; 30 stations cover more than 100 years (Figure 1). For 90% of the series, updates to 2006 or 2007 were available. Only a minority of stations end before 2005. The site network is fairly dense across the western, northern and eastern Mediterranean area. However, except for Alger there are no reliable data available from Northern African countries. For the correction of the TX summer series, we identify 22 subregions (Figure 1) consisting of clustered stations characterized by highly correlated summer maximum temperature series. A candidate temperature series was always corrected by a reference series of the same subregion. For break detection, also highly correlated reference series from outside a subregion were considered. These subregions are listed with their names, number of stations, and their key sites in Table 1. Key sites are representative stations for the subregion that were homogenized with higher reliability owing to higher correlated reference stations ($\rho > 0.90$ at daily scale), are usually longer than their reference series and do not show many missing or non valid values.

[12] The site density could be potentially improved in Italy, southern Spain, parts of the Balkan Peninsula and the

Middle East [e.g., Brunet *et al.*, 2007; Sensoy *et al.*, 2007]. Many series of these areas are available but were excluded after quality control owing to too many missing values and/or too short time series. These areas require further data rescue and digitization, data exchange and the application of a reliable homogenization method in order to produce more detailed and trustworthy regional climate studies. Significant efforts in data rescue and data digitization are made within the WMO-MEDARE-Initiative since 2007 [Brunet and Kuglitsch, 2008]. A central concern for future studies is the enlargement of the study area to Northern Africa and the Middle East [Sensoy *et al.*, 2007; Xoplaki, 2009]. Efforts to the both extend the maximum (TX) and minimum temperature (TN) database and apply daily homogenization methods should be pursued. However, an appropriate homogenization of daily TN (as also for nonsummer TX) is still not possible as currently not sufficient highly correlated station series are available.

2.1. Missing Values and Data Quality Control

[13] Analyzing daily data (e.g., the estimation of percentiles in a given time period) requires nearly complete and reliable time series. Therefore, when studying decadal variability and/or long-term trends of temperature it is essential to remove seasons with too many missing values and to ignore series with too many and/or clustered missing seasons during a certain period [Moberg and Jones, 2005]. For these reasons, and in order to include as many Mediterranean series as possible, we applied the criteria by Moberg and Jones [2005] in a slightly modified form for the selection of suitable stations: (1) A summer month (JJAS) is considered complete when there are ≤ 3 missing days. (2) A summer season is considered as available when all summer months are complete in respect to criterion 1. (3) A station series is considered as complete

Table 1. Subregions Used in the Analysis

Subregion	Number of Stations	Key Site (Length of TX Series)
<i>Western Mediterranean</i>		
South		
Atlantic	6	Lisbon, Portugal (1901–2007)
Continental	4	Salamanca, Spain (1950–2007)
Mediterranean	10	Valencia, Spain (1950–2007)
North		
Atlantic	8	San Sebastian, Spain (1929–2007)
Central	5	Deols Chateauroux, France (1921–2007)
Continental	8	Geneva, Switzerland (1901–2007)
Mediterranean	7	Montelimar, France (1925–2007)
Central Mediterranean		
Continental	9	Graz, Austria (1894–2004)
North	7	Verona, Italy (1951–2007)
South	7	Brindisi, Italy (1951–2007)
<i>Eastern Mediterranean</i>		
North		
Continental	20	Bucharest, Romania (1930–2005)
Black Sea West	5	Constanta, Romania (1961–2005)
East		
Northwestern Aegean	6	Larissa, Greece (1955–2007)
Northeastern Aegean and Marmara region	12	Bursa, Turkey (1931–2006)
Southwestern Aegean	9	Hellinikon, Greece (1955–2007)
Southeastern Aegean	5	Manisa, Turkey (1930–2006)
Turkey–Black Sea	7	Sinop (1943–2006)
Turkey–Western Anatolia	15	Kutahya (1930–2006)
Turkey–Eastern Anatolia	9	Erzurum (1930–2006)
Turkey–Southeastern Anatolia	6	Diyarbakir (1930–2006)
Turkey–Mediterranean	7	Antalya (1963–2006)
Israel	2	Jerusalem (1964–2004)

when no more than three consecutive summer seasons are missing.

[14] This results in the inclusion of approximately 20% more series than when only allowing two missing values per month. Beside the selection of time series in terms of missing values, detailed analyses of the daily data quality are an essential task before undertaking any homogeneity assessment and further data analysis. Therefore a complete data quality assessment on the full TX database recommended by *Aguilar et al.* [2003] was undertaken to identify suspicious values and outliers.

[15] The TX series were checked for consistency, tolerance, and temporal coherency. TX values exceeding (1) 50.0°C, (2) ± 4 standard deviations (σ) of the full length of the respective station series, and (3) a difference of 25°C between consecutive observations were replaced by missing values. Further, four or more equal consecutive values in a row were removed, and the number of days for each month was checked again for consistency. After applying these criteria to the whole data set we found the most common complete time period from 1962 and 1991 (for 174 stations) showing a maximum of 3 missing years. Therefore, this 30-year period rather than the climatological standard-normal period (1961–1990 or 1971–2000 [Scherrer et al., 2006]) has been used as the base period for the data homogenization.

2.2. Daily Data Homogenization

[16] A combined application of the penalized log likelihood procedure of *Caussinus and Mestre* [2004] and the nonlinear modeling method of *Della-Marta and Wanner* [2006] is used to detect an unknown number of breakpoints and for the correction of the time series respectively. This approach allows a reliable homogenization of long daily

maximum temperature measurements even when high-quality metadata that provide information on possible breakpoints (inhomogeneities) is limited or not available. However, both methods are based on highly correlated ($\rho > 0.8$) neighboring stations and usually allow a reliable homogenization only for regions with high site density and similar climate conditions. The minimum density depends on geographical features and generally has to be higher in continental areas than in coastal areas owing to typical decorrelation length scales. We call this new homogenization technique PENHOM. Within this study, PENHOM was applied to a Mediterranean daily summer maximum temperature database.

[17] In sections 2.2.1–2.2.3 we detail the break detection, break correction and the rebreak detection part of the PENHOM method. The method is illustrated by its application to the inhomogeneous daily maximum summer temperature series from Kutahya, Turkey. This temperature series is one of the longest (1930–2006) in western Anatolia. It shows only one missing day in 77 years and is highly correlated to many reference series in western Turkey.

2.2.1. Break Detection

[18] The Caussinus-Mestre method [*Caussinus and Mestre*, 2004] applied in this study accounts for the detection of an unknown number of multiple breakpoints in mean annual TX difference series on the basis of the following equation:

$$C_K(Y) = \ln \left\{ 1 - \frac{\sum_{j=1}^{k+1} n_j (\bar{Y}_j - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right\} + \frac{2k}{n-1} \ln(n), \quad (1)$$

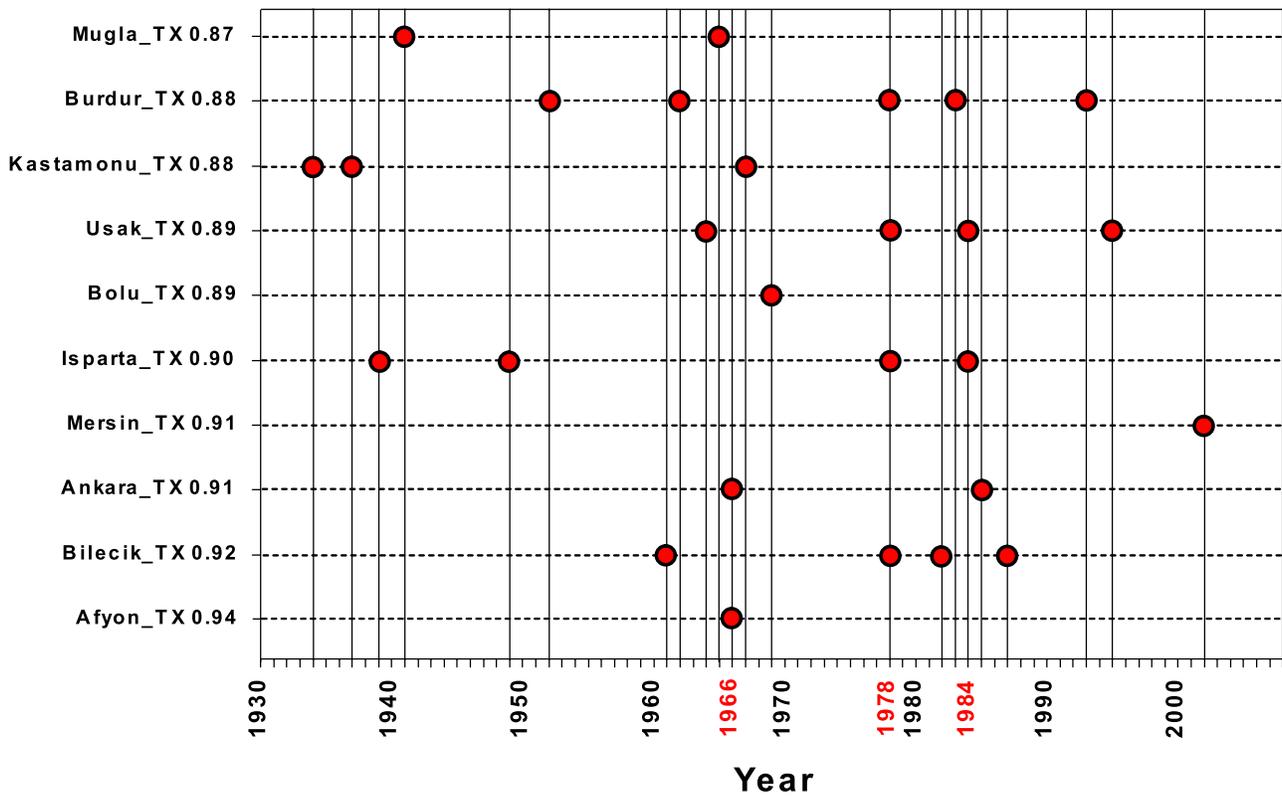


Figure 2. Synthesis of the detected breakpoints in the Kutahya series (raw data). The stations are ordered from bottom to top with respect to decreasing values of correlation. Hence the reliability of the comparisons increases from top to bottom. The red dots show the position of the detected breakpoints in the difference series for Kutahya (Turkey) versus the other stations in Turkey. The red-highlighted years indicate the chosen breakpoint years.

where n is the number of observations, \bar{y} denotes the mean value over the entire period (i.e., using all the n observations), k denotes the number of possible breakpoints located at $\{t_1, \dots, t_k\}$, n_j is the number of observations within the homogeneous subperiod $[t_{j-1}, t_j]$, and \bar{Y}_j is the mean of the series in that subperiod. The selected set K^* (i.e., our detected breakpoints) should satisfy: $\arg \min_K \{C_K(Y)\}$.

[19] These assumptions allow a reliable detection of breakpoints in long climate series even if they are not accompanied with metadata that provide information on potential breakpoints. It is assumed that between two breakpoints, a time series is homogeneous. Consequently, each time series consists of a number (≥ 1) of several Homogeneous Subperiods (HSPs). HSP1 is always the latest subperiod of a time series.

[20] Each single mean annual TX series (candidate station) is compared to its ten highest correlated neighboring series (reference stations) by producing difference series between the candidate and its references. Note that a reference station used in this step could be a station outside of the subregions shown in Figure 1. These difference series are tested for discontinuities according to equation (1). Years of breakpoints are assumed to be valid if three or more breakpoints are detected throughout the set of comparisons of a candidate station with its neighbors within two consecutive years. If there happens to be a clustering of break-detected points in several years close together then the year which contains the highest number of breakpoints in this cluster is chosen

as the breakpoint year. The break detection methodology is based on mean annual TX values since artificial shifts are more reliably detected in annual means than in monthly or daily time series [Easterling and Peterson, 1995; Alexandersson and Moberg, 1997; Vincent, 1998; Caussinus and Mestre, 2004]. The use of standardized difference series based on seasonal values usually gives comparable results. However, we avoided making any a priori assumptions on the reliability of HOM procedure (see the work of Della-Marta and Wanner [2006] for details). Indeed, where station densities are high, a reliable correction was also possible for May and/or October. However, we want to use a common and reliable methodology for the whole area, therefore we restricted our analysis to the summer season (JJAS). The correlation coefficient is calculated for the 30-year period 1962–1991 (for details, see section 2.1).

[21] The breakpoints detected in the difference series between Kutahya and its 10 highest correlated neighboring stations are summarized in Figure 2. The alignment of the breakpoints indicates at least three breakpoints in 1966, 1978, and 1984. We assumed these breakpoints are trustworthy even when a confirmation of the breakpoints detected with supplementary metadata was not possible.

2.2.2. Break Correction

[22] After break detection, the Higher-Order Moments (HOM) method of Della-Marta and Wanner [2006] was applied to adjust the breakpoints in the daily maximum

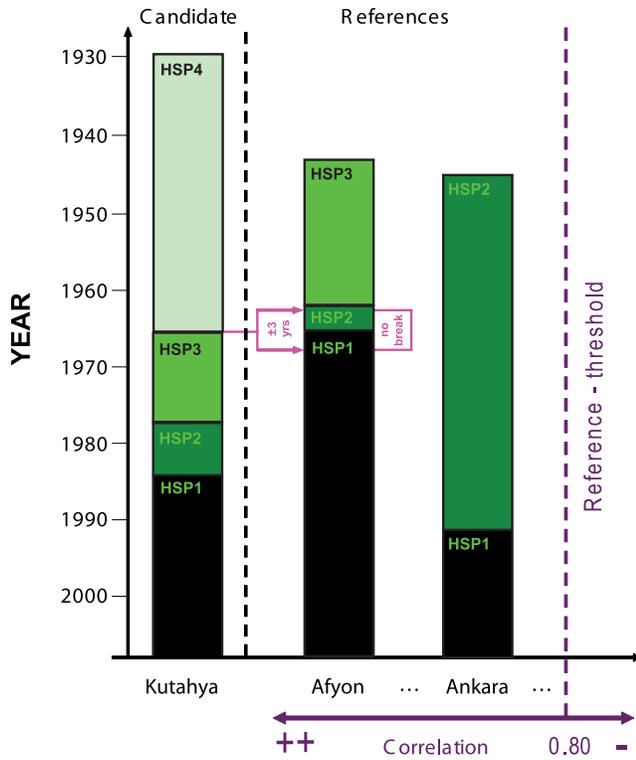


Figure 3. Schematic overview of the candidate (Kutahya, Turkey), its highest correlated reference series (Afyon, Turkey), and the reference series used (Ankara, Turkey) for correction due to the distribution of detected breakpoints. The periods between the detected breakpoints (HSP1-n) are assumed to be homogeneous. Only the highest correlated reference series (>0.8) are qualified for a reliable homogenization procedure.

well as 3 years before and after to get long enough overlapping periods [Della-Marta and Wanner, 2006]. The HOM method is applied to each month separately, therefore the minimum of 3 years of reference data either side of the breakpoint equals a minimum of approximately 90 daily values to be used in each nonlinear model. In some cases it is possible that a reliable homogenization using one reference station is only possible for some HSPs but not for the whole candidate time series owing to short overlapping periods. In this case, one should use a combination of more than one reference series [Della-Marta and Wanner, 2006] which was applied to a couple of stations in Turkey and Spain (see Figures 10a–10c). To correct the Kutahya series we used the maximum temperatures from Ankara as a reference. This was possible since (1) the summer TX of Ankara is highly correlated to Kutahya ($\rho = 0.95$), (2) the summer TX of Ankara only shows one breakpoint, and (3) HSP2 of Ankara adequately overlaps all HSPs of the candidate. In contrast, the highest correlated neighboring station, Afyon ($\rho = 0.97$), shows a breakpoint in the same year as the candidate (1966) and is therefore not suitable to homogenize the entire Kutahya series. However, it is also possible to correct the first and second HSPs of Kutahya using reference data from Afyon and reference data from Ankara for correcting the third and fourth HSPs (see Figure 3). There were, however, no significant changes detected after correcting the candidate series using either one or a combination of two references (not shown). A schematic overview of how to select an appropriate reference series for correction is shown in Figure 3.

temperature series. This method is based on a nonlinear model to estimate the relationship between a candidate station and a highly correlated ($\rho > 0.8$ at daily scale) reference station. However, this reference series should not show any breakpoints in the same years as the candidate as

[23] After this we fitted a nonlinear locally weighted regression (LOESS) [Cleveland and Devlin, 1988] to estimate the relationship between the candidate (response, y_i) and the reference (predictor, x_i) before the inhomogeneity (i.e., in the period of common overlap within HSP1 of the candidate; see Figure 3). The smoothing model is given by

$$y_i = g(x_i) + \varepsilon_i, \tag{2}$$

where g is the regression function, i is the i th observation from 1 to N_{mod} (the total number of observations), and ε_i are

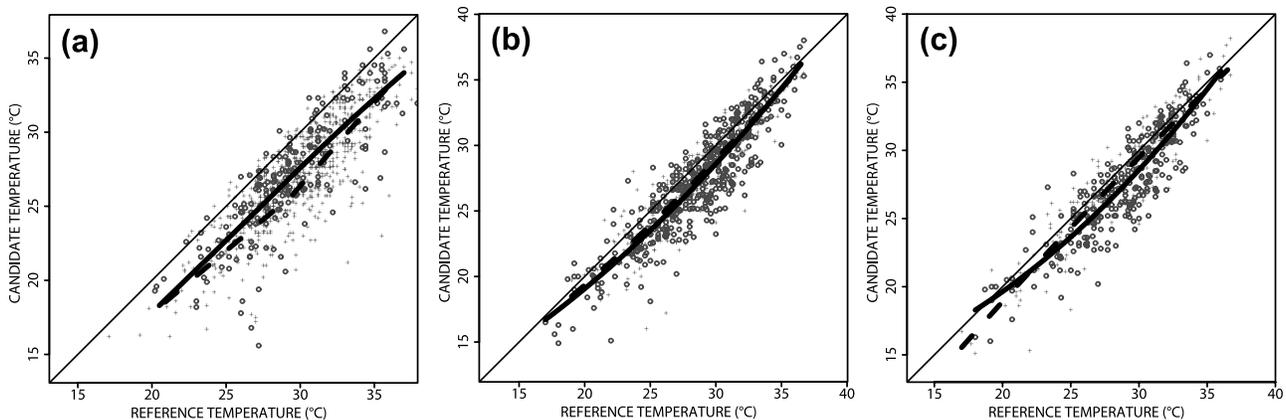


Figure 4. The relationship between Kutahya and the reference station Ankara before (gray circles) and after (gray crosses) with a locally weighted regression (LOESS)-fitted curve before (solid black curve) and after (dashed black curve) each inhomogeneity in HSP1 and (a) HSP2, (b) HSP3, and (c) HSP4 for July. The thin black line in each plot has a slope of 1 for comparison reasons.

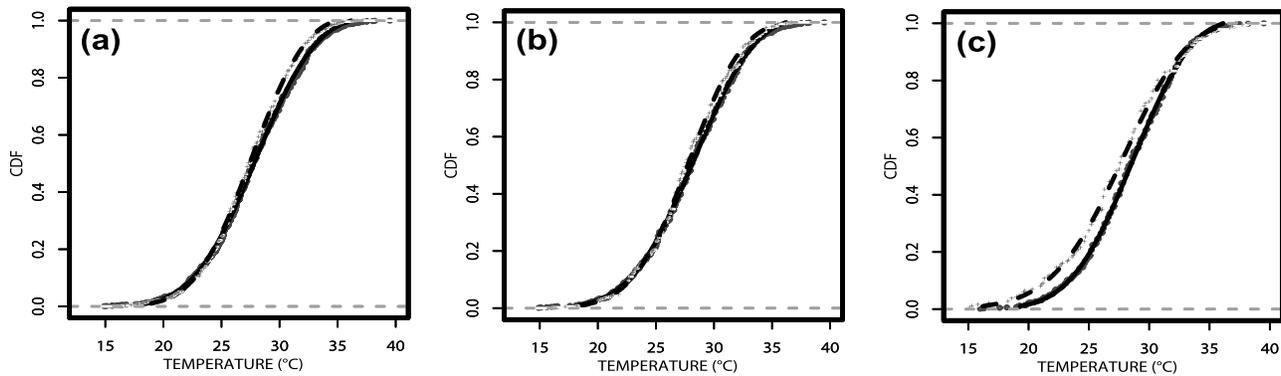


Figure 5. The fitted and sampled CDF before (solid black curve, gray dots) and after (dashed black curve, gray crosses) each inhomogeneity in Kutahya, July, using the generalized extreme value distribution: (a) HSP1 and HSP2, (b) HSP1 and HSP3, and (c) HSP1 and HSP4.

the random errors. A number of parameters control the regression function g : These are the smoothing parameter, α ($\alpha = 3$), the degree of the local fitted polynomial λ ($\lambda = 2$) and the Gaussian distribution of random errors. Sometimes these parameters were altered for individual stations to limit the amount of overfitting of the nonlinear models. See the work of *Della-Marta and Wanner* [2006] for more details on parameter settings and their sensitivity.

[24] Figure 4a shows the relationship of the candidate (Kutahya) versus the reference (Ankara) and the LOESS-fitted function for HSP1 (Figure 4a, solid black curve) and HSP2 (Figure 4a, dashed black curve) as an example for July. The other summer months show a very similar pattern (not shown). After correcting HSP2, this subperiod is merged with HSP1.

[25] Figures 4b and 4c present the same for the subperiod HSP1 and HSP3, and HSP1 and HSP4, respectively. The fitted curves are almost parallel in Figures 4a and 4b with the dashed black curve below the solid black curve on average, indicating that the response of the temperatures in HSP2–3 to the reference temperature are on average lower than in HSP1. Only the lowest temperatures in HSP2 are on average higher than in HSP1. In HSP4 (Figure 4c) we detected higher values for the majority of the records except for the upper and lower tail of the temperature distribution.

[26] The model shows (Figures 4a–4c) that before and after the inhomogeneities the overall corrections are small (on the order of 0.1–0.3°C). The model (e.g., Figure 4a, solid black line) was then used to estimate the observations at the candidate after the first inhomogeneity (i.e., Figure 3, in HSP2 of Kutahya) given homogeneous observations from the reference. The differences between the observed inhomogeneous values of the candidate (HSP2) and the model-fitted values after the inhomogeneity were binned according to which decile the model-fitted values were placed in the candidate observed cumulative distribution function (CDF), defined using the homogeneous temperatures before the inhomogeneity. The CDF was fitted by comparing the goodness of fit statistics of six distributions (Kolmogorov-Smirnov test). By comparing the July-CDF of the candidate in HSP1 and HSP2 (Figure 5a) using the generalized extreme value distribution (GEV), it is apparent that the dashed black curve CDF is on average cooler than the solid black curve CDF.

[27] The adjustments for Kutahya (Figure 6a) indicate that for HSP2 the largest shift in the mean of approximately 0.8°C is needed for decile 6, but also a change in skewness is required by making decile 1 less extreme by up to 0.8°C and decile 10 more extreme by 0.2°C in HSP2. After this step we fitted a LOESS function to the binned decile

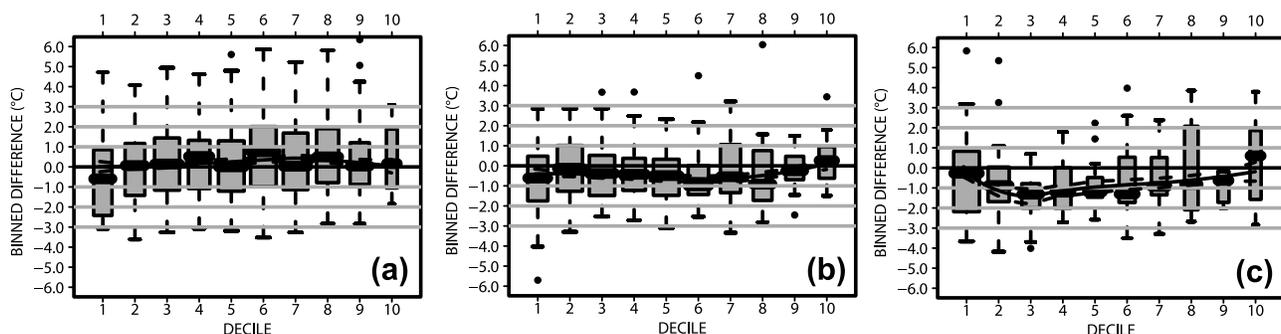


Figure 6. The smoothed adjustments (°C) for each decile in Kutahya, July, shown as a solid black curve fitted using a LOESS function for (a) HSP2, (b) HSP3, and (c) HSP4. The box plots indicate the mean of the binned differences (black line), the interquartile range (shaded area), 1.5 times the interquartile range (dashed black line), and outliers (black dots). The width of the box indicates the relative number of observations in each. The dashed black curves show the 95% confidence interval of the fitted curve.

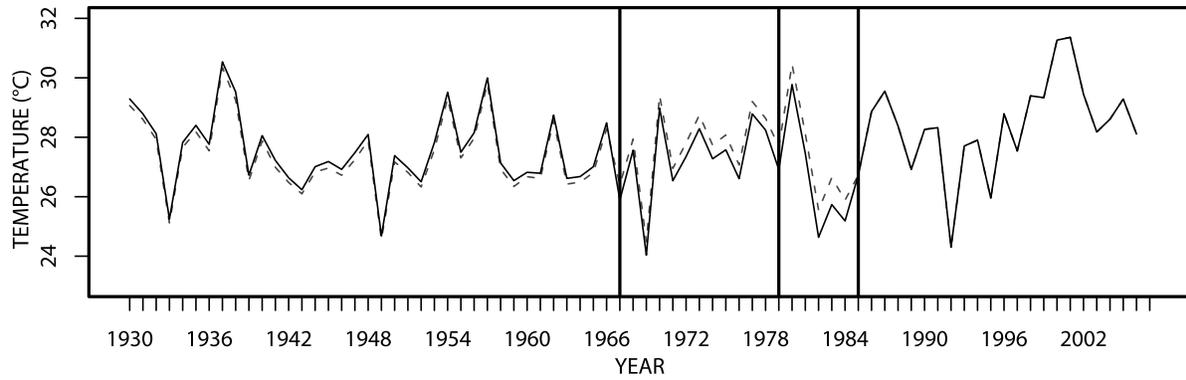


Figure 7. A comparison of the July monthly averaged daily Kutahya inhomogeneous time series (dashed gray curve) and the homogenized time series using PENHOM (solid black curve). Black vertical lines characterize the breakpoints detected and boundaries between the HSPs.

adjustments to obtain a smoothly varying set of adjustments between each decile. The mean of the adjustments given by the LOESS fit calculated over all deciles and all breakpoints is $+0.14^{\circ}\text{C}$.

[28] Figure 7 compares the July averaged daily maximum temperature series using PENHOM method with the non-homogenized July averaged raw series. It is apparent that the series between 1966 and 1984 has been made slightly cooler (-0.2°C on average) while the series between 1930 and 1966 has been made slightly warmer ($+0.2^{\circ}\text{C}$ on average) after PENHOM was applied. The mean adjustment of the coldest and warmest 10% of the temperature distri-

bution was $+0.16^{\circ}\text{C}$ and $+0.29^{\circ}\text{C}$, respectively. For this series the overall trend of July temperatures has not changed over the last 78 years (1930–2007) after homogenization.

2.2.3. Rebreak Detection

[29] After the correction of all breakpoint inhomogeneities, each of the homogenized candidate series is tested again for breakpoints using the methods described in section 2.2.1. This process uses the same reference series as used in the first application of PENHOM. The reanalysis of the homogenized Kutahya series shows few remaining and no more clustered breakpoints within two consecutive years (see Figure 8). Hence, we assume the Kutahya series

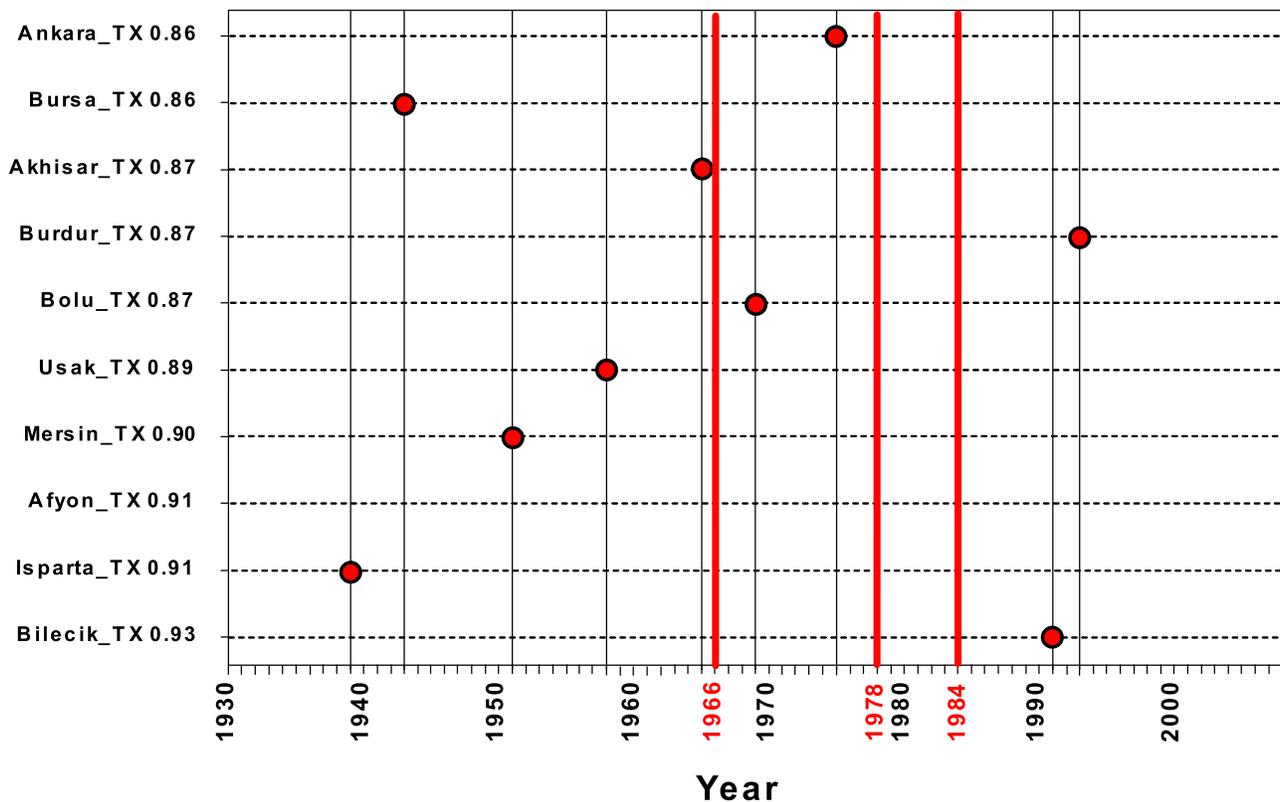


Figure 8. Synthesis of the detected breakpoints in the corrected Kutahya series (see Figure 2 for details) after the first correction.

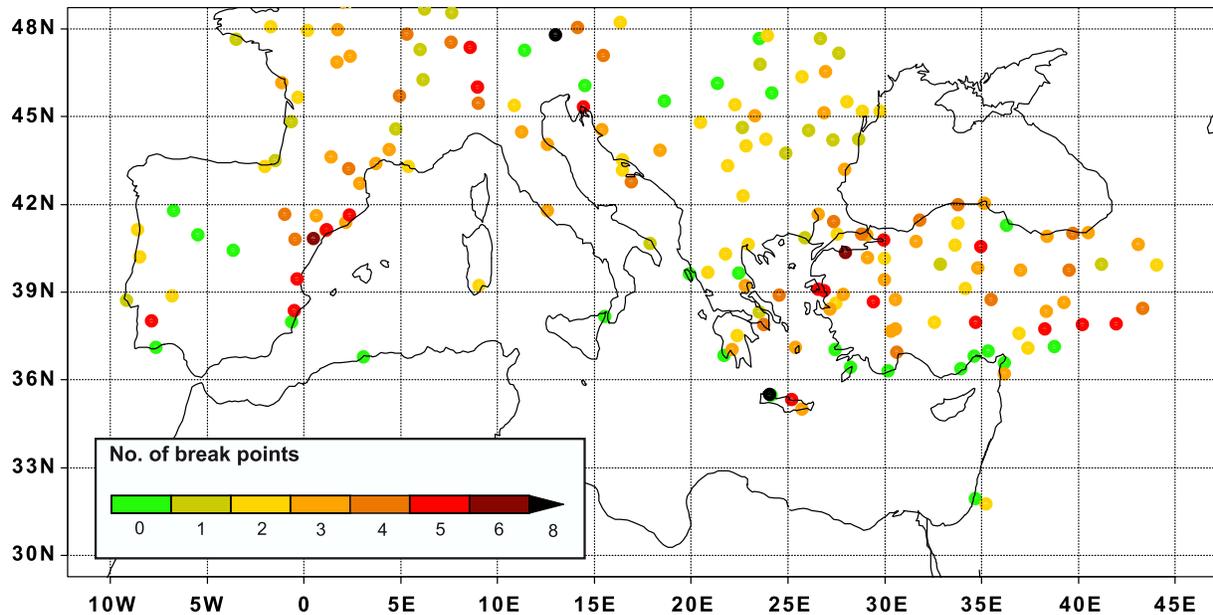


Figure 9. The number of breakpoints detected for each series indicated by color.

to be homogeneous. The single remaining breakpoints in different years are considered not to be reliable, and they might indicate (1) erroneous shifts of minimal size and/or (2) an effect of breakpoints in the reference series. An exclusion of them is needed to get long enough homogeneous subperiods and overlapping periods allowing a reliable correction.

[30] The break detection for many other series was not as clear as shown for Kutahya and often needed the inclusion of additional breakpoints after the first application of detection/correction. Taking the corrected series from the first application of the detection/correction we applied the detection method again. If we found additional breakpoints (using the criteria above) we considered them and corrected the raw series with this additional info and so on. These steps were repeated until no more clustered breakpoints were detected among the difference series. Typically, this procedure did not exceed four iterations.

3. Results and Discussion

[31] PENHOM has been applied to 240 daily summer maximum temperature series and gives reliable results for 174 series across the GMR. The need for highly correlated reference series means that our homogenization procedure produces the most reliable correction of TX in coastal areas and plains for the summer season (JJAS). 60 station series were excluded after application of PENHOM owing to (1) too many breakpoints detected (i.e., too short overlapping periods) and/or (2) the absence of highly correlated reference series either for breakpoint detection or breakpoint correction.

[32] Figure 9 summarizes the number of breakpoints detected for all 174 temperature series. In 22% of the series one breakpoint was found, in 47% of the series two to three breakpoints, in 23% of the series four to five breakpoints and in 2% of the series more than five breakpoints were identified. The western (one breakpoint in 36 years) and

central Mediterranean sites (one breakpoint in 34 years) show significantly less breakpoints per decade than eastern Mediterranean ones (one breakpoint in 24 years). Only 28 out of 174 original series (equals 16%) can be considered as homogeneous. Most of them are located in southern Turkey (7 series), in Greece (5 stations), the northern Balkan (5 series) and in the Iberian Peninsula (5 series) and are often not longer than 50 years.

[33] Figure 10 shows the mean adjustments applied to all daily summer maximum temperature data (Figure 10a), the highest 10% of the summer maximum temperature distribution (Figure 10b), and the lowest 10% of the summer maximum temperature distribution (Figure 10c) for all records homogenized. Owing to a unique history of each site there is no uniform trend of the adjustments over large areas of the Mediterranean detectable. However, the pattern of adjustments is more homogeneous over Anatolia, Romania and some smaller areas in the western Mediterranean. The following summary excludes Austrian and Swiss time series owing to a different pattern compared to more southerly located areas.

[34] The mean adjustment of all daily summer maximum data is $+0.03^{\circ}\text{C}$ ($\pm 0.38^{\circ}\text{C}$; standard error) for the western Mediterranean, $+0.16^{\circ}\text{C}$ ($\pm 0.52^{\circ}\text{C}$) for the central Mediterranean, and $+0.19^{\circ}\text{C}$ ($\pm 0.30^{\circ}\text{C}$) for the eastern Mediterranean, indicating a reduced increase in summer daytime temperature than assumed by analyzing raw data. 113 out of 174 stations show mean adjustments between -0.25°C and $+0.25^{\circ}\text{C}$. Only six stations were corrected by more than 1.0°C (Lastovo, Croatia; Cagliari, Italy; Vila Seca, Spain; Lugano and Zurich, Switzerland; Trabzon, Turkey). The mean adjustments of the warmest decile show quite conservative values of -0.02°C ($\pm 0.56^{\circ}\text{C}$) and -0.05°C ($\pm 0.40^{\circ}\text{C}$) for the western and eastern Mediterranean, respectively. Mean positive adjustments of $+0.15^{\circ}\text{C}$ ($\pm 0.66^{\circ}\text{C}$) indicate a reduced increase of hottest summer maximum temperatures in the central Mediterranean. 98 out of 174 stations show mean adjustment between -0.25°C and $+0.25^{\circ}\text{C}$. Sixteen

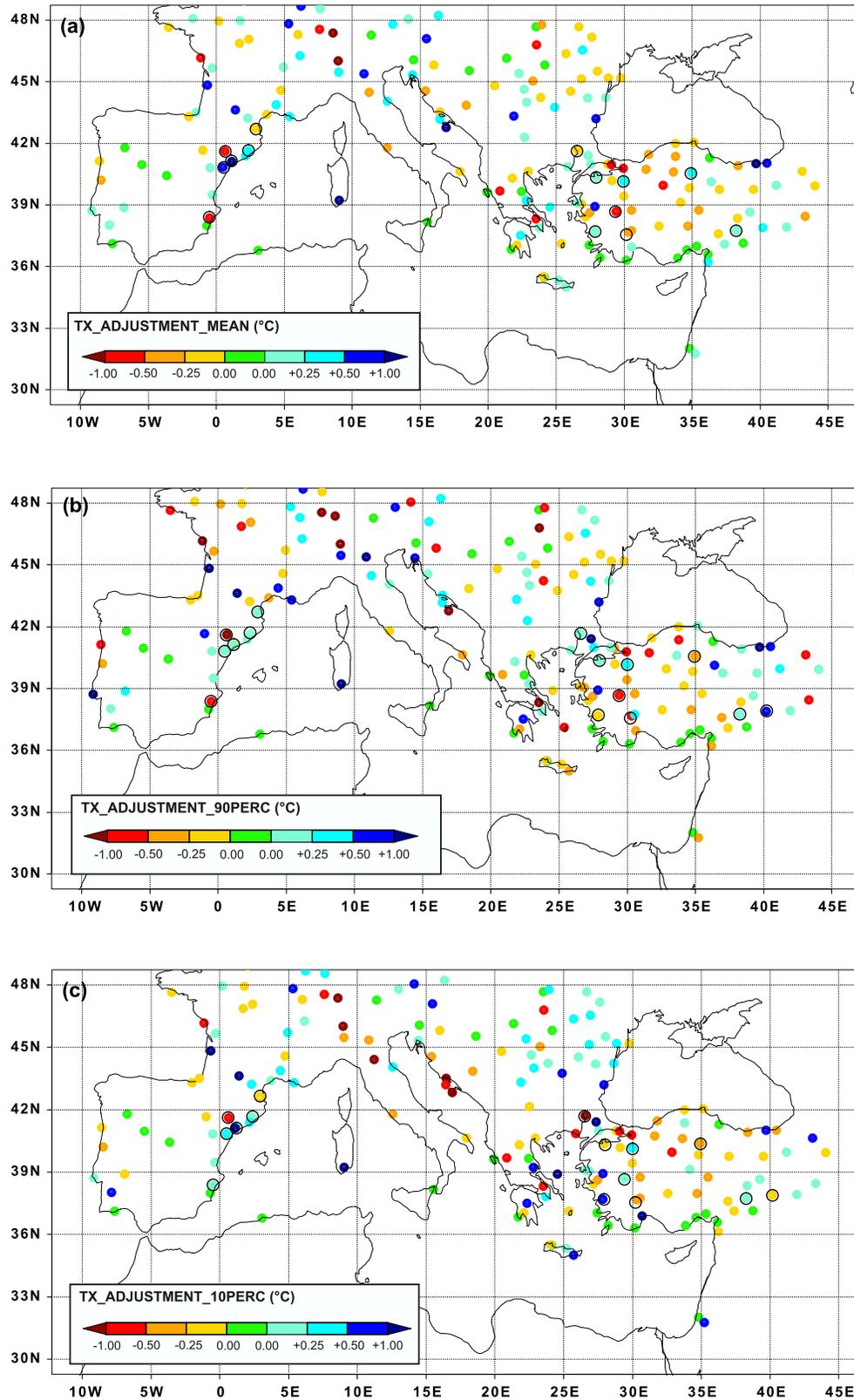


Figure 10. Mean adjustments (°C) applied to (a) all daily summer maximum temperature data, (b) the highest 10% of the maximum temperature distribution, and (c) the lowest 10% of the maximum temperature distribution. The black circles indicate station series that were corrected using more than one reference series.

Table 2. Summary of the Correction Quality Based on the Correlation Coefficient of the Reference Series and the Number of the Modeled and Predicted Temperature Values^a

Quality of Correction	Percentage of BPs Corrected, %	Correlation Coefficient	Number of Modeled (N_{mod}) + Predicted (N_{pred}) Values
Very good	27	$\rho \geq 0.90$	$N_{\text{mod}} + N_{\text{pred}} \geq 1000$
Good	52	$0.80 \leq \rho < 0.90$	$500 < N_{\text{mod}} + N_{\text{pred}} < 1000$
Acceptable	16	$0.80 \leq \rho < 0.90$	$180 < N_{\text{mod}} + N_{\text{pred}} \leq 500$
Poor	5	$0.80 \leq \rho < 0.90$	$N_{\text{mod}} + N_{\text{pred}} \leq 180$

^aSee text for details.

stations were corrected by more than 1.0°C. The adjustments of the coldest decile are between +0.11°C (±0.47°C), -0.05°C (±0.60°C) and -0.83°C (±1.28°C) in the western, eastern, and central Mediterranean, respectively. Also here, 98 out of 174 stations (56%) show mean adjustment between -0.25°C and +0.25°C. Fifteen stations were corrected by more than 1.0°C.

[35] The correction quality for each series and HSP is based on (1) the correlation coefficient between the candidate and the reference series and (2) the number of modeled (N_{mod}) and predicted values (N_{pred}) [Della-Marta and Wanner, 2006]. On the basis of the modeling study by Della-Marta and Wanner [2006], we formulated a set of correction quality criteria which resulted in each correction of each time series to be classified as either “very good,” “good,” “acceptable,” and “poor.” According to our criteria listed in Table 2, only 5% of the total number of corrections were identified as “poor,” and 79% of corrections were classified as either “good” or “very good.” The stations where “poor” corrections were made are scattered throughout the GMR, however there is a tendency for poor correction to be made for series located in southern Italy, Sardinia, and Israel. A summary of the correction quality is given in Table 2.

[36] For estimating changes in maximum temperature variability we calculated the difference between the mean adjustments (°C) applied to the highest 10% and the lowest 10% of the maximum temperature distribution. Figure 11 shows an area of increasing daily maximum temperature variability from southeastern France across the Adriatic area, the southern Balkan, and northern Greece toward northern Turkey.

[37] The mean July maximum temperatures before and after PENHOM and their changes in trends are shown for three key stations in the western (Lisbon, Portugal; Valencia, Spain; Montelimar, France), central (Graz, Austria; Verona and Brindisi, Italy), and eastern Mediterranean areas (Bucharest, Romania; Manisa and Diyarbakir, Turkey) in Figure 12. The overall effect of PENHOM is often accompanied by changes in the mean long-term maximum temperature trend. For Graz, Austria, temperature adjustments of more than 2°C result in a significantly reduced temperature increase. The mean summer maximum temperature trend changed from +0.21°C before to +0.13°C/decade after PENHOM was applied between 1900 and 2005. This agrees with the findings of Della-Marta and Wanner [2006], who also corrected the daily maximum temperature of Graz in this study. Besides Graz, summer TX was also corrected for

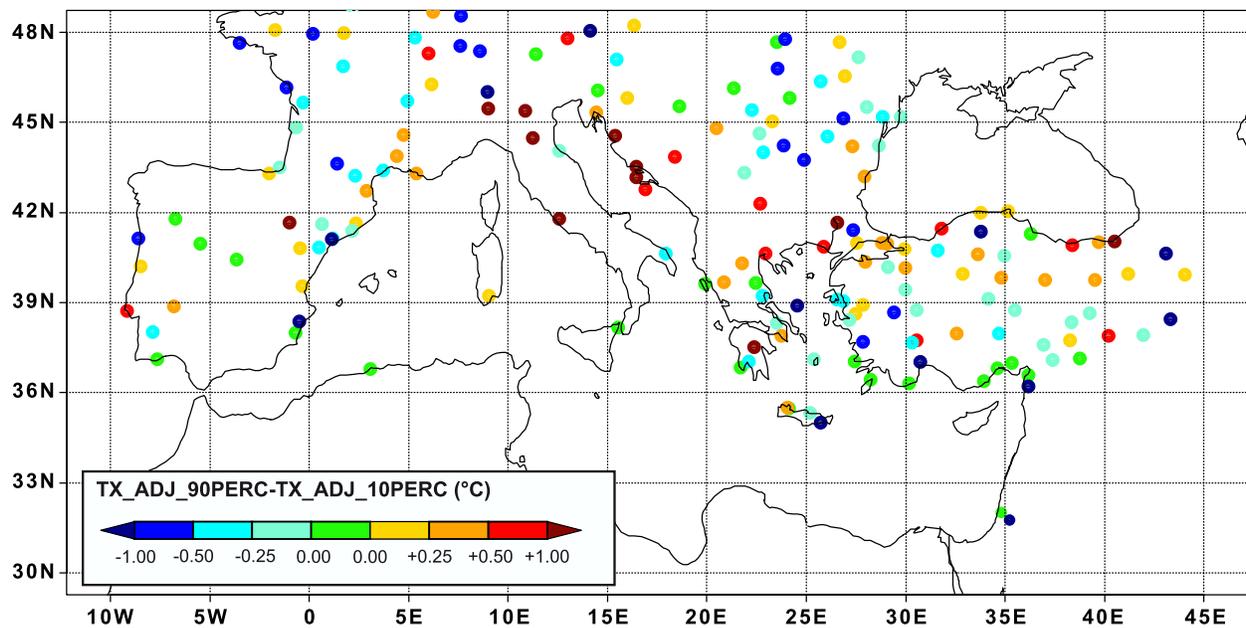


Figure 11. Difference between the mean adjustments (°C) applied to the highest 10% and the lowest 10% of the maximum temperature distribution. The black circles indicate station series that were corrected using more than one reference series. Warm colors indicate an increase in daily maximum temperature variance, and vice versa for cool colors.

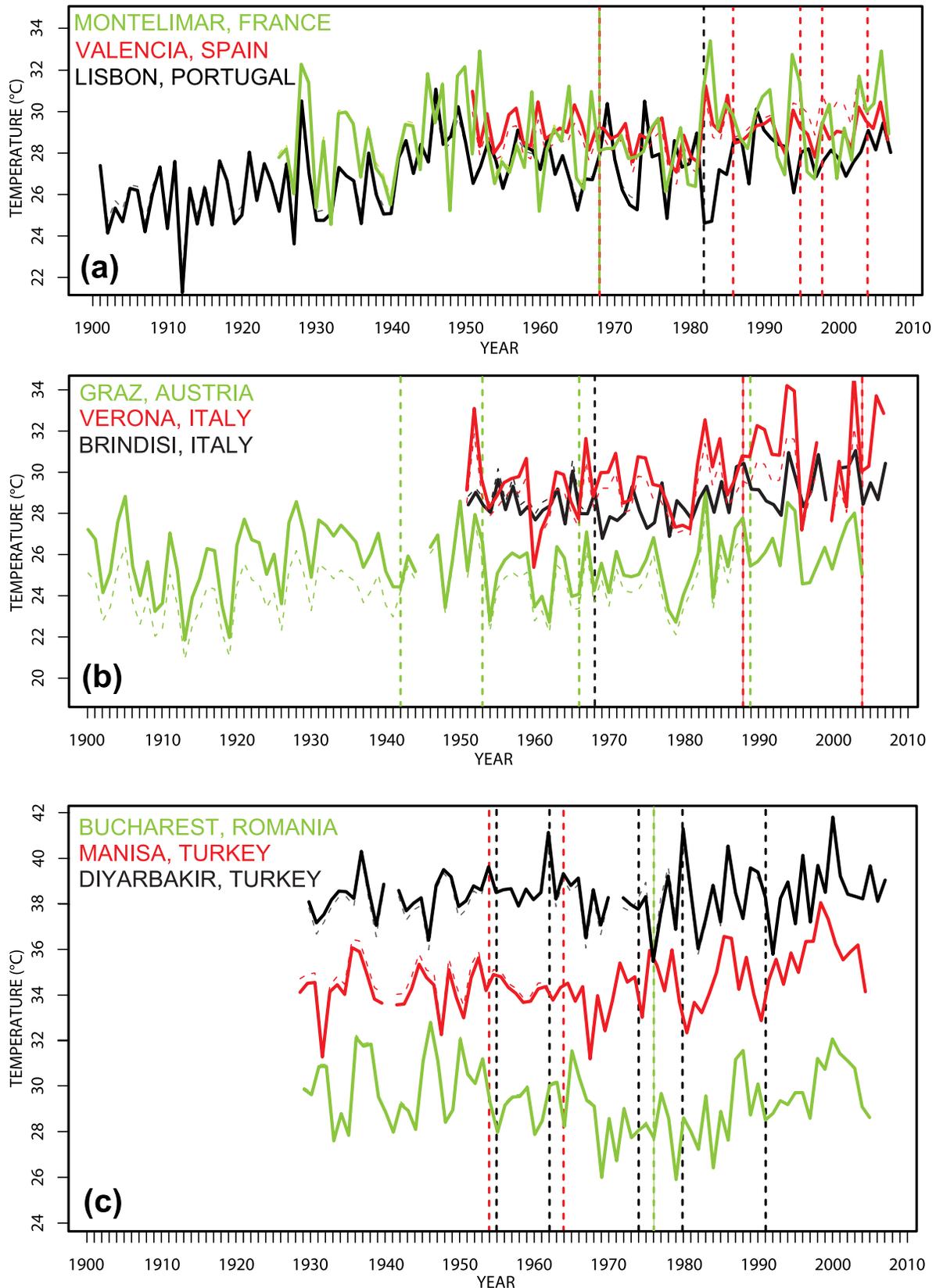


Figure 12. A comparison of the July monthly averaged daily inhomogeneous time series (dashed curves) and the homogenized time series using the PENHOM (solid curves) for three representative stations in (a) the western Mediterranean, (b) the central Mediterranean, and (c) the eastern Mediterranean. Dotted vertical lines characterize the breakpoints detected for the respective station.

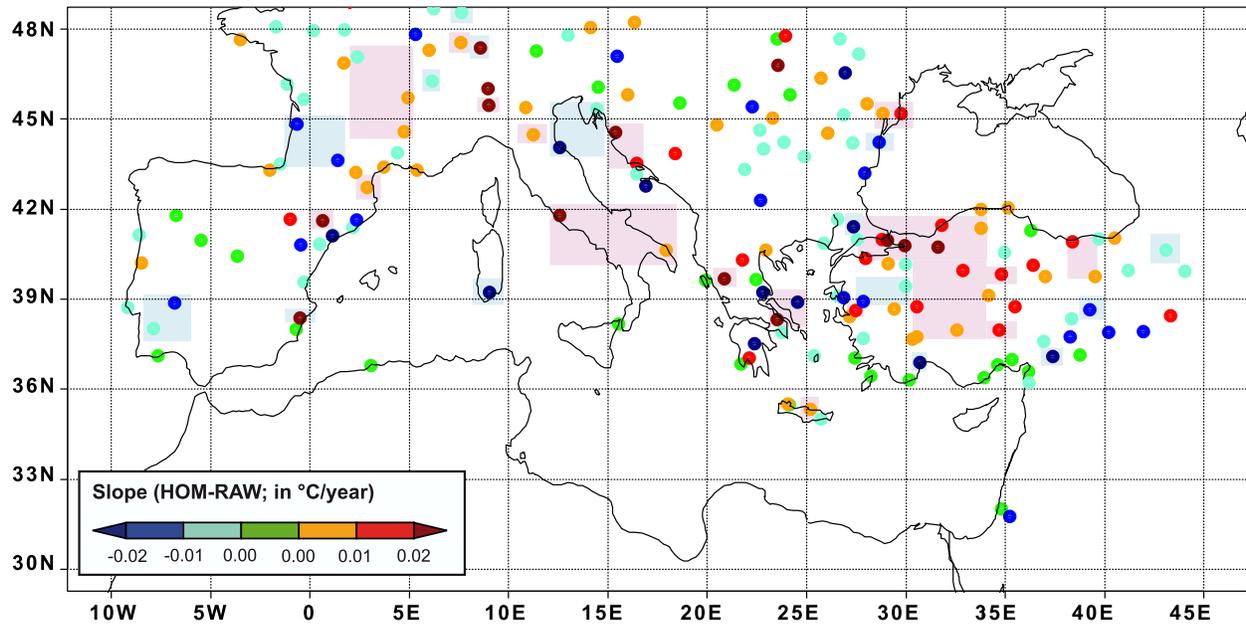


Figure 13. Differences in the slopes (PENHOM-RAW) are indicated by color for each site. Light red areas characterize records with significantly lower p values involving an increased temperature trend after PENHOM was applied.

Kremsmunster (Austria), Paris (France), Lisbon (Portugal), Basel, Geneva and Zurich (Switzerland) by *Della-Marta* [2006]. Comparison of this study with our results covering the same stations indicates close agreement (not shown).

[38] The linear trends (°C/decade) of the mean, the highest 10%, and the lowest 10% of the July maximum temperature distribution for the nine key stations are summarized in Table 3. Figure 13 summarizes the detected differences in slopes between PENHOM and RAW trends and identifies areas containing temperature records with significant changes in the p value of the trend after PENHOM was applied. Hence, red areas characterize records with either significantly lower ρ values involving an increased temperature trend after the data were homogenized. Such trends were identified for some stations in the northwestern parts (Basel, Lleida, Lyon, Montelimar, Perpignan, and Vichy-Charmeil) in Italy and the Adriatic region (Bologna, Brindisi, Gospic, Ioannina, Milan, Rome, and Split), many stations in Western Anatolia (Afyon, Ankara, Bolu, Cankiri,

Erzincan, Giresun, Isparta, Konya, Nigde, Yozgat, and Zonguldak), in the Marmara region (Bandirma, Florya/Istanbul, Goztepe/Istanbul, Istanbul, and Izmit), in parts of the Aegean (Heraklion, Manisa, Skyros, and Tanagra), and along the Romanian Black Sea coast (Sulina and Tulcea). Significantly reduced temperature trends are usually found for more isolated stations and not for larger areas (see Figure 13, blue areas).

[39] The station series that were excluded after PENHOM was applied are primarily located in mountainous areas that are affected by orographical effects (e.g., Foehn) that lead to correction problems. As an example for a nonreliable homogenization, Figure 14 illustrates the smoothed adjustments (°C) for July and each decile in Ransol, Andorra, a station in 1645 amsl that could be homogenized neither by lowland nor by neighboring mountainous series. Highly nonlinear adjustments between -4.1°C (median, decile 1) and $+2.6^{\circ}\text{C}$ (median, decile 9) for July cannot be assumed to be valid and underlines the need for highly correlated

Table 3. Long-Term Trends of the Mean, the Highest 10%, and the Lowest 10% of the July Maximum Temperature Distribution for Nine Key Stations Before and After PENHOM Was Applied^a

Station, Country	Trend of Mean July TX (°C/decade)		Trend of Mean July TX90PERC (°C/decade)		Trend of Mean July TX10PERC (°C/decade)	
	Before PENHOM	After PENHOM	Before PENHOM	After PENHOM	Before PENHOM	After PENHOM
Lisbon, Portugal	0.22	0.25	0.13	-0.06	0.22	0.36
Valencia, Spain	0.18	0.12	0.19	0.17	0.14	0.12
Montelimar, France	0.12	0.16	0.13	0.14	0.00	0.12
Graz, Austria	0.21	0.13	0.27	0.08	0.19	0.23
Verona, Italy	0.18	0.15	0.13	0.27	0.11	0.16
Brindisi, Italy	0.10	0.20	0.20	0.29	0.24	0.24
Bucharest, Romania	-0.05	-0.03	-0.05	-0.04	-0.08	-0.08
Manisa, Turkey	0.06	0.12	0.11	0.16	0.08	0.10
Diyarbakir, Turkey	0.06	-0.09	0.04	-0.11	0.12	-0.08

^aValues for observed trends significant at the 5% level (Mann-Kendall test) are shown in bold.

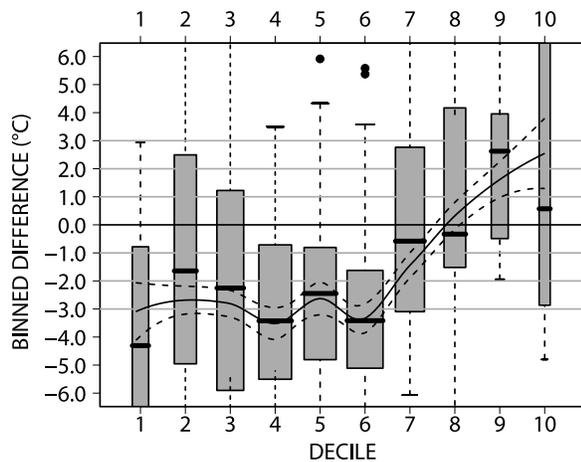


Figure 14. The smoothed adjustments ($^{\circ}\text{C}$) for July and each decile in Ransol, Andorra (see Figure 6 for further details).

reference series. Similar difficulties appear when correcting TX out of summer season or daily minimum temperatures (TN) owing to lower spatial correlations. A reliable homogenization of these parameters requires much higher station densities.

[40] Since the method only uses reference station data for correcting the candidate data, it is important that all breakpoints in those series are accurately determined. If several suitable reference series are available to homogenize a candidate series, a comparison between the adjustments from each might be helpful. In Romania where many series show only one breakpoint, or are even homogeneous, a correction using more than one reference series was possible for a couple of station records (Botosani, Bucuresti, Calarasi, Cluj Napoca, Constanta, and Iasi). The mean adjustments of all values and the highest/lowest deciles varied by less than 0.05°C underlining the significance of the correction procedure in this region. Unfortunately, similar investigations are not possible for other Mediterranean areas. Therefore a confirmation of the breakpoints detected using statistical procedures from metadata is highly recommended [Peterson *et al.*, 1998; Wijngaard *et al.*, 2003], but in many cases this is also not possible owing to missing information or this information is very difficult to be accessed in the archives.

[41] The results show that 84% of all temperature series analyzed are affected by artificial breakpoints. Even if the mean adjustments for all sites are only small and slightly positive ($+0.1^{\circ}\text{C}$), some series changed dramatically after PENHOM was applied leading to significant changes in regional temperature trends.

[42] A direct comparison with many other previously mentioned homogenization studies [e.g., Böhm *et al.*, 2001; Begert *et al.*, 2005; Brunetti *et al.*, 2006; Auer *et al.*, 2007] that have estimated their shifts to the mean from a reference series that generally is a weighted average of many surrounding stations, is often not possible. A single reference station has the advantage of representing the variability of climate in a smaller scale, whereas weighted reference series are more representative of a climatic region. The disadvantage of weighted reference series is that they are assumed to be homogeneous even when they are usually

based on nonhomogenized individual stations records. In order to take into account more than one reference series in the correction procedure and the application of advanced statistical methods (e.g., universal kriging) are currently under investigation.

4. Conclusion and Outlook

[43] The aim of this study was to homogenize a comprehensive and high-quality daily maximum temperature database for the Greater Mediterranean Region. The lack of metadata providing information about station history and potential artificial breakpoints directed us to develop an improved homogenization method (PENHOM) based on statistical methods only.

[44] We show that a combination of the break detection method by *Caussinus and Mestre* [2004] and the correction model by *Della-Marta and Wanner* [2006] is a suitable way to homogenize daily meteorological series reliably (see results in Table 2) given that highly correlated neighboring reference series are available. The need for highly correlated reference series results in our procedure being most reliable in coastal areas, plains and during the summer season (JJAS).

[45] PENHOM has been applied to 240 daily summer maximum temperature series and gives reliable results for 174 series in 16 countries covering the last 50–150 years across the GMR. Each of the 174 series has a maximum of three missing values per month per season (JJAS) and is homogenized using a reference series with which it is highly correlated (greater than 0.8).

[46] Results show that 86% of all TX series are affected by one or more artificial breakpoints. Station records in the eastern Mediterranean show significantly more breakpoints than records in the western Mediterranean. Owing to a unique history of each site there is no uniform trend of the adjustments across the whole GMR detected. The mean adjustment of all daily summer maximum temperatures is $+0.03^{\circ}\text{C}$ ($\pm 0.38^{\circ}\text{C}$) for the western Mediterranean, $+0.16^{\circ}\text{C}$ ($\pm 0.52^{\circ}\text{C}$) for the central Mediterranean, and $+0.19^{\circ}\text{C}$ ($\pm 0.30^{\circ}\text{C}$) for the eastern Mediterranean, indicating a reduced increase of the mean summer maximum temperature compared to findings from analyzing raw data.

[47] This is the first time that one coherent daily temperature homogenization technique has been applied to such a large amount of data on a continental scale. Using criteria based on the sensitivity study of *Della-Marta and Wanner* [2006], 79% of corrections made to the 174 series were classified as “good” or “very good.”

[48] Forthcoming work focuses on the detailed analysis of trends in maximum temperature extremes and homogenizing Mediterranean daily minimum temperature (TN) database. With such data sets the detailed analysis of heat waves and related temperature based heat stress indices in the GMR over the last decades will be possible. Other benefits of this data set would allow a better understanding of dynamical processes causing extreme temperature events on different time and space scales. Furthermore homogenized TX and TN series are a fundamental necessity for any kind of climate change impact research.

[49] Beside an enhanced collaboration of the NMHS, the WMO-MEDARE initiative (www.omm.urv.cat/MEDARE/)

index.html) aims to improve both the Mediterranean data exchange and the recovery of climate data and metadata during the next years. This implies more information concerning site history, a validation and comparison of breakpoints detected with documentary records, and more reliable corrections due to higher station density.

[50] **Acknowledgments.** We are grateful to Gerard van der Schrier and partners from the European Climate Assessment and Data set (ECA&D) at the KNMI, Franco Desiato from the ISPRA (SCIA project), Enric Aguilar, Pere Esteban, Marc Prohom, Dimitra Founda, Ozan Mert Gokturk, and the Turkish State Meteorological Service for providing data. We thank Olivier Mestre for his break detection routine and interesting discussions. This research was funded by the European Union integrated project Climate Change and Impact Research: The Mediterranean Environment (CIRCE) and the European COST Action HOME ES0601 (Advances in homogenization methods of climate series: An integrated approach—HOME). Jürg Luterbacher acknowledges support from the 7th EU Framework program Assessing Climate Impacts on the Quantity and Quality of Water (ACQWA, <http://www.acqwa.ch/>) and MedClivar (<http://www.medclivar.eu/>). The three reviewers made useful comments and suggestions and helped to improve the quality of this study.

References

- Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa (2003), Guidance on metadata and homogenization, *WMO-TD 1186*, 51 pp., World Meteorol. Organ., Geneva, Switzerland.
- Alexander, L. V., et al. (2006), Global observed changes in daily climate extremes of temperature and precipitation, *J. Geophys. Res.*, *111*, D05109, doi:10.1029/2005JD006290.
- Alexandersson, H., and A. Moberg (1997), Homogenization of Swedish temperature data. Part 1: Homogeneity test for linear trends, *Int. J. Climatol.*, *17*, 25–34, doi:10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J.
- Alpert, P., et al. (2002), The paradoxical increase of Mediterranean extreme daily rainfall in spite of decrease in total values, *Geophys. Res. Lett.*, *29*(11), 1536, doi:10.1029/2001GL013554.
- Auer, I., et al. (2005), A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002, *Int. J. Climatol.*, *25*, 139–166, doi:10.1002/joc.1135.
- Auer, I., et al. (2007), HISTALP: Historical instrumental climatological surface time series of the Greater Alpine Region, *Int. J. Climatol.*, *27*, 17–46, doi:10.1002/joc.1377.
- Begert, M., T. Schlegel, and W. Kirchhofer (2005), Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000, *Int. J. Climatol.*, *25*, 65–80, doi:10.1002/joc.1118.
- Beniston, M., et al. (2007), Future extreme events in European climate: An exploration of regional climate model projections, *Clim. Change*, *81*, 71–95, doi:10.1007/s10584-006-9226-z.
- Böhm, R., I. Auer, M. Brunetti, M. Maugeri, T. Nanni, and W. Schoner (2001), Regional temperature variability in the European Alps: 1760–1998 from homogenized instrumental time series, *Int. J. Climatol.*, *21*, 1779–1801, doi:10.1002/joc.689.
- Brandsma, T., and G. P. Können (2006), Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level, *Int. J. Climatol.*, *26*, 75–89, doi:10.1002/joc.1236.
- Brown, S. J., J. Caesar, and C. A. T. Ferro (2008), Global changes in extreme daily temperature since 1950, *J. Geophys. Res.*, *113*, D05115, doi:10.1029/2006JD008091.
- Brunet, M., and F. G. Kuglitsch (Eds.) (2008), *Proceedings of the International Workshop on Rescue and Digitization of Climate Records in the Mediterranean Basin*, WMO-TD 1432, 219 pp., World Meteorol. Organ., Geneva, Switzerland.
- Brunet, M., O. Saladie, P. D. Jones, J. Sigro, E. Aguilar, A. Moberg, D. Lister, A. Walther, D. Lopez, and C. Almarza (2006), The development of a new dataset of Spanish daily adjusted temperature series (1850–2003), *Int. J. Climatol.*, *26*, 1777–1802, doi:10.1002/joc.1338.
- Brunet, M., P. D. Jones, J. Sigro, O. Saladie, E. Aguilar, A. Moberg, P. M. Della-Marta, D. Lister, A. Walther, and D. Lopez (2007), Temporal and spatial temperature variability and change over Spain during 1850–2005, *J. Geophys. Res.*, *112*, D12117, doi:10.1029/2006JD008249.
- Brunetti, M., M. Maugeri, F. Monti, and T. Nanni (2006), Temperature and precipitation variability in Italy in the last two centuries from homogenized instrumental time series, *Int. J. Climatol.*, *26*, 345–381, doi:10.1002/joc.1251.
- Caussinus, H., and F. Lyazrhi (1997), Choosing a linear model with a random number of change-points and outliers, *Ann. Inst. Stat. Math.*, *49*, 761–775, doi:10.1023/A:1003230713770.
- Caussinus, H., and O. Mestre (2004), Detection and correction of artificial shifts in climate series, *J. R. Stat. Soc., Ser. C*, *53*, 405–425, doi:10.1111/j.1467-9876.2004.05155.x.
- Cleveland, W. S., and S. J. Devlin (1988), Locally weighted regression: An approach to regression-analysis by local fitting, *J. Am. Stat. Assoc.*, *83*, 596–610, doi:10.2307/2289282.
- Cocheo, C., and D. Camuffo (2002), Corrections of systematic errors and data homogenisation in the daily temperature Padova series (1725–1998), *Clim. Change*, *53*, 77–100, doi:10.1023/A:1014950306015.
- Della-Marta, P. M. (2006), Extreme temperature events in western Europe from 1880–2003, Ph.D. thesis, Inst. of Geogr., Univ. of Bern, Bern, Switzerland.
- Della-Marta, P. M., and H. Wanner (2006), A method of homogenizing the extremes and mean of daily temperature measurements, *J. Clim.*, *19*, 4179–4197, doi:10.1175/JCLI3855.1.
- Della-Marta, P. M., D. Collins, and K. Braganza (2004), Updating Australia's high-quality annual temperature dataset, *Aust. Meteorol. Mag.*, *53*, 75–93.
- Della-Marta, P. M., J. Luterbacher, H. von Weissenfluh, E. Xoplaki, M. Brunet, and H. Wanner (2007a), Summer heat waves over western Europe 1880–2003: Their relationship to large scale forcings and predictability, *Clim. Dyn.*, *29*, 251–275, doi:10.1007/s00382-007-0233-1.
- Della-Marta, P. M., M. R. Haylock, J. Luterbacher, and H. Wanner (2007b), Doubled length of western European summer heat waves since 1880, *J. Geophys. Res.*, *112*, D15103, doi:10.1029/2007JD008510.
- Demarée, G. R., P. J. Lachaert, T. Verhoeve, and E. Thoen (2002), The long-term daily Central Belgium Temperature (CBT) series (1767–1998) and early instrumental meteorological observations in Belgium, *Clim. Change*, *53*, 269–293, doi:10.1023/A:1014931211466.
- Diffenbaugh, N. S., J. S. Pal, F. Giorgi, and G. Xuejie (2007), Heat stress intensification in the Mediterranean climate change hotspot, *Geophys. Res. Lett.*, *34*, L11706, doi:10.1029/2007GL030000.
- Easterling, D. R., and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, *15*, 369–377, doi:10.1002/joc.3370150403.
- Fang, X. Q., A. Y. Wang, S. K. Fong, W. S. Lin, and J. Liu (2008), Changes of reanalysis-derived Northern Hemisphere summer warm extreme indices during 1948–2006 and links with climate variability, *Global Planet. Change*, *63*, 67–78, doi:10.1016/j.gloplacha.2008.06.003.
- Founda, D., and C. Giannakopoulos (2009), The exceptionally hot summer of 2007 in Athens, Greece: A typical summer in the future climate?, *Global Planet. Change*, *67*, 227–236, doi:10.1016/j.gloplacha.2009.03.013.
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. Tank, and T. Peterson (2002), Observed coherent changes in climatic extremes during the second half of the twentieth century, *Clim. Res.*, *19*, 193–212, doi:10.3354/cr019193.
- Giorgi, F. (2006), Climate change hot-spots, *Geophys. Res. Lett.*, *33*, L08707, doi:10.1029/2006GL025734.
- González-Rouco, J. F., J. L. Jimenez, V. Quesada, and F. Valero (2001), Quality control and homogeneity of precipitation data in the southwest of Europe, *J. Clim.*, *14*, 964–978, doi:10.1175/1520-0442(2001)014<0964:QCAHOP>2.0.CO;2.
- Hegerl, G. C., T. J. Crowley, W. T. Hyde, and D. J. Frame (2006), Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, *440*, 1029–1032, doi:10.1038/nature04679.
- Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, New York.
- Jones, G. S., P. A. Stott, and N. Christidis (2008), Human contribution to rapidly increasing frequency of very warm Northern Hemisphere summers, *J. Geophys. Res.*, *113*, D02109, doi:10.1029/2007JD008914.
- Klein Tank, A., et al. (2002), Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *Int. J. Climatol.*, *22*, 1441–1453, doi:10.1002/joc.773.
- Kostopoulou, E., and P. D. Jones (2005), Assessment of climate extremes in the Eastern Mediterranean, *Meteorol. Atmos. Phys.*, *89*, 69–85, doi:10.1007/s00703-005-0122-2.
- Maugeri, M., L. Buffoni, B. Delmonte, and A. Fassina (2002), Daily Milan temperature and pressure series (1763–1998): Completing and homogenising the data, *Clim. Change*, *53*, 119–149, doi:10.1023/A:1014923027396.
- Moberg, A., and P. D. Jones (2005), Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99, *Int. J. Climatol.*, *25*, 1149–1171, doi:10.1002/joc.1163.
- Moberg, A., et al. (2006), Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, *J. Geophys. Res.*, *111*, D22106, doi:10.1029/2006JD007103.

- Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review, *Int. J. Climatol.*, *18*, 1493–1517, doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.
- Scherrer, S. C., C. Appenzeller, and M. A. Liniger (2006), Temperature trends in Switzerland and Europe: Implications for climate normals, *Int. J. Climatol.*, *26*, 565–580, doi:10.1002/joc.1270.
- Sensoy, S., T. Peterson, L. Alexander, and X. Zhang (2007), Enhancing Middle East climate change monitoring and indexes, *Bull. Am. Meteorol. Soc.*, *88*, 1249–1254, doi:10.1175/BAMS-88-8-1249.
- Stott, P. A., D. A. Stone, and M. R. Allen (2004), Human contribution to the European heatwave of 2003, *Nature*, *432*, 610–614, doi:10.1038/nature03089.
- Szentimrey, T. (1999), Multiple analysis of series for homogenisation (MASH), in *Proceedings of the Second Seminar for Homogenisation of Surface Climatological Data, WMO-TD 962*, pp. 27–46, World Meteorol. Organ., Geneva, Switzerland.
- Toreti, A., and F. Desiato (2008), Temperature trend over Italy from 1961 to 2004, *Theor. Appl. Climatol.*, *91*, 51–58, doi:10.1007/s00704-006-0289-6.
- Trewin, B. C., and A. C. F. Trevitt (1996), The development of composite temperature records, *Int. J. Climatol.*, *16*, 1227–1242, doi:10.1002/(SICI)1097-0088(199611)16:11<1227::AID-JOC82>3.0.CO;2-P.
- Vincent, L. A. (1998), A technique for the identification of inhomogeneities in Canadian temperature series, *J. Clim.*, *11*, 1094–1104, doi:10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2.
- Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg (2002), Homogenization of daily temperatures over Canada, *J. Clim.*, *15*, 1322–1334, doi:10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2.
- Wang, X. L. (2003), Comments on “Detection of undocumented change-points: A revision of the two-phase regression model”, *J. Clim.*, *16*, 3383–3385, doi:10.1175/1520-0442(2003)016<3383:CODOUC>2.0.CO;2.
- Wijngaard, J. B., A. Tank, and G. P. Konnen (2003), Homogeneity of 20th century European daily temperature and precipitation series, *Int. J. Climatol.*, *23*, 679–692, doi:10.1002/joc.906.
- Xoplaki, E. (2009), Effects of extreme climate on Mediterranean societies, *Eos Trans. AGU*, *90*(15), 131, doi:10.1029/2009EO150006.
- Xoplaki, E., J. F. González-Rouco, J. Luterbacher, and H. Wanner (2003), Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs, *Clim. Dyn.*, *20*, 723–739, doi:10.1007/s00382-003-0304-x.
- Xoplaki, E., J. Luterbacher, and J. F. González-Rouco (2006), Mediterranean summer temperature and winter precipitation, large-scale dynamics, trends, *Nuovo Cim.*, *29*, 45–54.
- Zhang, X. B., L. A. Vincent, W. D. Hogg, and A. Niitsoo (2000), Temperature and precipitation trends in Canada during the 20th century, *Atmos. Ocean*, *38*, 395–429.

P. M. Della-Marta, Federal Office for Meteorology and Climatology, MeteoSwiss, CH-8044 Zurich, Switzerland.

F. G. Kuglitsch, A. Toreti, H. Wanner, and E. Xoplaki, Oeschger Centre for Climate Change Research, University of Bern, CH-3012 Bern, Switzerland. (franz.kuglitsch@giub.unibe.ch)

J. Luterbacher, Climatology, Climate Dynamics, and Climate Change Section, Department of Geography, Justus-Liebig University of Giessen, D-35390 Giessen, Germany.