

An assessment of the role of homogenization protocol in the performance of daily temperature series and trends: application to northeastern Spain

A. El Kenawy,^{a,b,*} J. I. López-Moreno,^a P. Stepanek^c and S. M. Vicente-Serrano^a

^a Instituto Pirenaico de Ecología-CSIC, Campus de Aula Dei, P.O. Box 13034, Zaragoza, 50059, Spain

^b Department of Geography, Mansoura University, Mansoura, Egypt

^c Czech Hydrometeorological Institute, Brno, Czech Republic

ABSTRACT: This paper gives the complete details of the protocols applied for developing a spatially and temporarily high-resolution dataset of temperature for northeastern Spain. Our methodologies used data from a large number of observatories (1583) spanning some portions of the period between 1900 and 2006. The raw dataset was first tested for internal and external consistency to check data quality. To improve data completeness, a linear regression model was then utilized to infill gaps in the daily temperature series using the best correlated data from nearby sites. Discontinuities in the reconstructed series were determined by combining the results of three homogeneity-relative tests: the Standard Normal Homogeneity Test (SNHT), the Easterling and Peterson two-phased regression method and the Vincent test. To assess the possible impact of data homogenisation on trends and statistical properties of the final series, a set of tests (e.g. semivariance models and L-moment statistics) was applied to the series before and after correction. Semivariance models suggest a significant improvement in the spatial dependence of the corrected dataset on both seasonal and annual timescales. Also, L-moments gave no evidence of significant changes in the probability distribution of daily temperature series after correction. Taken together, the newly compiled dataset seems to be more robust and reveals more coherent spatial and temporal patterns of temperature compared with the original dataset. From the temporal and spatial perspectives, the new dataset comprises the most complete register of temperature in northeast Spain (1900–2006), with a reasonably spatial coverage. Accordingly, this database can provide a more reliable base for studying temperature changes and variability in the region. This dataset can also be of particular relevance to a number of meteorological, ecological, hydrological and agricultural applications on local, regional and continental scales. Copyright © 2011 Royal Meteorological Society

KEY WORDS daily temperature; homogenisation; quality control; reconstruction; trends; semivariance; Spain

Received 28 April 2010; Revised 18 October 2011; Accepted 25 October 2011

1. Introduction

Complete, reliable and spatially dense climatic datasets are mandatory for different types of climatic analyses (Eischeid *et al.*, 2000). An appropriate analysis of climate variability and trends necessitates climatic datasets of fine spatial and temporal resolution. Nonetheless, the usefulness of many daily datasets is still degraded across many areas of the world. Owing to various institutional and financial constraints, numerous datasets still include gaps that are distributed at random in space and time (Eischeid *et al.*, 2000). Other datasets suffer from lack of data quality assurance or a sparse network of stations.

The Mediterranean is one of the regions that is most sensitive to climate change. The region is characterized by large topographic contrasts and high variability of climate. In addition, this region encompasses a

climatic gradient between mid-latitude and subtropical regimes. For these reasons, the Mediterranean climate has recently been the main focus of intensive research on climate change and variability. However, the utility of daily climate datasets is still degraded in much of the Mediterranean countries, including Spain, as a consequence of relocations of the meteorological observatories. The Mediterranean has extensive archives of instrumental climate records which, in many cases, date back to the middle of the 19th century. This promising amount of data allows for potential reconstruction of very useful long-term climate data. Over the last few decades, developing complete and homogenous climatic datasets has been of considerable interest in the Mediterranean (e.g. Eischeid *et al.*, 2000; Klein Tank *et al.*, 2002; Brunetti *et al.*, 2006; Haylock *et al.*, 2008; Vicente-Serrano *et al.*, 2010; Semenov *et al.*, 2010). The rationale behind all these works was to improve quality, temporal extension and spatial coverage of climatic time series. Nonetheless, most of these works were mainly devoted to precipitation (Romero *et al.*, 1998; Eischeid *et al.*, 2000; Brunetti

* Correspondence to: A. El Kenawy, Instituto Pirenaico de Ecología-CSIC, Campus de Aula Dei, Zaragoza 50059, Spain.
E-mail: kenawy@ipe.csic.es

et al., 2006; Vicente-Serrano *et al.*, 2010) and to a less extent to temperature (Eischeid *et al.*, 2000; Brunet *et al.*, 2006).

In many regional, continental and global datasets (e.g. Gleason, 2002; Klein Tank *et al.*, 2002; Hofstra *et al.*, 2009), Spain had a very few number of temperature observatories. Details and sources about different available datasets are listed in Table I. Recently, Brunet *et al.* (2006) developed a new daily adjusted dataset of 22 observatories of maximum and minimum temperatures in the whole Spain. Although this dataset covers a long period (1850–2003), the coverage is still inadequate for several spatial studies. Overall, the insufficient number of temperature stations and their uneven spatial distribution has been the main feature of many recent studies focusing on air temperature change and variability in the Iberian Peninsula (Morales *et al.*, 2005; Brunet *et al.*, 2006, 2007), Owing to this, there is urgent need to develop temperature datasets with high spatial and temporal resolution which can contribute to a better understanding of temperature variability and changes in the Iberian Peninsula, the Mediterranean and southwestern Europe. Moreover, it can be a very useful tool in a wide variety of applications including, for example, characterisation of extreme events (e.g. hot and cold spells), climate risk assessment (e.g. frost), hydrological and

environmental modelling, and verification of numerical model simulations.

This paper describes in totality, the protocol for developing a complete, reliable and homogenous daily temperature dataset for northeastern Spain. The main goal of this study is to improve the spatial and temporal coverage of temperature time series in the study domain using a very dense network of 1583 raw time series. The next goal is to assess and compare the impact of the protocol that has been used to adjust breaks on trends, spatial and temporal coherence, and statistical properties of the final series. The assessment is based on the application of different statistical methods (e.g. semivariance models, L-moment statistics and Pearson correlation) to draw a comparison between the series before and after adjustments. The paper is structured into four main parts. The data sources and a detailed description of the methodology employed to check data quality, infill gaps and test homogeneity of daily temperature time series are outlined in Section 1. An assessment of the possible impact of data adjustment (correction) on spatial and temporal characteristics of the final temperature time series is included in Section 2. Results are presented in Section 3. Discussion of the results and conclusions are given in Section 4.

Table I. List of sources and description of selected datasets covering northeastern Spain.

Reference	Variables	Temporal resolution	Region	Type	Time period
The Global Historical Climatology Network version2 (GHCN)(Vose <i>et al.</i> , 1992)	Precipitation, temperature, sea-level pressure and air pressure	monthly	Globe	land-based	to 1950 near present
The Global daily Climatology Network version 1, US National Climate Data Centre, NCDC, (Gleason, 2002)	Precipitation and temperature	daily	Globe	land-based	2001–1840
Herrera <i>et al.</i> , 2010	Precipitation	daily	Spain (0.2° resolution)	Gridded dataset	1950–2003
Haylock <i>et al.</i> , 2008	Precipitation and temperature	daily	Europe (different resolutions)	Gridded dataset	1950–2006
ECA&D (Klein Tank <i>et al.</i> 2002 and Klok & Tank, 2009)	Precipitation and temperature	daily	Europe	Land-based	1900–2006
ELPIS(Semenov <i>et al.</i> 2010)	Precipitation. Temperature and solar radiation	daily	Europe (25 km)	Gridded dataset	1982–2008
E-OBS (Hofstra <i>et al.</i> , 2009)	Precipitation and temperature	daily	Europe (0.25° resolution)	Gridded dataset	1950–2006
EMULATE (Moberg <i>et al.</i> , 2006)	Precipitation and temperature	daily	Europe	Land-based	to 1850 near present
SDATS (Brunet <i>et al.</i> , 2006)	Temperature	daily	Spain	Land-based	1850–2003
NESATv2 (northeastern Spain Adjusted Temperature) (http://www.urv.cat/centres/Departaments/geografia/clima/archive.htm)	Temperature	monthly	Catalonia	Gridded dataset	1869–1998

Table II. Summary statistics of the flagged data following the quality control procedure. The numbers refer to the fraction of the flagged data in daily maximum (Tmax) and minimum (Tmin) temperature series. The mean indicates the average of the flagged data for the whole dataset, while the lowest (highest) shows the maximum (minimum) percentage of flagged data at the station-based level.

		Winter	Spring	Summer	Fall	Annual
Tmax	Lowest	0.00	0.00	0.00	0.00	0.00
	Highest	0.47	0.97	0.55	0.60	1.99
	Mean	0.01	0.04	0.02	0.02	0.09
Tmin	Lowest	0.00	0.00	0.00	0.00	0.00
	Highest	0.52	1.41	1.35	0.99	3.85
	Mean	0.02	0.02	0.04	0.03	0.11

2. Data and methods

2.1. Study area and dataset description

This study employs a database of 1583 daily maximum and minimum temperature time series spanning some selected periods between 1900 and 2006. This dataset was provided by the Spanish Meteorological Agency (Agencia Estatal de Meteorología, AEM). As shown in Figure 1, this dataset is distributed across an 18-province region located in northeast Spain. It covers a spatial domain of 159424 km² between the latitudes 39°43'N and 43°29'N, and the longitudes 05°01'W and 03°17'E. The study area is heterogeneous in terms of terrain complexity, with elevation up to 3000 m a.s.l. (the Pyrenees). Similarly, the climate is very diverse with clear wet/dry and oceanic/continental contrasts. The network of stations is shown in Figure 1. As depicted, the spatial density of stations is inhomogeneous across the administrative provinces. The most data-rich province is Barcelona in the east, with a density of 1 station/482 km². In contrast, the coarser coverage is found in Soria in the west, with a density of 1 station/3430 km². Temporarily,

the stations' density showed a sharp increase from 1970 onwards (Figure 2). About 21.3% of stations extended back to 1950 or earlier, while 98.7% of the observatories have their most recent records until 2002. In general, this dataset suffers from outliers, missing values and breaks with changes in stations opening, closing and their locations being the main reasons for this.

In the following section, a full description of the multistep approach used for quality control, reconstruction and homogenisation of the daily temperature series is detailed. The different steps are also summarized in Figure 3.

2.2. Quality control of data series

Quality control is a fundamental task to remove incorrect data and to check for data consistency (Feng *et al.*, 2004). In this work, the original series were subjected to several quality checks. First, typical tests were performed to identify systematic errors, which resulted from different sources (e.g. archiving, transcription, and digitisation). This can include nonexistent dates, $T_{min} \geq T_{max}$, $T_{max} > 50^\circ\text{C}$, $T_{min} < -50^\circ\text{C}$ and runs of at least 7 consecutive days with identical records. Next, the data were screened for internal consistency by comparing the value in question against other values in the same time series following Reek *et al.* (1992). Finally, the data were checked for external consistency by comparing each time series with nearby sites. The rationale behind this procedure is to trim outliers that markedly differ from the majority of neighbours while retaining the valuable extreme information. To accomplish this, daily data of each testable observatory were compared with a minimum of 3 nearby observatories located within a distance of 30 km. More specifically, the rank of a value in each testable time series is obtained as a quantile of all values in this series. Next, these quantiles were compared with the quantiles of nearby stations for each particular day. Values which exceeded a defined

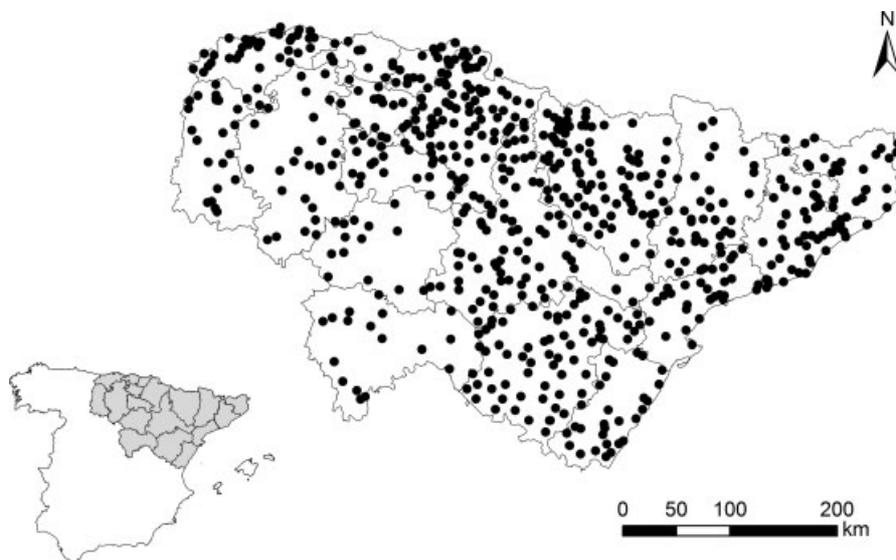


Figure 1. Location of the study area and spatial distribution of temperature observatories in the original dataset ($N = 1583$).

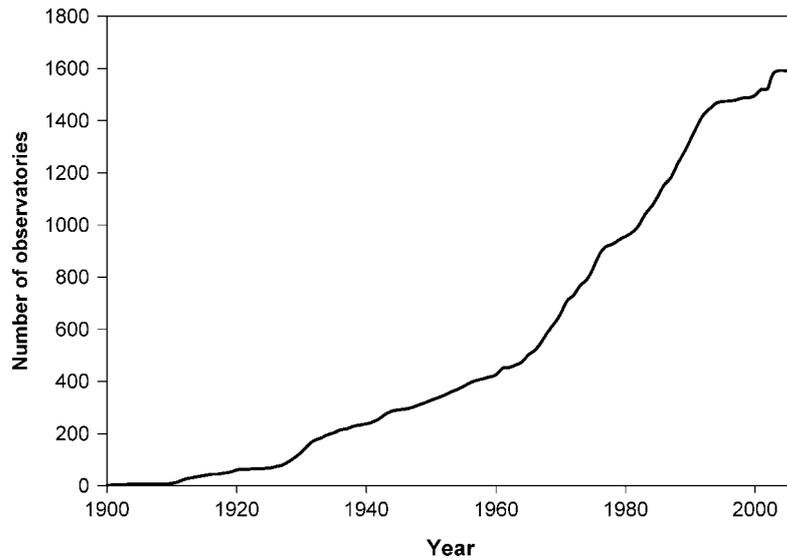


Figure 2. Temporal evolution of the number of available daily maximum–minimum temperature observatories in the original dataset from 1900 to 2006.

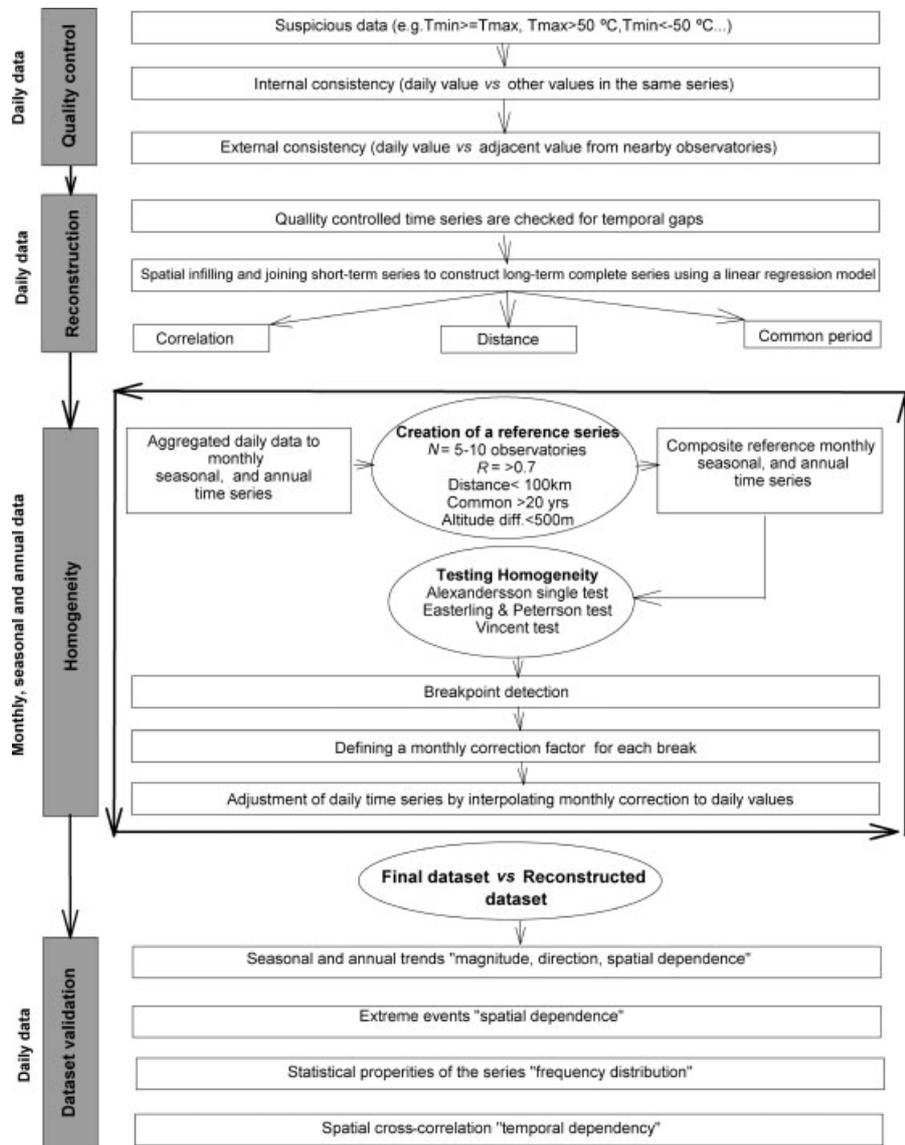


Figure 3. Schematic representation of the integrated approach applied to quality control, reconstruction and homogenization of daily temperature time series in the study area. Shaded areas indicate the main steps.

difference threshold in this between-station comparison were flagged and set to missing values. In order to screen for appropriate interquartiles range threshold, we randomly tested different thresholds (e.g. 0.1, 0.2, 0.3, etc.) using a subset of 1% of the time series. Following the results (not shown), a threshold of 0.5 was proved as a more suitable compromise for all seasons apart from winter, whereby a more restricted threshold (0.3) was set. We did so principally to account for the thermal and dynamical conditions originating from topography influences in winter. In areas of complex terrain, the spatial dependency among observatories can markedly degrade over short distances. A similar approach has recently been adopted by Stepanek *et al.* (2009) and Vicente-Serrano *et al.* (2010).

2.3. Reconstruction of data series

Short-term and fragmented series may introduce noise to estimates of long-term climate changes. This is the typical case in many climatic datasets worldwide, where many gaps are introduced in the series. Peterson and Vose (1997) indicated that short or fragmented series may significantly alter the magnitude and sign of climate variability trends, particularly at regional-scale resolution. For this reason, many methods have recently been applied for serial reconstruction in climatology including, for example, the neural artificial network (Rigol *et al.*, 2001), interpolation algorithms (Vicente-Serrano *et al.*, 2010), and the Singular Spectrum Analysis (Ghil *et al.*, 2002). The basic idea behind these methods is to obtain long-term complete datasets using all available data within a region. This can be carried out as a function of the weighted distance (Shen *et al.*, 2001) and/or correlation (Young, 1992). Nevertheless, these approaches differ in their applicability according to terrain complexity and spatial density of stations. For instance, many interpolation techniques (e.g. optimal interpolation, Kriging and splines-surface fitting) can give good results in regions with no gradients (Jarvis and Stuart, 2001a, 2001b). Contrarily, their performance in terms of both bias and amount of variance is predominantly poor in areas of complex terrain or uneven density of observatories (Hubbard and You, 2005). In such environments, the regression-based methods can be a superior solution. These methods take into account data from adjoining stations with the best temporal correlation and the highest spatial dependence. In addition, they show less sensitivity to outliers (Romero *et al.*, 1998), and can, therefore, better represent systematic temperature differences associated with topographic influences in heterogeneous regions (Hubbard and You, 2005). Numerous previous studies proved the advantages of the regression methods to reconstructing temperature series with respect to other methods (Hubbard and You, 2005; You *et al.*, 2008).

In this research, we have taken advantage of the large number of available observatories in the original dataset to rebuild long series following a standard linear regression model. It is assumed that information introduced

from nearby stations identically reflects uniform climatic conditions of the target station. To accomplish this task, the daily temperature series of the original dataset were firstly divided in two broad groups. The first category included all the series that are still in operation (2004 or more recently) and labelled as target series. This step resulted in 668 target observatories. The remaining data (915 observatories) served as a pool for the reconstruction process. Since the final output of the reconstruction procedure is largely influenced by the selected neighbouring stations, a set of similarity measures (distance and correlation) is used to rebuild the target series. The procedure for the preselection of the adequate neighbours only comprises stations which fulfill a set of restricted multi-criteria, including: (1) A minimum common observing period of 4 years shared by the target station, (2) A positive correlation coefficient exceeding 0.90, and (3) Location within a radius of 15 km around the target station. This procedure helps avoid suspicious information and only sacrifice quality-assured data that experience similar climate conditions to the target observatory. In order to maximize the series length, priority of infilling a target station was first given to the longest nearby series that best correlates with the target station. For this reason, the stations were first ranked on the basis of the available record length and data completeness. When data from an adjacent series were used to reconstruct a target station, they were automatically discarded from the bulk series to avoid introduction of redundant information. However, it is worthwhile indicating that the regression model was applied iteratively to maximize the value of information used from the nearby observatories. Nevertheless, the number of surrounding stations for each candidate station was not fixed in time depending on data availability on each particular day.

2.4. Homogeneity testing

Most long-term climatological time series have been affected by a number of non-climatic factors that make these data unrepresentative of the actual climate variations. These factors include changes in observatory locations, observers, surrounding environment, measurement practices or instruments (Peterson *et al.*, 1998; Costa and Soares, 2009). Historically, less attention has been devoted towards testing homogeneity of climatic data at daily and sub-daily resolution (Vincent *et al.*, 2002; Wijngaard *et al.*, 2003). This task is challengeable due to more noise and high variability introduced in daily data compared with monthly and annual data. A comprehensive review of different homogenisation approaches is given by Peterson *et al.* (1998) and Costa and Soares (2009). In the absence of accurate, complete and well-documented metadata (station history), the relative tests are favoured to identify supporting evidence of significant changes in observational routines. Our dataset shows a typical example. The relative homogeneity test procedure includes two main steps: the creation of a reliable reference series and the application of proper homogeneity tests.

Building reference time series is mandatory for undertaking the relative homogeneity tests. The reference time series is assumed to reveal the same variability of weather and climate exhibited in the candidate station. In this regard, we used composite reference series to test homogeneity of each candidate series. This is mainly because temperature shows less spatial variations compared with other climatic variables, such as precipitation, which can vary greatly over space. Similar to previous works (e.g. Peterson and Easterling, 1994; Klein Tank *et al.*, 2002; Vicente-Serrano *et al.*, 2010), this work brings together factors of distance, correlation, elevation dependence and stations' density in selecting the most useful sites to create reliable composite reference series. More specifically, the reference series were composited from nearby localities lying within a maximum distance of 100 km with correlation greater than 0.7. In practice, a distance threshold of 100 km was considered as high enough so as to secure enough number of nearby stations in sparse network areas (e.g. southwestern portions). In the same sense, although a higher correlation coefficient is desirable for building a reliable reference series, a threshold of 0.7 between the candidate and the reference series still suits our purpose, given that the correlation was computed between the first difference series ($X_{t+1} - X_t$) to avoid additional inhomogeneities from the reference series. Moreover, this threshold is still within the range being considered in previous studies (Peterson *et al.*, 1998; Costa and Soares, 2006; Vicente-Serrano and Lopez Moreno, 2006; Stepanek *et al.*, 2008; Savic *et al.*, 2010). In addition to distance and correlation thresholds, a minimum overlap period of 20 years of records between the candidate and nearby observatories was also set. In fact, this common period is long enough to construct reliable composite reference series. Furthermore, it helps overcome the problem of discontinuities in the nearby series. Also, a maximum altitude difference of 500 m between the candidate and the nearby sites was adopted to limit the influence of topographical gradients.

On the basis of the aforementioned criteria, the reference series were created using a weighted average of data from the highly correlated neighbouring stations following the standard procedure of Peterson and Easterling (1994). However, values of the selected nearby stations were first standardized by their mean and standard deviation for each month independently. This procedure reduces possible biases from the reference series. In general, the procedure to create the reference series was run automatically using the PROCLIM software (Stepanek, 2008). Herein, it is worthwhile to indicate that although our main target was to obtain a homogenous dataset at daily resolution, the reference series were created on monthly, seasonal and annual scales. This helps avoid problems emanating from the high variability of daily data on small spatial scales. Indeed, the daily data may comprise nonlinear events resulting from very local processes such as orography and surface albedo, which are difficult to capture in the daily reference series.

Detecting breakpoints in the time series is one of the most relevant problems among those addressed by the relative homogeneity tests. In practice, there is no one single test to be recommended as optimal to detect breaks in all situations (Costa and Soares, 2009). Thus, application of different statistical methods could largely improve the degree of certainty related to the detectable breaks in the series, particularly when metadata is unavailable (Wijngaard *et al.*, 2003). This approach has been favoured by many researchers to verify the possible discontinuities in climatic time series (Wijngaard *et al.*, 2003; Costa and Soares, 2009). In this research, three well-established relative tests were chosen to test homogeneity of the series, including: the SNHT for a single break (Alexandersson, 1986), the Easterling and Peterson test (Easterling and Peterson, 1995) and the Vincent method (Vincent, 1998). The SNHT has increasingly been recommended to define the most significant break in climatic series (Slonosky *et al.*, 1999; Wijngaard *et al.*, 2003). On the other hand, the Easterling and Peterson two-phase regression-based technique has the advantage of detecting multiple breaks in the series, specifically when they are very close in time (Easterling and Peterson, 1995). More recently, Vincent (1998) proposed a multiple linear regression approach to identify multiple breaks in temperature series, particularly undocumented change points. Vincent *et al.* (2002) used this test to identify multiple breaks in the daily Canadian temperature series. In our study, all these homogeneity tests were carried out on monthly, seasonal and annual timescales. These temporal resolutions fulfill normality, which is a prerequisite for the SNHT. Moreover, seasonal resolution allows a better detection of inhomogeneities, given that some sources of inhomogeneity show different impacts from season to season (e.g. summer *vs* winter). The AnClim software developed by Stepanek (2004) was used for testing homogeneity. Only detectable breaks at a confidence level of 95% ($p < 0.05$) were considered statistically significant. Following this approach, all possible inhomogeneities in the series can finally be determined and grouped together. However, given that the underlying hypothesis of the relative homogeneity tests is based on comparing the mean from two sides of the data, inhomogeneities found within the first and the last five years of the series were not considered. As a consequence of the reduced sample size close to the boundaries of the series, there is an increasing probability to have high T values (Hawkins, 1977). Previous studies reject inhomogeneities of five years or more at the start and end of the series, particularly if they are not fully explained by metadata (Gonzalez-Rouco *et al.*, 2000; Gokturk, *et al.*, 2008).

The homogeneity adjustments make it possible to consider temporal variations of temperature time series as caused only by climate processes. Therefore, an adjustment (correction) model was applied to account for statistically significant abrupt changes. The correction factor was computed for each month individually based on a comparison of percentiles of the differences between the candidate and the reference series before and after any detected break. To obtain daily corrections, the

predefined monthly corrections were interpolated into the daily time series following the approach described by Sheng and Zwiers (1998) and recommended by Vincent *et al.* (2002). This procedure is advantageous in various ways. First, it reduces discontinuities on the first and last days of each month (Vincent *et al.*, 2002). Second, it maintains characteristics of daily extreme events, such as frequency and intensity. Finally, it preserves temperature trends and variability presented in monthly temperatures. Accordingly, the homogenized daily temperature series will be compatible with those of monthly resolution.

Finally, it is worthwhile indicating that testing homogeneity was applied iteratively where all time series in the dataset were considered repeatedly as candidate and reference series. However, in this iterative procedure, we first used the multiple reference series to test homogeneity as each candidate time series was tested independently and iteratively against each of the best correlated five neighbouring stations. Although this procedure involves intensive use of the data, it reduces the possible effects of temperature spatial variation originating from terrain complexity in the study domain. However, in following iterations, homogeneity was tested again using the composite reference series applying restricted criteria (e.g. $r > 0.9$, number of observatories ≤ 10). This procedure mainly aims to verify the regional consistency among nearby sites. Overall, combining both multiple and composite reference series and testing all the time series several times helps to ensure that the final dataset is relatively free from any significant breaks.

2.5. Impacts of the adjustment on trends and statistical properties of the series

Evaluation of the impact of series adjustment is of great importance to determining the reliability of the final time series for different climate analyses such as trends and climate extremes. Recalling that homogenisation procedures are more subjective and, therefore, can have adverse and complex influences on time series, many authors (e.g. Peterson *et al.*, 1998; Vincent *et al.*, 2002) highlighted the importance of identifying the possible impacts of inhomogeneity adjustment on climatic data. In this research, we used a set of techniques to account for these possible impacts. Our main hypothesis is that the degree to which a station's data reveals spatial and temporal coherence with their immediate surroundings can be a good indication of the reliability of the applied methodology. Therefore, the validity of the final dataset can be evaluated by examining spatial and temporal characteristics of the new dataset, as compared with the original dataset. All the assessment tests were applied to a set of 98 observatories covering the period from 1950 to 2006.

First, trends of maximum and minimum temperatures were calculated on seasonal and annual scales before and after eliminating inhomogeneities. The main purpose was to assess the relative influence that the homogeneity procedure exerts on the trend assessment. The

significance of the trends was assessed using the non-parametric Spearman (*Rho*) test at a significance level of 95% (p value < 0.05). This test is robust to outliers and does not assume prior probability distribution of the residuals. The slope was estimated using the ordinary least squares (OLS) fitting and expressed in °C per decade. Seasonal averages were obtained from monthly data for each year and defined as winter (December–February), spring (March–May), summer (June–August), and autumn (September–November). A comparison between the trends sign (direction) before and after corrections was conducted by means of the cross-tabulation analysis, which illustrates pairwise relationships between the categories of the trend signs (i.e. significant positive (+), significant negative (−), and statistically insignificant (N)). In this context, the pivot tables were constructed to represent the cross-categorized frequency data in a matrix format following the results of the trend assessment.

To assess the degree of spatial dependence between the seasonal and annual trends, the semivariance models were computed for the magnitude of change. Semivariance analysis has widely been applied in ecology (Urban *et al.*, 2000) and climatology (Vicente-Serrano *et al.*, 2010) to detect scales of variability in spatial data. The semivariance describes the spatial variance as a function of distance between the observatories. The value of the semivariance decreases as the pair of data points is separated by a short distance, and it increases when the distance increases. A detailed explanation of semivariance analysis is given in Webster and Oliver (2001). In this study, this geostatistic is employed to analyze variations in spatial structure of the trends before and after homogeneity correction. It is expected that the semivariance will be lower after correcting inhomogeneities as a consequence of introducing a relatively free-of-error and homogeneous series. Spatial semivariance was obtained, as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i) - Z(s_i - h)]^2 \quad (1)$$

where $\gamma(h)$ is semivariance at lag distance h , $N(h)$ is the number of station pairs that are separated by h distance, $Z(s_i)$ and $Z(s_{i+h})$ are the values of Z variable in stations s_i and s_{i+h} .

Secondly, adjustment of the series may alter the probability distribution of extreme temperature (e.g. frequency and intensity). Therefore, we used two extreme temperature indices to assess possible impacts of adjustment on extreme values. In practice, the annual count of warm and cold days before and after adjustment was computed for each time series with respect to the period from 1950 to 2006. The warm (cold) day was defined as the day higher (lower) than the 90th (10th) percentile of the average maximum (minimum) temperature. The definition of extreme values based on the percentiles is objective, site-independent and facilitates direct comparisons between regions of diverse climates (Jones *et al.*, 1999). However,

it is worthwhile indicating that the key question behind this analysis was to assess the way in which the adjustment procedure affects spatial continuity of the trends rather than the trends themselves. Therefore, the spatial continuity of the extreme temperature trends before and after homogenisation was compared using semivariance of the magnitude of change.

Thirdly, to determine how the adjustment procedure affects the statistical distribution of daily temperature values in the series, we analysed some statistical indicators of the time series before and after correction. In this regard, L-moment statistics were computed for each independent daily maximum and minimum time series before and after correction. These statistics provide information on the scale (L-coefficient of variance (τ_2)), shape (L-coefficient of skewness (τ_3)), and peakedness of the series (L-coefficient of kurtosis (τ_4)). L-moment statistics are independent of the sampling size and are also resistant to outliers. Moreover, they show less bias in comparison with other conventional product moments, such as standard deviation, skewness and kurtosis. These statistics have been used in climatological and hydrological studies (e.g. Guttman, 1993; Guttman *et al.* 1993; Vicente Serrano *et al.*, 2010). More details on the L-moment theory are given by Asquith (2003) and the calculation procedure can be found in Hosking (1990) and Hosking and Wallis (1997). To get a picture of the spatial structure of L-moment coefficients, values of these statistics were also plotted against distance by means of the semivariance models before and after inhomogeneities correction.

Lastly, while the statistical methodologies described above can help provide a guidance to assess the reliability of the final dataset, perhaps the most outstanding feature in assessing the effect of the homogeneity procedure is to account for the correlation between the time series before and after correcting breaks. In this research, an assessment of this effect was achieved by deriving the Pearson correlation matrix between temperature series at different distance orders. The correlation matrix was computed for both the original series (raw and corrected) and the series of first differences ($X_{t+1} - X_t$) (raw and corrected). Considering the correlation between the first difference series is important to give insights into the strength of the temporal dependency among the series after isolating the influence of inhomogeneities introduced in the independent series. Overall, introducing inhomogeneities in the series is assumed to degrade correlation among nearby sites and, therefore, the inter-station correlation could act as a kind of filter for unadjusted series.

3. Results

3.1. The final dataset

Table II summarizes the percentage of flagged data following the quality control procedure. On average, the percentage of erroneous data was higher for minimum (0.11%) than for maximum temperature (0.09%). The

highest errors for maximum and minimum temperatures occurred during spring and summer, respectively. Spatially, the fraction of flagged data was greater in station-dense areas (e.g. Catalonia and Cantabria), compared with relatively sparse areas (e.g. Burgos and La Rioja).

After quality control checks, a linear regression model was used to compile the nearby and best correlated series of short-term span in order to rebuild long-term time series. Figure 4 illustrates an example of the reconstruction procedure. As presented in Figure 4(a), the wintertime daily minimum temperature in the observatory of Calatorao Cooperative (9428E, Zaragoza) has records from 1940 to 2006 as a consequence of joining data from two nearby sites (9425I and 9432). The temporal evolution of the final reconstructed series fits well that of the adjacent localities (9425I: $r = 0.92$, 9432: $r = 0.93$). This clearly implies that our procedure can provide a robust basis for the series reconstruction. The quality controlled and reconstructed daily dataset of 668 maximum–minimum temperature series covering some periods between 1900 and 2006 was then subjected to homogeneity testing and adjustment.

Figure 5(a) illustrates the temporal evolution of the number of available series in the newly compiled dataset. The temporal coverage of the new dataset has clearly been improved in terms of series completeness. The number of complete and homogenous time series that dates back to 1920, 1950, 1960 and 1970 is 19, 98, 128 and 189 respectively. The spatial distribution of the final dataset is given in Figure 5(b). As illustrated, the spatial density of the series is satisfactory across much of the study area (e.g. Catalonia and Lerida in the east, and Cantabria and Vizcaya in the northwest), with the exception of the southern and south-central portions (e.g. Soria and Guadalupe). Vertically, the observatories are mainly located in the elevation range of 6–1920 m a.s.l., with the majority of them (70.1%) being placed at elevations below 800 m a.s.l. As illustrated in Figure 5(b), the distribution of observatories across much of the mountainous areas (e.g. the Pyrenees and the Iberian system) is irregular since only 14.4% of observatories are located above 1000 m. This uneven distribution is expected given that higher elevations are not adequately covered by the data in the original dataset.

3.2. Impacts of the adjustment procedure on trends and statistical properties of the series

In the following section, an evaluation of the reliability of the newly developed dataset is presented to assess the sensitivity of temperature series in terms of their statistical properties (e.g. means, variance, extremes and statistical distribution) to the adjustment procedure.

3.2.1. Homogeneity results

One of the most common problems in handling long-term climate data is the presence of inhomogeneities. Box plots summarising the criteria used to build the composite reference series to define inhomogeneities in the

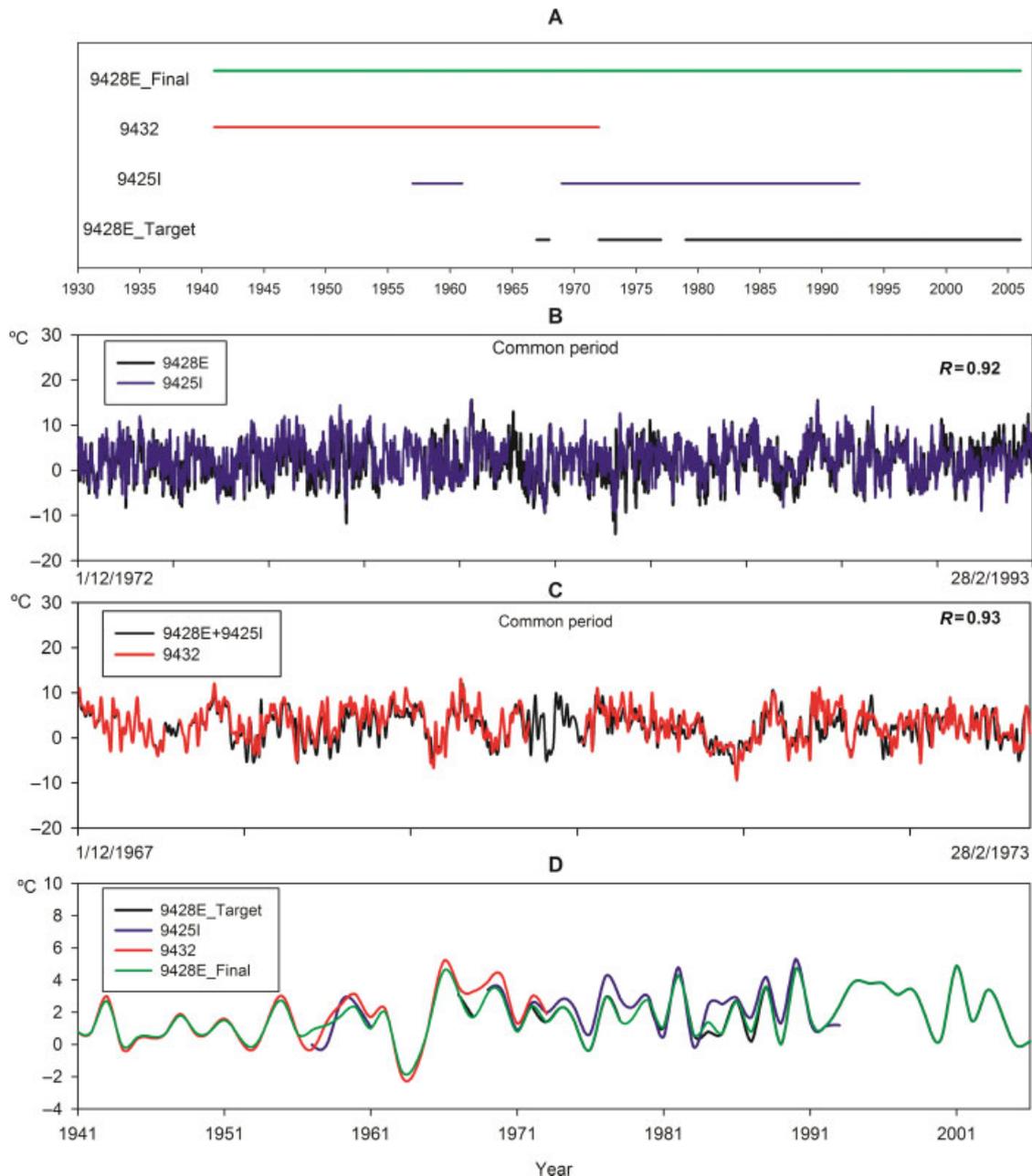


Figure 4. The reconstructed wintertime minimum temperature series of the observatory of Calatorao Cooperative (9428E, Zaragoza) based on joining data from two nearby sites (9425I and 9432). Pearson coefficient (r) indicates the correlation between the reconstructed time series and each of the nearby observatories. The green line denotes the final time series. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

series are illustrated in Figure 6. On average, values of the 10 highest correlated neighbouring stations with average distance under than 50 km were considered to create the reference series for both maximum and minimum temperature series. More importantly, the Pearson correlation coefficient was generally higher than 0.90 for a vast majority of observatories, with only very few exceptions. Taken together, these findings comply with the common standards required to build reliable reference series in homogeneity studies.

Table III summarizes the homogeneity testing results for maximum and minimum temperature series. The results demonstrate that 307 (46%) and 302 (45.2%)

of minimum and maximum temperature time series, respectively, did not show inhomogeneities. However, 82% (79%) of the minimum (maximum) temperature series classified as homogenous were relatively short (<30 years). A majority of the series longer than 30 years suffered from some homogeneity problems. For instance, maximum and minimum temperature time series of more than 40 (50) years of records had an average of 2.7 (3) and 2.3 (2.6) breakpoints, respectively. Table III also indicates that the number of inhomogeneities was larger for maximum temperature ($n = 969$) than for minimum temperature ($n = 865$). This can largely be attributed to the high spatial correlation among maximum temperature

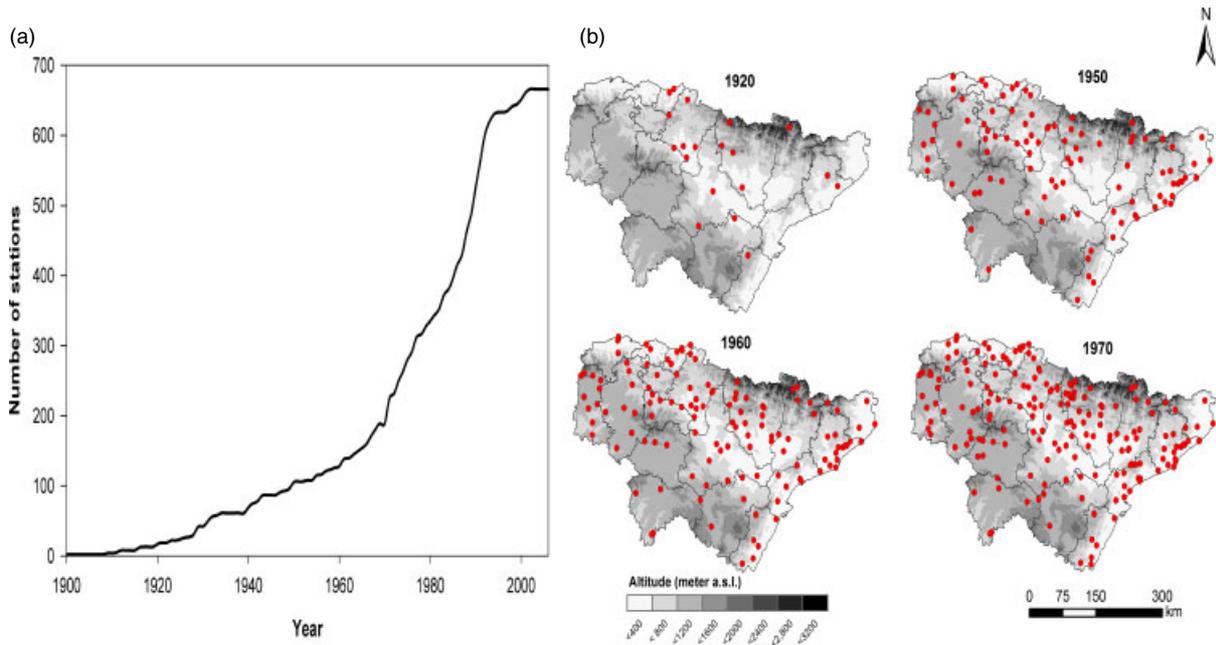


Figure 5. (a) The number of the final complete and homogenous temperature series for the period from 1900 to 2006; and (b) their spatial distribution. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

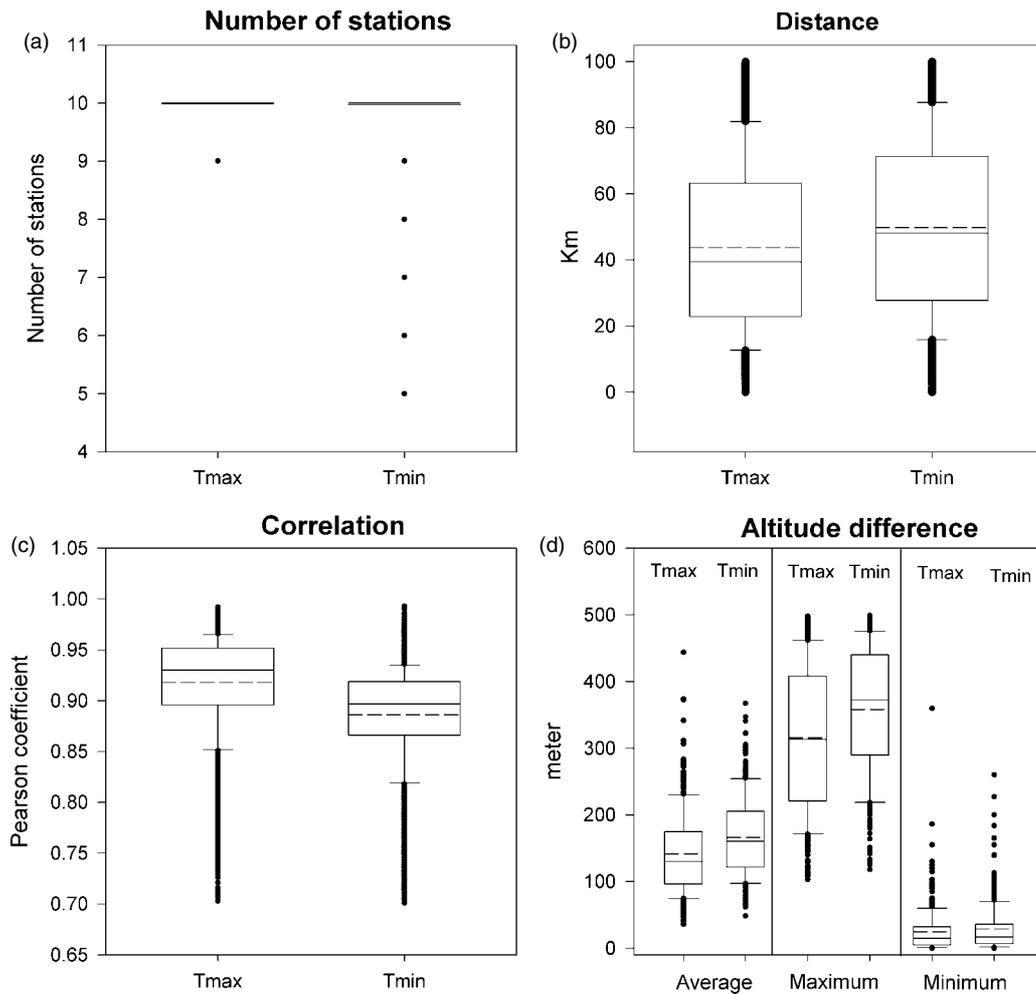


Figure 6. Box plots summarising (a) the number of nearby observatories (b) distance threshold, (c) Pearson correlation threshold and, (d) altitude difference of the series used to create the composite reference series for break detection. Dotted line indicates the mean. The median, 10th, 25th, 75th and the 90th percentiles are plotted as vertical boxes with errors bar.

Table III. Summary of the homogeneity testing results (significance level = 95%).

Minimum temperature	
Number of testable series	651
Number of non-testable series	17
Total number of the series	668
Number of homogenous series	307
Number of series with significant breaks	344
Total number of significant breaks	865
Maximum temperature	
Number of testable series	656
Number of non-testable series	12
Total number of the series	668
Number of homogenous series	302
Number of series with significant breaks	354
Total number of significant breaks	969

time series, which makes it easier to detect even small shifts in the series. Additionally, the impact of station relocations, as a main cause of inhomogeneities, is expected to be minimal during nighttime due to low vegetation activity (Stepanek and Mikulová, 2008). The reconstruction procedure only accounted for 10.3 and 10.4% of the detectable inhomogeneities in maximum and minimum temperature, respectively. These breaks often occur when joining very short-term fragments from different time series. Temporarily, the majority of the breaks dated back to the 1970s and 1980s for both maximum and minimum temperatures, which agrees well with other earlier studies (e.g. Tuomenvirta, 2001; Wijngaard *et al.*, 2003; Auer *et al.*, 2005). Spatially, inhomogeneities were randomly distributed for both maximum and minimum temperature suggesting that these breaks can be attributed to

changes in instruments or observation practices. Nonetheless, it is noteworthy that most of the strong inhomogeneities in terms of the amount of change were mainly localized in regions with complex terrain (e.g. the Pyrenees and the Iberian system). This is reasonable given that stations in higher elevated regions are more vulnerable to changes in locations and/or surrounding environment (Tuomenvirta, 2001).

A comparison between the performances of the three relative homogeneity tests reveals that the SNHT was markedly efficient in detecting the most significant break in the series. Although the gradual breaks were not determined using the single-shift SNHT, the Easterling and Peterson method and the Vincent method were proven to have more power in defining such breaks. Overall, the statistically significant breaks ($p < 0.05$) defined by the three tests were combined for each particular observatory. Then, a monthly correction model was applied to account for each detectable break. Figure 7 shows the frequency distribution of the correction factors applied to monthly maximum and minimum temperature datasets. The correction factor did not vary greatly from one season to another. However, as revealed in Figure 7, there was a bias to decrease the temperature means after correction. The average value of the correction factor was generally higher for minimum temperature than for maximum temperature. The decrease in the means of maximum temperature largely occurred during hot seasons: summer (-0.04°C) and spring (-0.05°C). In all seasons, a correction factor higher than 1°C was needed only in very few series. Figures 8 and 9 illustrate the results of the SNHT applied to two different time series. The results corresponding to summer maximum temperature in the observatory of Fuenterrabia aeropuerto (Guipuzoca) are

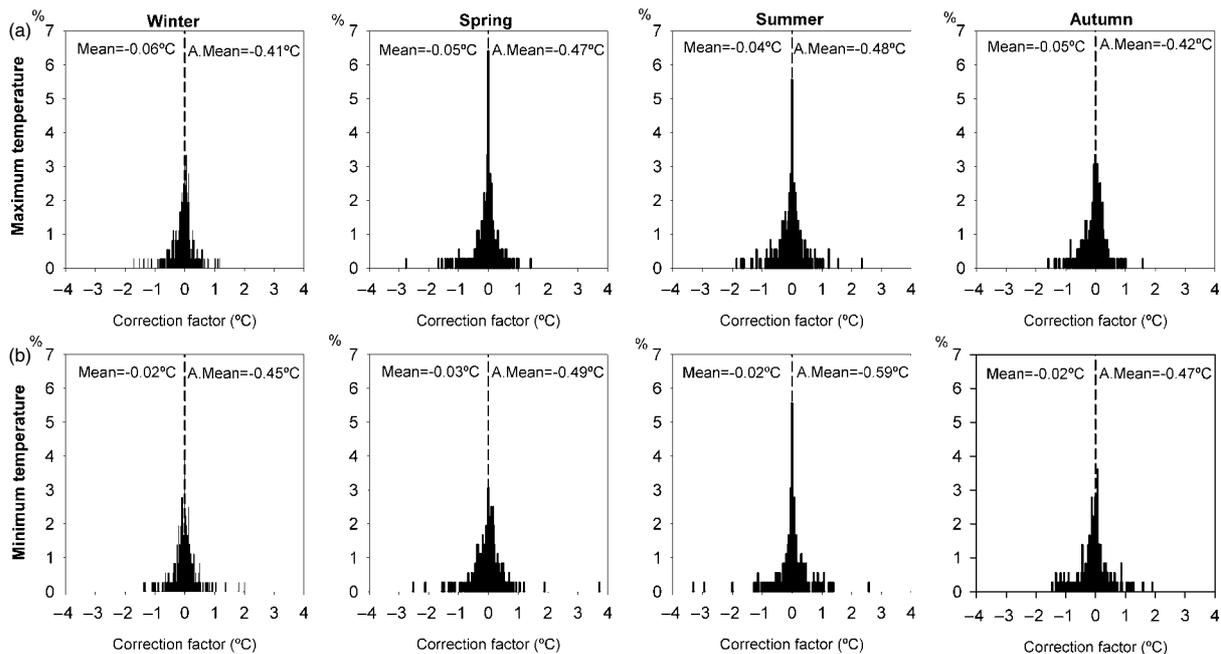


Figure 7. Frequency distribution of the correction factors ($^\circ\text{C}$) applied to (a) maximum, and (b) minimum temperature series averaged for each season. The values of the mean (absolute mean) of the correction factors averaged for all time series are also provided.

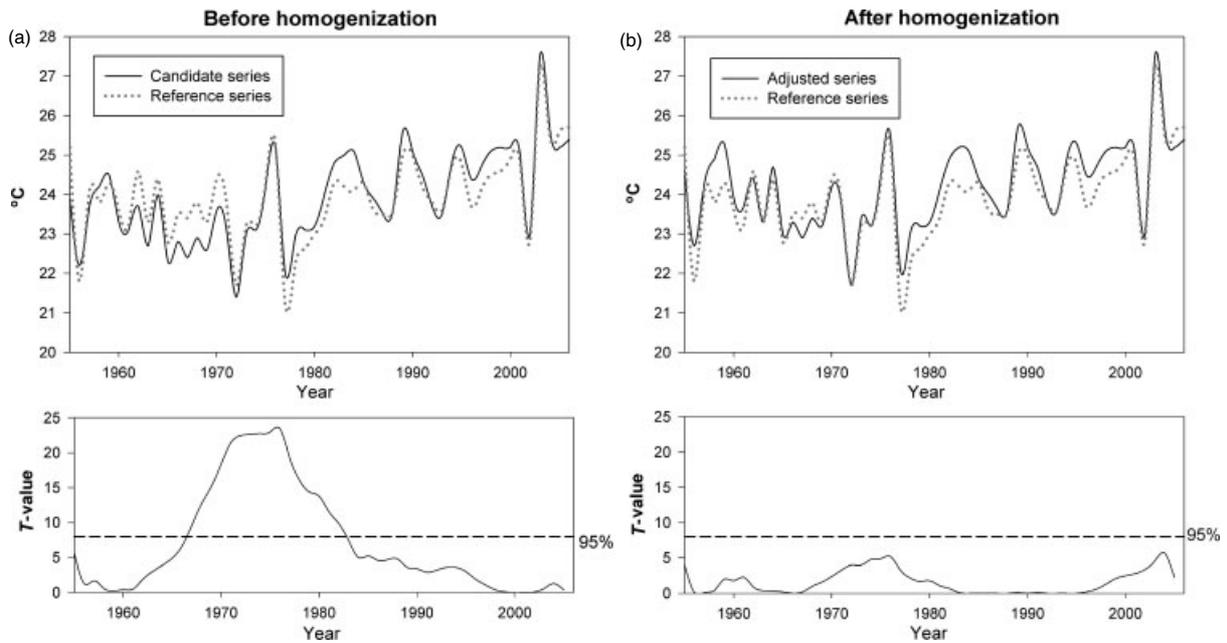


Figure 8. Test results of the SNHT applied to summer maximum temperature series in Fuenterrabia aeropuerto, (Guipuzcoa) (a) before and (b) after homogeneity corrections. Dashed lines indicate the 95% significant level. The test statistic (T) is plotted against the critical value.

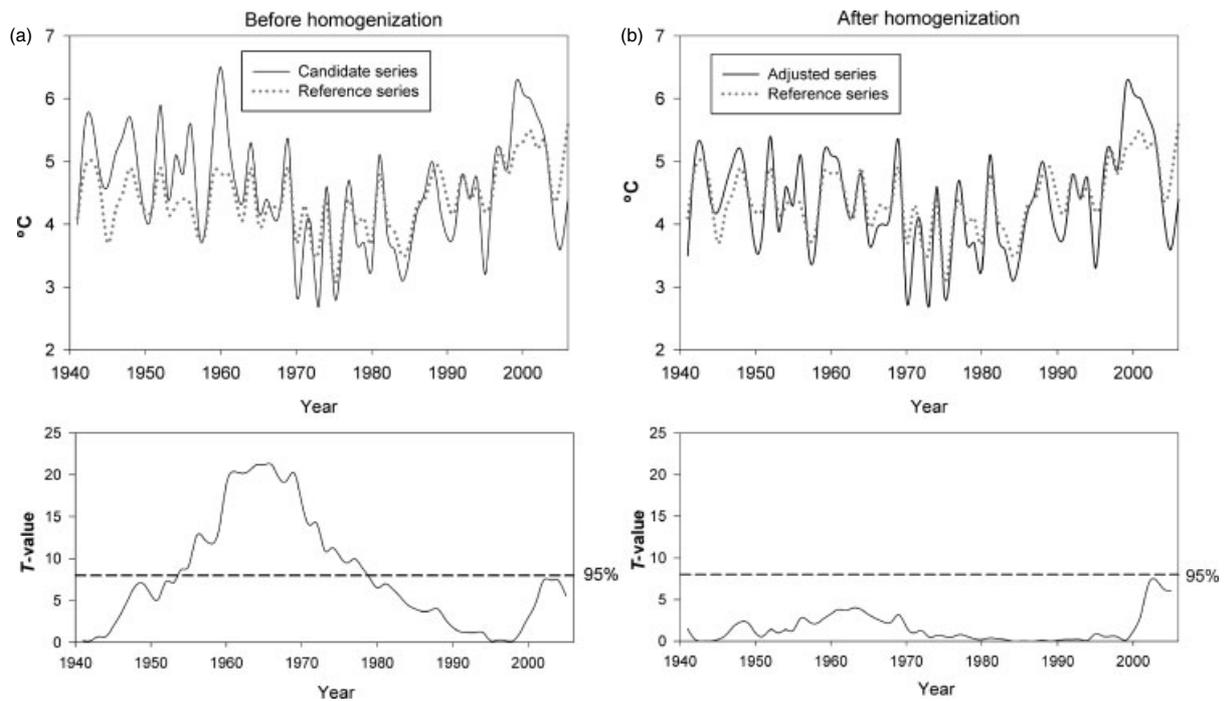


Figure 9. Test results of the SNHT applied to spring minimum temperature series in Els Hostalets de Balenya (Barcelona) (a) before and (b) after homogeneity corrections. Dashed lines indicate the 95% significant level. The test statistic (T) is plotted against the critical value.

given in Figure 8. As shown, T statistic reached its critical value (95% significance level) around the year 1968. Nonetheless, the displacement from the mean disappeared after eliminating the detected inhomogeneity. Another example is presented in Figure 9, which corresponds to spring minimum temperature in Els Hostalets de Balenya (Barcelona). As illustrated, a statistically significant break is presented close to the year 1956. The differences between the candidate and the reference series and their

t -test results indicate that T -value for the adjusted data was below the confidence limit.

3.2.2. Impact of the homogeneity on trends

Table IV presents the results of the cross-tabulation analysis applied to trends in the annual maximum and minimum temperature for each pair of stations (i.e. before and after homogeneity correction). In general, the results did not reflect considerable differences in the sign (direction)

Table IV. Results of the cross-tabulation analysis applied to trends in annual maximum and minimum temperature series before and after homogenization. Significance is assessed at the 95% level. Numbers between brackets indicate the fraction of observatories.

<i>After homogenization</i>	Before homogenization					
	<i>Maximum temperature</i>			<i>Minimum temperature</i>		
	<i>Positive</i>	<i>Negative</i>	<i>Insignificant</i>	<i>Positive</i>	<i>Negative</i>	<i>Insignificant</i>
<i>Positive</i>	63 (64.3%)	2 (2%)	18 (18.4%)	49 (50%)	5 (5.1%)	36 (36.8%)
<i>Negative</i>	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Insignificant</i>	5 (5.1%)	0 (0%)	10 (10.2%)	1 (1%)	2 (2%)	5 (5.1%)
<i>Total</i>		98			98	

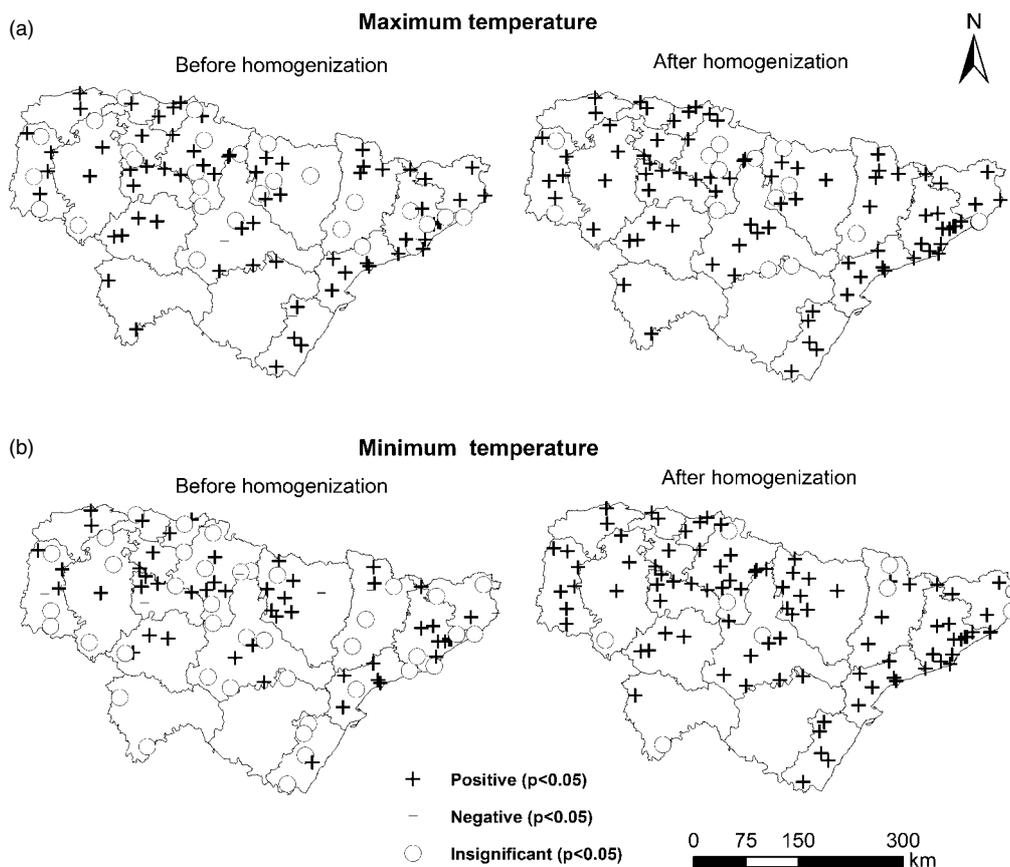


Figure 10. Spatial distribution of the trends in the annual (a) maximum, and (b) minimum temperature time series before and after homogeneity corrections. Trend calculation is based on the period 1950–2006, and statistical significance is assessed at the 95% level.

of the trends. The oppositely directed trends were only evident in 13 (13.26%) and 35 (35.71%) of maximum and minimum temperature series, respectively. This clearly implies that in most cases the adjustment had no discernible effect on the direction of the gained trends. Perhaps this arises from the small correction factors applied to the majority of observatories on the monthly scale, which in turn disappear when aggregated to the annual timescale. Another possible explanation can be linked to the fact that most of the detectable inhomogeneities were found in the 1970s and 1980s, a period which exhibited a remarkable warming trend worldwide (Jones and Moberg, 2003). Under this warming, the magnitude of the correction failed to alter the direction of the observed

variability in the series. Spatial differences between the trends in the raw and corrected annual temperature series are illustrated in Figure 10. It seems that the adjustments had a very local impact on trends, whereby observatories located along the Mediterranean coast experienced more warming, particularly for minimum temperature. In fact, the most convincing impact of adjustments on trends is mainly related to changes in the magnitude rather than the direction of the trend. Figure 11 depicts scatter plots showing the differences in linear trends between the original and adjusted time series. For most seasons, the trends, after correction, were not linearly consistent with those before correction. This suggests a considerable impact of the adjustment procedure on the slope of both seasonal

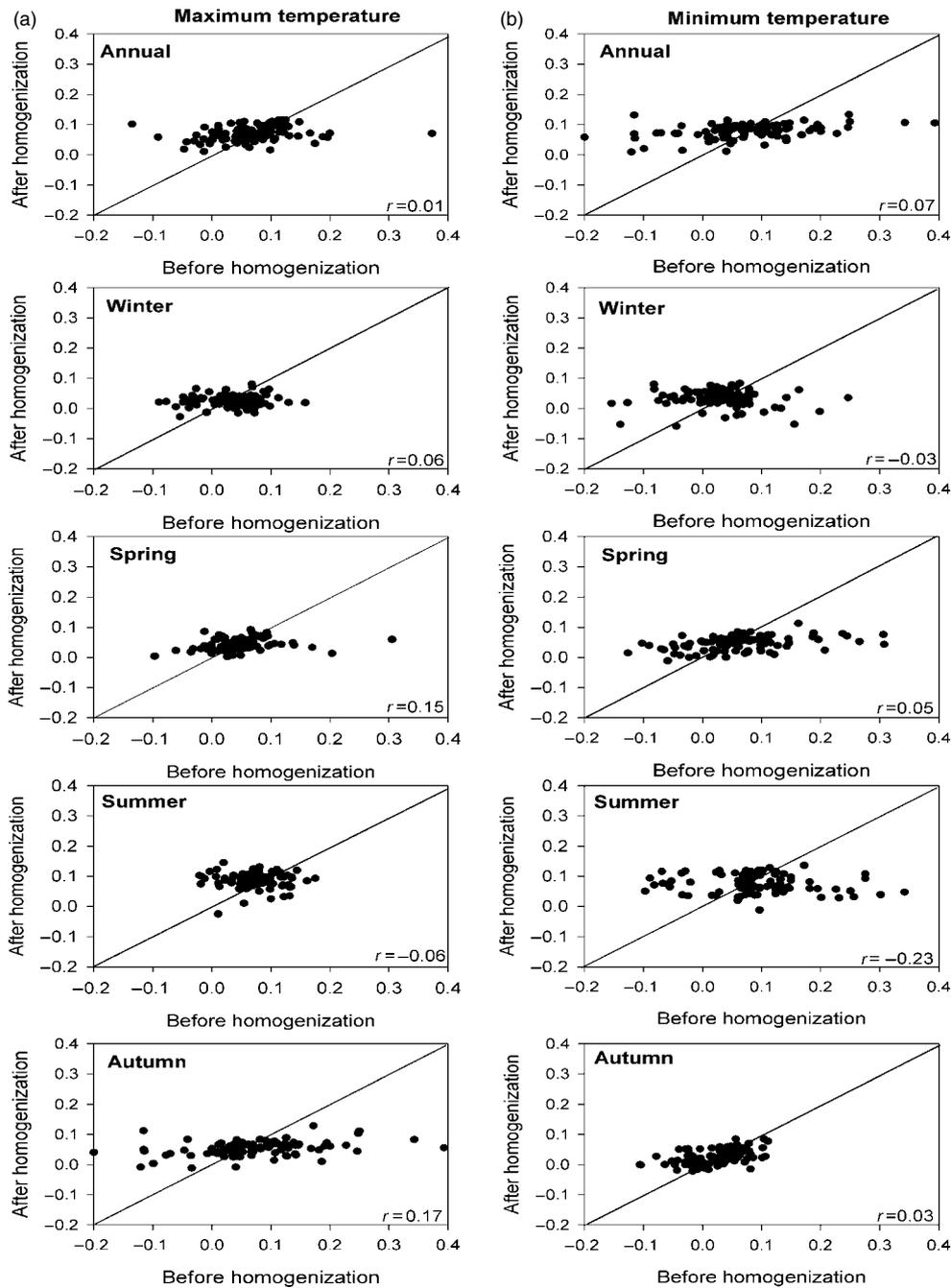


Figure 11. Scatter plots of the magnitude of the trends ($^{\circ}\text{C decade}^{-1}$) as derived from the seasonal and annual trend analysis for (a) maximum, and (b) minimum temperature before and after homogeneity corrections for the period from 1950 to 2006.

and annual time series. One typical example corresponding to the observatory of Yerba de Basa (Huesca) is shown in Figure 12. The linear fit of the raw and homogenized series indicates that both series showed uptrends over the period from 1950 to 2006. Nevertheless, the tendency toward warmer conditions was weaker after adjusting the series ($0.1^{\circ}\text{C decade}^{-1}$), compared with the series prior to correction ($0.3^{\circ}\text{C decade}^{-1}$). In the same sense, Figure 13 strongly suggests that the temporal behaviour of the adjusted series as predicted by the semivariance models was spatially more dependent compared with the raw series. This higher spatial continuity was markedly apparent in all seasons for both maximum and minimum

temperature series. Given that eliminating inhomogeneities likely reduces signal-to-noise ratio in the time series, it can thus be expected that the homogeneity adjustment is accounting for the improvement in the spatial continuity of the trends after correction.

3.2.3. Impact of the homogeneity on extreme events

The homogeneity procedure applied in this work may affect extreme temperatures in different ways like frequency, intensity and persistence. Figure 14 compares the semivariance models of the trends in hot and cold days indices before and after correction. A quick visual inspection of the semivariance models indicates that the

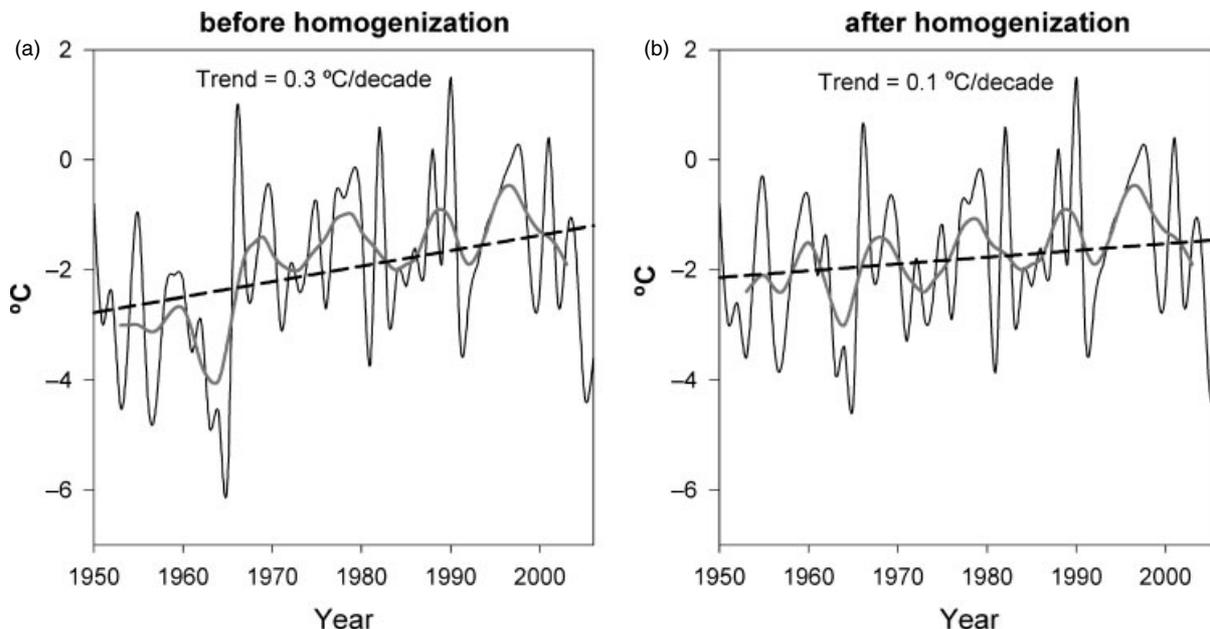


Figure 12. Trends in wintertime minimum temperature in the observatory of Yebra de Basa (Huesca) (1950–2006) (a) before, and (b) after homogeneity corrections. Grey line corresponds to a low-pass filter of 9 years.

spatial heterogeneity of the series markedly decreased after applying the homogeneity correction. Trends in extreme events did show high levels of spatial consistency for the adjusted series, as neighbouring observatories tended to have more identical patterns. In contrast, the raw series exhibited certain abrupt jumps over small distances. This implies that the variance of the raw data is higher, suggesting a weaker spatial component. By contrast, the high spatial continuity of the corrected data simply indicates that the spatial coherence among observatories is predominantly attributed to similar temporal evolution at short distances. It is worth indicating that the influence of adjustments on spatial dependency of cold days' climatology was less apparent compared with hot days (Figure 14). This can primarily be linked to the fact that minimum temperature during wintertime shows higher spatial variability compared with other seasons as a consequence of the joint effect of strong circulation influences and high topography-induced thermal contrasts. Overall, the high degree of spatial coherence in temperature extremes after correcting inhomogeneities can provide a strong guidance on the reliability of the adjusted dataset.

3.2.4. Impact of the homogeneity on statistical properties of the series

Figure 15 shows the relationships between L-moment statistics (i.e. variance, skewness and kurtosis) before and after homogenisation. A comparison of the parent distribution of L-moments permits the assumption that the frequency distribution of temperature series before and after adjustment generally coincided with a clear linear well-fit. This was evident for both maximum and minimum temperature time series, whereby strong and

statistically significant positive correlations were remarkably apparent between the series before and after adjustment ($r > 0.8$). This clearly implies that the statistical attributes of the adjusted series in terms of variance, skewness and kurtosis were almost equal to similar values, as inferred from the series prior to the homogeneity adjustment. The observed bias in the frequency distribution of a few time series is generally negligible. Taking these results together, we can note that the homogenous dataset preserves the same statistical distribution of the original dataset. The same finding was also confirmed by means of the semivariance models as illustrated in Figure 16. The semivariance models of L-moment statistics reveal that the statistical properties of the adjusted series were rather similar to those of the raw series. In addition, the spatial association of L-moment coefficients still suggests a decrease in the spatial association between the series as distance increases. Distant observatories are likely to have larger variance which in turn points toward heterogeneous spatial patterns. This gives good indication on spatial dependence of the statistical attributes of the series after correction.

3.2.5. Impact of the homogeneity on inter-station correlation

Inter-station correlation is one of the most important features that can describe the improvement in the temporal dependency of the dataset after eliminating inhomogeneities. Figure 17 depicts the inter-station correlation before and after correcting inhomogeneities. It clearly reveals higher inter-station correlations for the adjusted series with relevance to the raw data for both maximum and minimum temperature. This was particularly the case for correlation matrices calculated for the original as well as the first difference series. As Peterson and Easterling

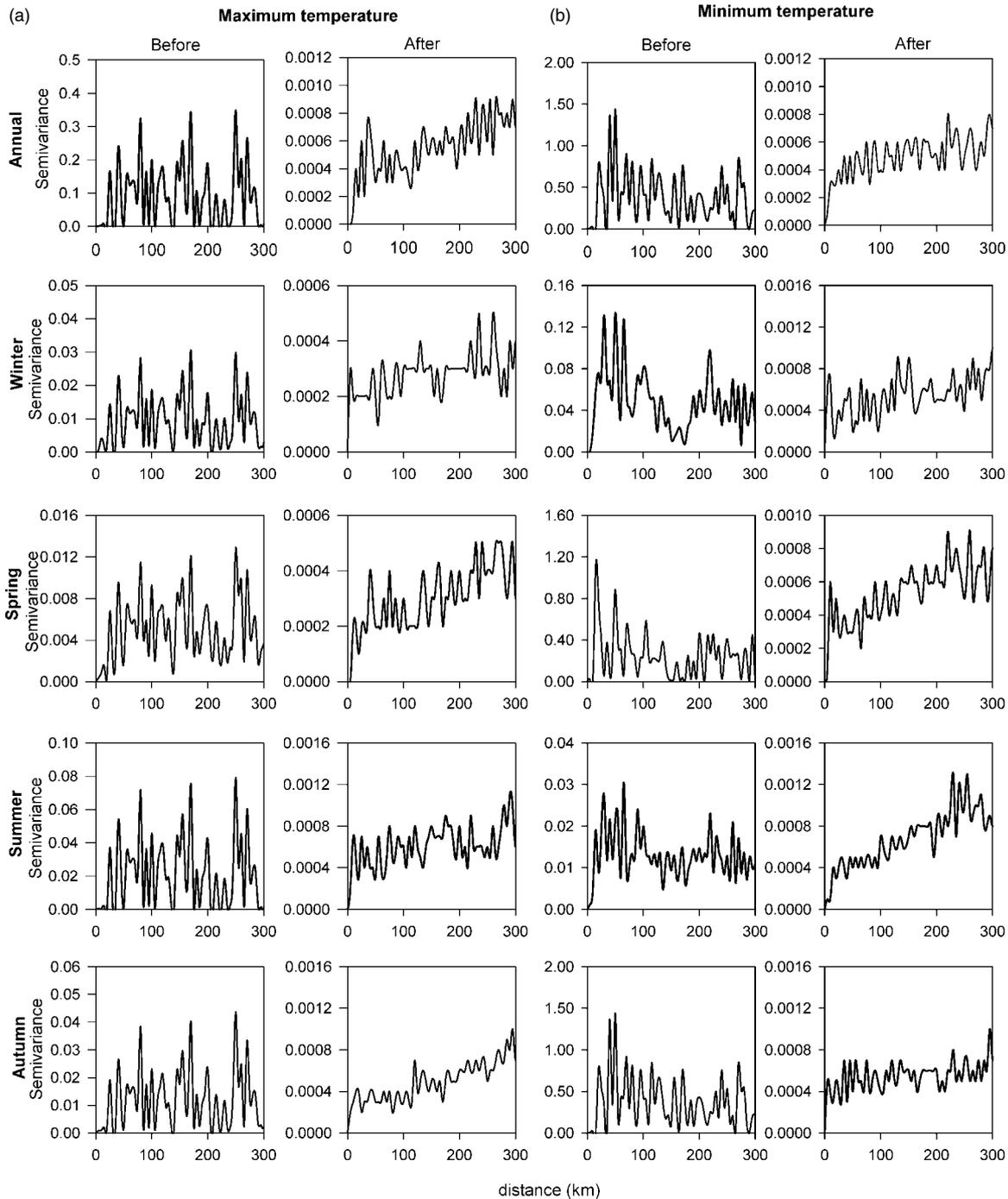


Figure 13. Semivariance of the magnitude of the annual and seasonal trends for (a) maximum, and (b) minimum temperature before and after homogeneity corrections.

(1994) demonstrated that correlation coefficients among the series are very sensitive to the presence of inhomogeneities, the inter-station correlation was improved as a consequence of the removal of the noise from the series after adjusting the detectable breaks. Figure 17 also indicates that the inter-station correlation was higher among nearby locations, while it gradually decreased over greater distances. This demonstrates that the homogenized dataset is spatially dependent, with more similar temporal variability among nearby sites. This finding implies that the new compiled dataset is more robust in

capturing the regional variability of temperature across the study region.

4. Discussion and conclusions

In this work, a dense daily temperature database spanning the period between 1900 and 2006 has been developed for northeastern Spain. The main focus of this study was to employ all available information in the original dataset (1583 raw series) provided by *the Spanish Meteorological Agency* to build a spatially

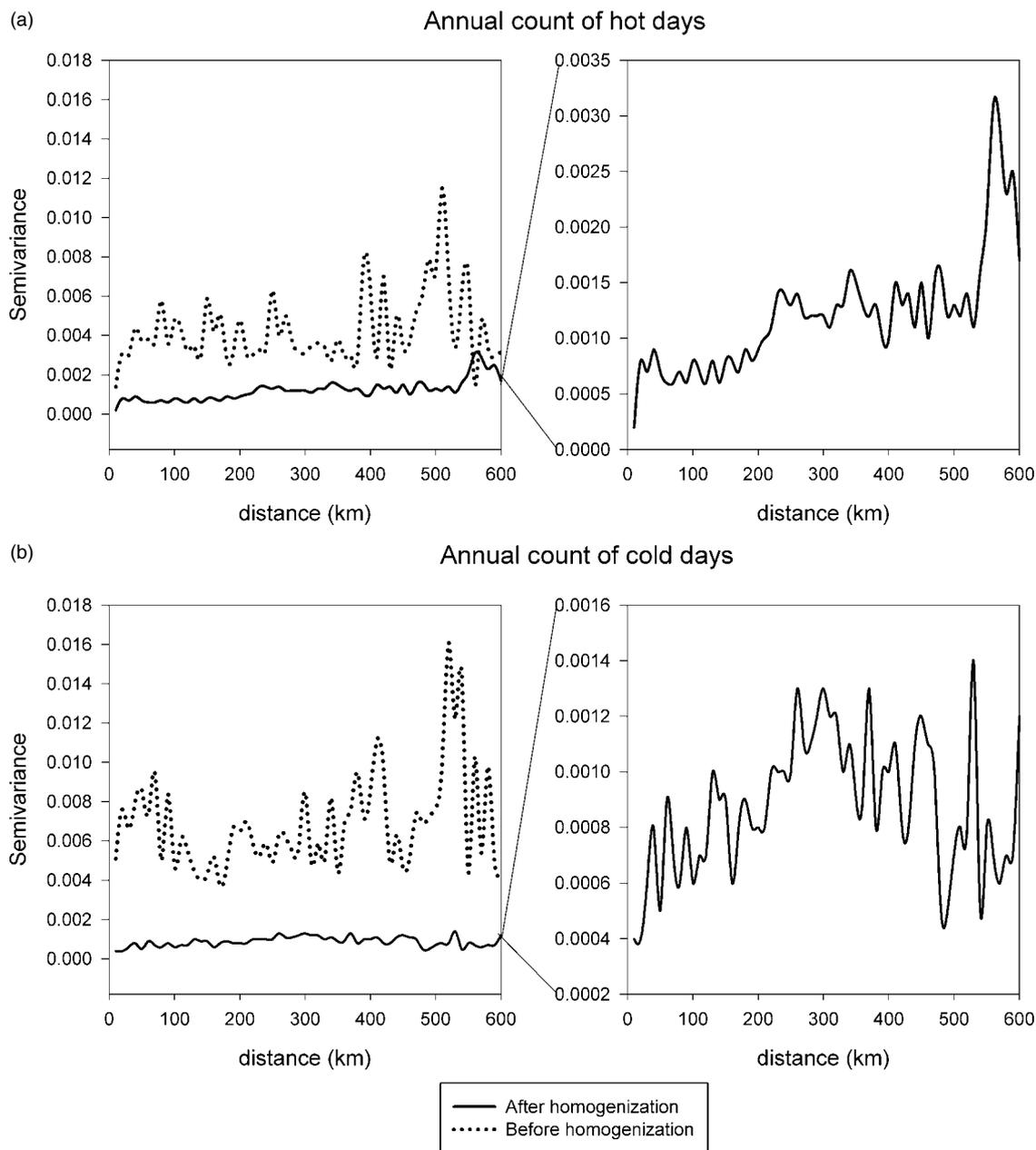


Figure 14. Semivariance of the magnitude of trends in the annual count of (a) hot, and (b) cold days calculated for the 57-year time series (1950–2006) before and after homogeneity corrections.

and temporarily high-resolution temperature dataset. To accomplish this task, the original dataset was first quality controlled to remove anomalous and suspicious data. This procedure proved to be sensitive to trim outliers and kept only valuable extremes. A linear regression model was then undertaken to infill gaps in daily temperature series using information captured from the surrounding observatories. This regression model is simple, straightforward, applicable and robust when dealing with extreme values. Furthermore, it accounted for the dependency between temperature and topography gradients, particularly in areas of complex terrain. To account for possible breakpoints in the reconstructed time series, three well-established relative tests were used to test homogeneity of the series on monthly, seasonal and annual timescales.

A monthly correction factor was calculated and interpolated to daily values to eliminate inhomogeneities from the daily series. A combination of the results of three homogeneity tests was advantageous because these tests had different sensitivities to defining discontinuities in the time series. Moreover, this approach helped determine not only strong breakpoints in temperature time series, but also small shifts.

In this work, a great deal of effort being put into developing ways to assess how the homogeneity methodology affected spatial and temporal characteristics of the final time series. For this reason, we provided a suite of statistical tests for screening of different aspects in which the break correction can affect temperature series. By means of the spatial semivariance statistic and L-moment

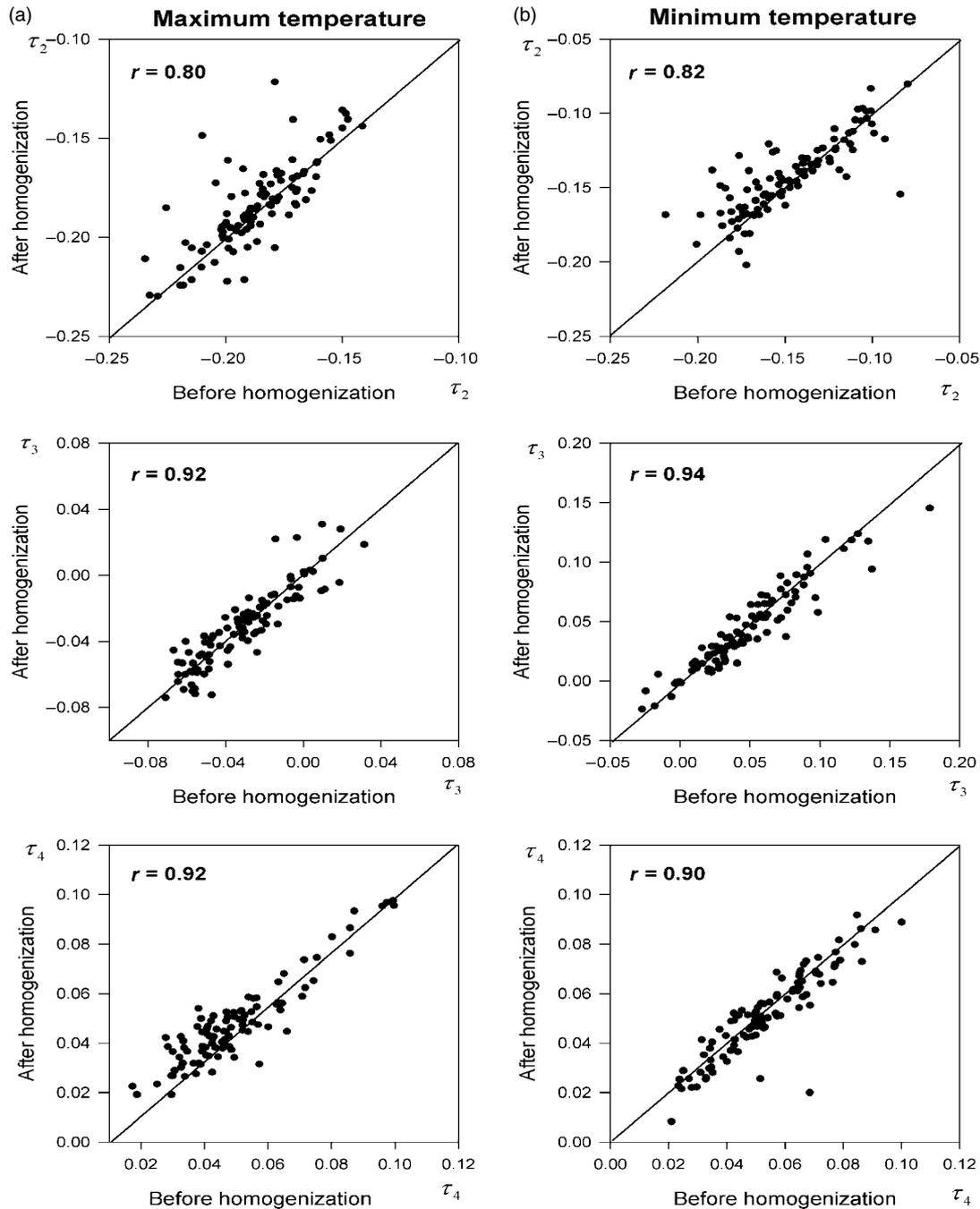


Figure 15. L-moment coefficients of (a) maximum, and (b) minimum temperature series before and after homogeneity corrections.

statistics, it was possible to look at a broad array of time series characteristics (e.g. means, extreme values and frequency distribution). These methodologies were applied comparatively to the series before and after adjustment to realistically evaluate impacts of the break correction on the final dataset. In practice, the semivariance models were efficient in describing and analyzing spatial structure of temperature means and extremes. L-moment statistics also provided a consistent tool to assess changes in the statistical properties of the series (i.e. variance, kurtosis and skewness). Following these techniques, it was clearly evident that the homogeneity adjustment significantly improved the spatial and temporal structure of

both temperature means and extremes. Given that evaluation of the impact of homogeneity routines on final climatological products has not attracted much attention in the literature, our methodologies can offer a useful approach for the assessment of homogeneity impacts on climate time series. These methodologies are objective, flexible and reproducible, and can therefore be applied in similar environments.

The newly compiled database comprises the most complete and homogenous time series of maximum and minimum temperature for northeastern Spain. From the spatial and temporal perspectives, this dataset is unique and represents an improvement when compared with previous

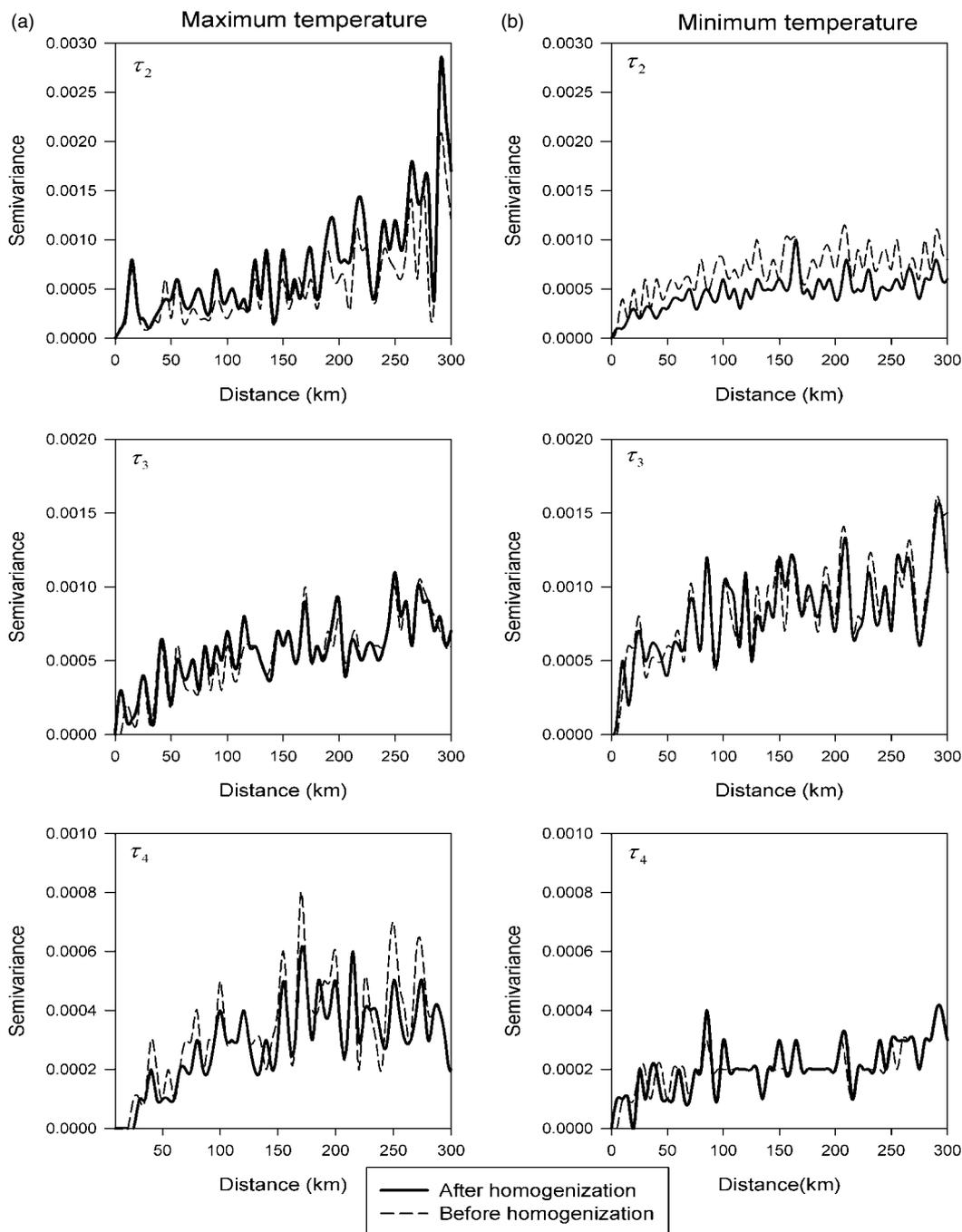


Figure 16. Semivariance of L-moment coefficients of (a) maximum, and (b) minimum temperature series before and after homogeneity corrections.

datasets available for air temperature in the Iberian Peninsula. Therefore, this dataset can contribute to better understanding of space–time variability of temperature and its driving causes on both local and regional scales. Moreover, this dataset can be useful in understanding the causes and impacts of local and regional climate changes on hydrological systems, ecosystems, natural resources and human activities. This feature is of great importance given the complex topography and diverse climates of the study domain. Additionally, this climatology can enhance the grid resolution of any climatic study in future with more potential to validate climate simulations from Regional Climate Models (RCMs).

Acknowledgements

We are indebted to the anonymous reviewers for their constructive comments which were most helpful in improving this paper. We would like to thank the Agencia Estatal de Meteorologia for providing the temperature data used in this study. This work has been supported by the research projects CGL2006-11619/HID, CGL2008-01189/BTE, CGL2011-27574-CO2-02, CGL2011-27753-CO2-01 and CGL2011-27536 financed by the Spanish Commission of Science and Technology; and also FEDER, EUROGEOS (FP7-ENV-2008-1-226487) and ACQWA (FP7-ENV-2007-1- 212250) financed by the

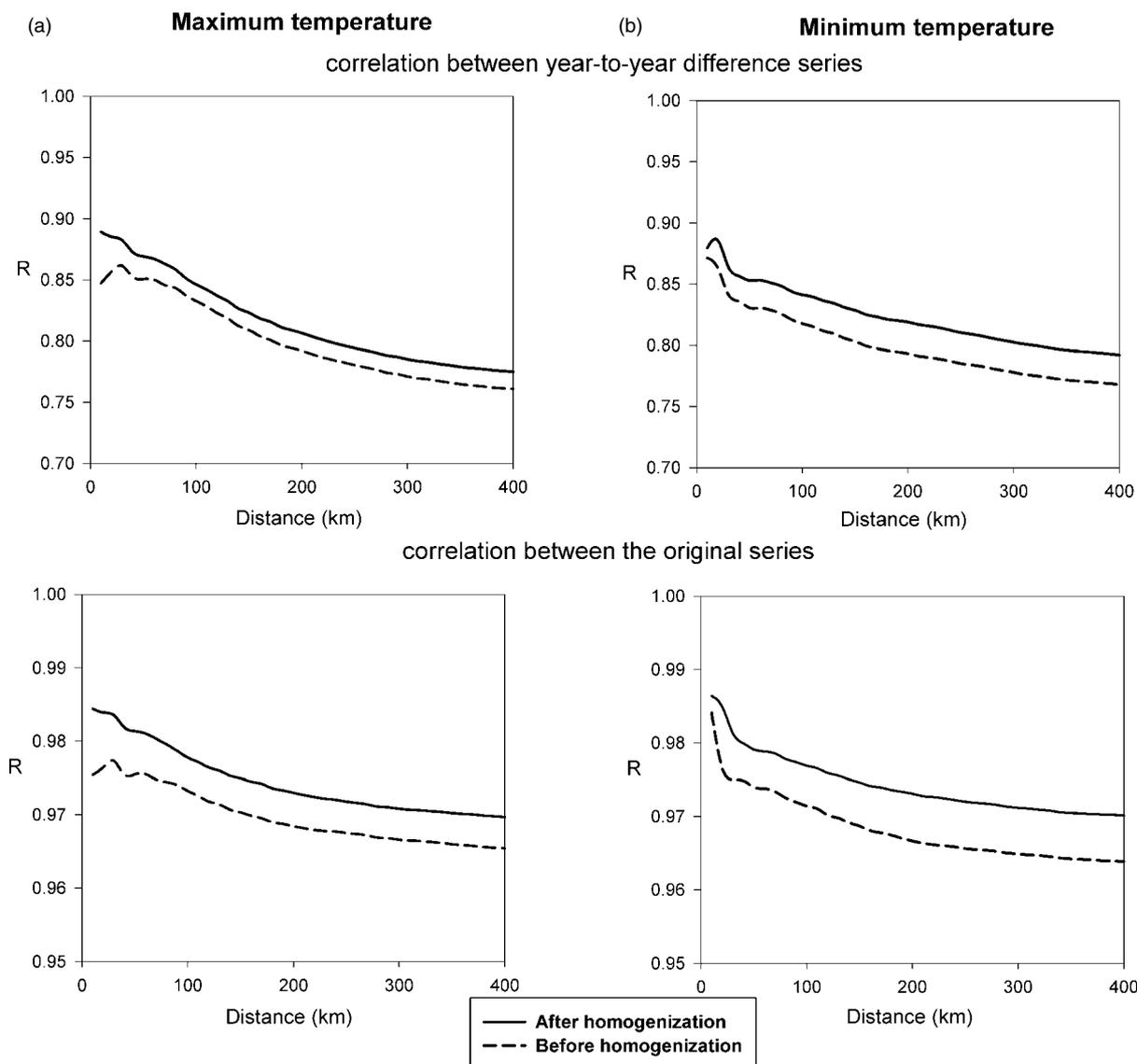


Figure 17. Average Pearson correlation coefficient of (a) maximum, and (b) minimum temperature computed for all station pairs as a function of distance before and after homogeneity corrections. The upper (lower) panel belongs to the first-difference (original) temperature dataset.

VII Framework Programme of the European Commission, La nieve en el Pirineo Aragonés y su respuesta a la variabilidad climática, and Efecto de los escenarios de cambio climático sobre la hidrología superficial y la gestión de embalses del Pirineo Aragonés, financed by Obra Social La Caixa and the Aragón Government and Influencia del cambio climático en el turismo de nieve, CTP01/10, financed by the Comisión de Trabajo de los Pirineos.

References

- Alexandersson H. 1986. A homogeneity test to precipitation data. *International Journal of Climatology* **6**: 661–675.
- Asquith WH. 2003. *Modeling of runoff-producing rainfall hyetographs in Texas using L-moment statistics*, PhD dissertation, University of Texas at Austin, Austin, Texas.
- Auer I, Böhm R, Jurković A, Orlik A, Potzmann R, Schöner W, Ungersböck M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, Moisselin JM, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajić-Čapka M, Zaninović K, Majstorović Z, Szalai S, Szentimery T and Mercalli L. 2005. A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002. *International Journal of Climatology* **25**: 139–166.
- Brunet M, Jones PD, Sigró J, Saladié O, Aguilar E, Moberg A, Della-Marta PM, Lister D, Walther A, Lopez D. 2007. Temporal and spatial temperature variability and change over Spain during 1850–2005. *Journal of Geophysical Research* **112**: D12117, DOI: 10.1029/2006JD008249.
- Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, Moberg A, Lister D, Walther A, Lopez D, Almarza C. 2006. The development of a new dataset of Spanish daily adjusted temperature series SDATS 1850–2003. *International Journal of Climatology* **25**: 1777–1802.
- Brunetti M, Maugeri M, Monti F, Nanni T. 2006. Temperature and precipitation variability in Italy in the last two centuries from homogenized instrumental time series. *International Journal of Climatology* **26**: 345–381.
- Costa AC, Soares A. 2006. Identification of inhomogeneities in precipitation time series using SUR models and the Ellipse test. *In Proceedings of Accuracy 2006–7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Caetano M, Painho M (eds). Instituto Geográfico Português, 419–428.
- Costa AC, Soares A. 2009. Homogenization of climate data: review and new perspectives using Geostatistics. *Mathematical Geosciences* **41**: 291–305.

- De Luis M, González-Hidalgo JC, Longares LA, Štěpánek P. 2009. Seasonal precipitation trends in the Mediterranean Iberian Peninsula in second half of 20th century. *International Journal of Climatology* **29**: 1312–1323.
- Easterling DR, Peterson TC. 1995. A new method of detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369–377.
- Eischeid JK, Pasteris PA, Diaz HF, Plantico MS, Lott NJ. 2000. Creating a serially complete national daily time series of temperature and precipitation for the western United States. *Journal of Applied Meteorology* **39**: 1580–1591.
- Feng S, Hu Q, Qian Q. 2004. Quality control of daily meteorological data in China 1951–2000: a new dataset. *International Journal of Climatology* **24**: 853–870.
- Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi F, Yiou P. 2002. Advanced spectral methods for climatic time series. *Reviews of Geophysics* **40**(1): 3.1–3.41, DOI: 10.1029/2000GR000092.
- Gleason B. 2002. *Global Daily Climatology Network, V1.0*. National Climatic Data Center: Asheville, NC.
- Gokturk OM, Bozkurt D, Sen OL, Karaca M. 2008. Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes* **22**: 3210–3218.
- Gonzalez-Rouco JF, Jimenez JL, Quesada V, Valero F. 2000. Quality control and homogeneity of precipitation data in the Southwest of Europe. *Journal of Climate* **14**: 964–978.
- Guttman NB. 1993. The use of L-Moments in the determination of regional precipitation climates. *Journal of Climatology* **6**: 2309–2325.
- Guttman NB, Hosking JRM, Wallis JR. 1993. Regional precipitation quantile values for the continental United States computed from L-Moments. *Journal of Climatology* **6**: 2326–2340.
- Hawkins DM. 1977. Comments on “A new statistic for testing suspected outliers”. *Communications in Statistics A* **6**: 435–438.
- Haylock MR, Hofstra N, Klein Tank AM, Klok EJ, Jones PD, New M. 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research* **113**: D20119, DOI: 10.1029/2008JD010201.
- Herrera S, Gutierrez JM, Ancell R, Pons MR, Frias MD, Fernandez J. 2010. Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *International Journal of Climatology*. DOI: 10.1002/joc.2256.
- Hofstra N, Haylock M, New M, Jones P. 2009. Testing E-OBS European high-resolution gridded dataset of daily precipitation and surface temperature. *Journal of Geophysical Research* **114**: D21101, DOI: 10.1029/2009JD011799.
- Hosking JRM. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical Society, B Series* **52**: 105–124.
- Hosking JRM, Wallis JR. 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press: Cambridge, UK, 224.
- Hubbard KG, You J. 2005. Sensitivity Analysis of Quality Assurance Using the Spatial Regression Approach: A Case Study of the Maximum/Minimum Air Temperature. *Journal of Atmospheric and Oceanic Technology* **22**(10): 1520–1530.
- Jarvis CH, Stuart N. 2001a. A comparison between strategies for interpolating maximum and minimum daily air temperatures: The selection of ‘guiding’ topographic and land cover variables. *Journal of Applied Meteorology* **40**: 1060–1074.
- Jarvis CH, Stuart N. 2001b. A comparison between strategies for interpolating maximum and minimum daily air temperatures: The interaction between number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology* **40**: 1075–1084.
- Jones PD, Horton EB, Folland CK, Hulme M, Parker DE, Basnett TA. 1999. The use of indices to identify changes in climatic extremes. *Climatic Change* **42**: 131–149.
- Jones PD, Moberg A. 2003. Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of Climate* **16**: 206–223.
- Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier P, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Míetus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, López JA, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P. 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology* **22**: 1441–1453.
- Klok EJ, Tank A. 2009. Updated and extended European dataset of daily climate observations. *International Journal of Climatology* **29**: 1182–1191.
- Moberg A, Jones PD, Lister D, Walther A, Brunet M, Jacobeit J, Alexander LV, Della-Marta PM, Luterbacher J, Yiou P, Chen D, Klein Tank AMG, Saladié O, Sigró J, Aguilar E, Alexandersson H, Almaraz C, Auer I, Barriendos M, Begert M, Bergström H, Böhm R, Butler CJ, Caesar J, Drebs A, Founda D, Gerstengarbe F-W, Micela G, Maugeri M, Österle H, Pandzic K, Petrakis M, Srncic L, Tolasz R, Tuomenvirta H, Werner PC, Linderholm H, Philipp A, Wanner H, Xoplaki E. 2006. Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *Journal of Geophysical Research* **D111**: D22106, DOI: 10.1029/2006JD007103.
- Morales CG, Ortega MT, Labajo JL, Piorno A. 2005. Recent trends and temporal behavior of thermal variables in the region of Castilla-Leon (Spain). *Atmosfera* **18**: 71–90.
- Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrely T, Salinger J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones P and Parker D. 1994. Creation of homogeneous composite climatological reference series. *International Journal of Climatology* **14**: 671–680.
- Peterson TC, Vose RS. 1997. An overview of the global historical climatology network temperature database. *Bulletin of the American Meteorological Society* **78**: 2837–2850.
- Peterson TC and Others. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18**: 1493–1517.
- Reek T, Doty SR, Owen TWA. 1992. Deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bulletin of the American Meteorological Society* **73**: 753–762.
- Rigol JP, Jarvis CH, Stuart N. 2001. Artificial neural networks as a tool for spatial interpolation. *International Journal of Geographical Information Science* **15**(4): 323–344.
- Romero R, Guijarro JA, Ramis C, Alonso S. 1998. A 30-year 1964–1993 daily rainfall data base for the Spanish Mediterranean regions: first exploratory study. *International Journal of Climatology* **18**(10): 541–560.
- Savic S, Petrovic P, Milovanovic B. 2010. Homogenization of Mean Air Temperature Data Series from Serbia. *Geophysical Research Abstracts* (vol. 12), EGU2010-5521-1.
- Semenov MA, Donatelli M, Stratonovitch D, Chatzidaki E, Baruth B. ELPIS. 2010. A dataset of local-scale daily climate scenarios for Europe. *Climate Research* **44**: 3–15.
- Shen SSP, Dzirkowski P, Li G, Griffith D. 2001. Interpolation of 1961–1997 daily climate data onto Alberta polygons of ecodistrict and soil landscape of Canada. *Journal of Applied Meteorology* **40**: 2162–2177.
- Sheng J, Zwiers F. 1998. An improved scheme for time-dependent boundary conditions in atmospheric general circulation models. *Climate Dynamics* **14**: 609–613.
- Slonosky VC, Jones PD, Davies TD. 1999. Homogenization techniques for European monthly mean surface pressure series. *Journal of Climate* **12**: 2658–2672.
- Stepanek P. 2004. *AnClim- software for time series analysis (for windows)*, Department of Geography, Faculty of Natural Sciences, Masaryk University: Brno, 1.47 MB.
- Stepanek P. 2008. ProClimDB – software for processing climatological datasets CHMI regional office Brno: <http://www.climahomeu/ProcData/html>
- Stepanek P, Mikulová K. 2008. Homogenization of air temperature and relative humidity monthly means of individual observation hours in the area of the Czech and Slovak Republic. In *5th Seminar for Homogenization and Quality Control in Climatological Databases*. Hungarian Met. Service: Budapest, Hungary.
- Stepanek P, Zahradnický P, Skalák P. 2009. Data quality control and homogenization of the air temperature and precipitation series in the Czech Republic in the period 1961–2007. *Advances in Science and Research* **3**: 23–26.
- Stepanek PR, Reznickova L, Brazdil R. 2008. Homogenization of daily air pressure and temperature series for Brno (Czech Republic)

- in the period 1848–2005. In *Proceedings of the Fifth seminar for homogenization and quality control in climatological databases* (Budapest, 29 May–2 June 2006). WCDMP, WMO, Geneva, CD-ROM.
- Tuomenvirta H. 2001. Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data. *International Journal of Climatology* **21**: 495–506.
- Urban DL, Miller C, Stephenson NL, Halpin PN. 2000. Forest pattern in Sierran landscapes: the physical template. *Landscape Ecology* **15**: 603–620.
- Vincent L. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11**: 1094–1104.
- Vincent LA, Zhang B, Bonsal R, Hogg WD. 2002. Homogenization of daily temperatures over Canada. *Journal of Climate* **15**: 1322–1334.
- Vicente-Serrano SM, Beguería S, López-Moreno J, Garcia-Vera M, Stepanek P. 2010. A complete daily precipitation database for North-East Spain: reconstruction, quality control and homogeneity. *International Journal of Climatology* **30**: 1146–1163.
- Vicente-Serrano SM, Lopez-Moreno JJ. 2006. The influence of atmospheric circulation at different spatial scales on winter drought variability through a semiarid climatic gradient in north east Spain. *International Journal of Climatology* **26**: 1427–1456.
- Vose RS, Schmoyer RL, Steurer PM, Peterson TC, Heim R, Karl TR, Eischeid J. 1992. The global historical climatology network: long-term monthly temperature precipitation sea level pressure and station pressure data. ORNL/CDIAC-53 NDP-041 Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory: Oak Ridge, TN.
- Webster R, Oliver MA. 2001. *Geostatistics for Environmental scientists*. Wiley: Chichester, UK, 271.
- Wijngaard JB, Klein Tank AMG, Können GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**: 679–692.
- You J, Hubbard KG, Goddard S. 2008. Comparisons of methods for spatially estimating station temperatures in a quality control system. *International Journal of Climatology* **28**: 777–787.
- Young KC. 1992. A three-way model for interpolating monthly precipitation values. *Monsoon Weather Review* **120**: 2561–2569.