For more information, please contact:
World Meteorological Organization

Observing and Information Systems Department
Tel.: +41 (0) 22 730 82 68 – Fax: +41 (0) 22 730 80 21
E-mail: wcdmp@wmo.int

7 *bis*, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland
www.wmo.int

World Meteorological Organization
Weather • Climate • Water

WMO-TD No. 1425

Climate Data and Monitoring
WCDMP-No. 66

A Case-Study/Guidance on the Development of Long-term Daily Adjusted Temperature Datasets

A CASE-STUDY/GUIDANCE ON THE DEVELOPMENT OF LONG-TERM DAILY ADJUSTED TEMPERATURE DATASETS

By Manola Brunet[1, 2, ¥*], Oscar Saladié[1], Phil Jones[2], Javier Sigró[1], Enric Aguilar[1], Anders Moberg[3], David Lister[2], Alexander Walther[4] and Carlos Almarza[5]

* Corresponding author:

Manola Brunet
Climate Change Research Group, University Rovira i Virgili
Pza. Tarraco, 1, 43071 – Tarragona, Spain
Tel. +34 977 55 95 83, Fax +34 977 55 95 97
manola.brunet@urv.net

This work is funded by:

- The Commission of the European Union (contract EVK2-CT-2002-00161) (EMULATE)
- The Comisión Interministerial de Ciencia y Tecnología (CICYT), Spain (reference number REN2002-0091/CLI) (SCREEN).

[1] Climate Change Research Group, University Rovira i Virgili, Tarragona, 43071, Spain
[2] Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, UK
[3] Department of Physical Geography and Quaternary Geology, Stockholm University, SE-106 91 Stockholm, Sweden
[4] Earth Sciences Centre, Göteborg University, SE-40530 Göteborg, Sweden
[5] Instituto Nacional de Meteorología, Servicio de Desarrollos Climatológicos, Leonardo Prieto 8, Madrid, 28040, Spain
¥ Co-chair WMO/CCl OPAG 2 Monitoring and Analysis of Climate Variability and Change

Table of Contents

Introduction

Long-term, high-quality and reliable instrumental climate records are indispensable pieces of information required for undertaking robust and consistent studies to better understand, detect, predict and respond to global climate variability and change. Moreover, the development of the most appropriate environmental and societal climate change adaptation and mitigation strategies also requires high quality climate data. In this latter context, scientists, decision makers and application communities require the best data for their particular needs. High quality and high-resolution climate data are also essential for regional detection/attribution studies of climate change (integrating observational and modelling activities), the calibration of satellite data or the generation of high-quality gridded climate data and reanalysis. Furthermore, the shortage of high-quality integrated climate products generally impedes the development of optimum strategies to mitigate and/or adapt the territories to the negative impacts of climate change.

Improved instrumental data, therefore, means better knowledge of long-term changes in the mean and extreme states of the climate and their societal and economic effects, enhanced regional detection/attribution studies of climate change, better calibration of satellite estimates, more consistent high-quality climate reanalysis, more robust assessments of the environmental and socio-economic impacts associated with climate variability and change, better definition of the national strategies for adapting national economies in the face of climate change's hazards and better design of national development agendas for ensuring environmental sustainability and eradicating poverty.

Over most regions of the world, instrumental data extend back in time at least to the 19th century and over a few of them to the late 17th century (i.e. Europe). Historical climate data have been recorded in different places at different time scales ranging from a few sub-daily and daily observations to multiple hourly observations. Huge amounts of data have been recorded since the earliest observational days in the different countries. These have been collected by National Meteorological and Hydrological Services (NMHSs), private organizations and individuals or by the scientific community. However, the existing data heritage is largely under-exploited, mainly due to the different political, social and economic situations that exist among the different regions. Although some NMHSs have undertaken data rescue activities aimed at transferring historical climate records from fragile media (paper forms) to new media (imaging), fewer long-term records than are needed are readily available in digital form. This reality is preventing many regions from developing more accurate assessments of regional climate variability and change.

The glaring lack of readily available and accessible datasets from global to local scales is hampering our knowledge of long-term climate variability and change, its forcing factors, and the environmental and socio-economic impacts associated with current, very likely man-made, climate change. This is a common feature of many meteorological regional and national networks around the world due to different situations and circumstances. Easily accessible digital climate data are mostly restricted to the second half of the 20th century, although this varies from country to country. From the rich observational history and for a few countries, some station records for some Essential Climate Variables (ECVs) (mainly air pressure, air temperature and precipitation) have been recovered, digitized, quality controlled and homogenized. However, this has mostly been done on a monthly basis. Even though some progress has been made at this time scale, there still remain huge amounts of key monthly, daily and hourly observations to be located, recovered, preserved, digitized, quality controlled, adjusted and analyzed. This historical information is retained in fragile media (paper forms) in a wide range of documentary sources and locations, ranging from the NMHSs historical central and local archives to local, national and international libraries and archives in the different regions.

Lack of funding, both at the local/national level and at the regional/international scale, and not the lack of climate data, can partially explain this unsatisfactory current situation. In developing countries, an even greater lack of human, economic and technical resources, together with the isolation of some of them due to historical and political circumstances, explains why present and past observational data remain dispersed and in fragile media. This situation should and could be solved by encouraging international bodies, NMHSs and scholars to make common efforts to locate/recover/digitize/adjust the instrumental data currently held in hard-copy and fragile media in different kinds of national and international archives and libraries. This effort should be undertaken in honour of our forebear's dedication and heroic efforts to consciously and regularly monitor our atmosphere with the scarce resources that were then available. Therefore, a first and essential step ahead is undertaking data rescue missions aiming to locate, recover, preserve and make accessible old climate data, as recently recommended in the WMO World Climate Data and Monitoring Program - WCDMP's Guidelines on Climate Data Rescue (Tan *et al*. 2004).

Furthermore, long-term, high-quality, homogeneous and in this context reliable climate datasets are needed before undertaking any climatic analysis, particularly those aiming to assess climate variability and change. However, it is well known that long climatological time series often contain variations that are not only due to the vagaries of the weather or climate. Mostly these variations are related to non-climatic factors, such as the introduction of new instrumentation, relocation of weather stations, changes in exposure of instruments or in observing practices, modification of the environment surrounding the meteorological stations, etc. At the same time, wrong or aberrant observations are common in most observational systems. All these factors reduce the quality of original data and compromise their homogeneity. Some of these changes can be the cause of the presence of outliers and abrupt discontinuities in the time series (i.e. stations relocations, changes in exposures, in formulae used to calculate averages, changes in observing times, etc.) while others can induce artificial trends or gradual biases in the data (i.e. changes in the environment around the station such as new developments, growing trees gradually shading instruments, urban heat island effects, impacts related to the development of new irrigated fields around the stations, etc.), all widely recognized in the relevant scientific literature and recently assessed in the WMO/WCDMP's Guidance on Metadata and Homogeneity (Aguilar *et al*., 2003).

These non-climatic factors affecting the meteorological records make these data less suitable for the assessment of actual climate variations, through the reduced reliability of the time series. Many problems of data reliability and quality are difficult to deal with, particularly with long records. Long climate time series are likely to be more affected by spurious biases due to, for example, changes in the observing site. However, short records can be negatively influenced by some of the causes of inhomogeneities. In addition, since biases in time series frequently have a similar magnitude as the climate signal (i.e. long-term variations, such as trends or cycles), the use of uncorrected data might lead to misinterpretations about the evolution of the climate. Therefore the identification and minimization of these breaks in time series' homogeneity is essential before any reliable climate study can be carried out for a meaningful assessment of changes in climate. At present, a remarkable variety of tests for assessing homogeneity and adjusting time series are available on a monthly basis (see Peterson *et al.* 1998 for a full review of them), but only a few approaches to homogenize daily data have been developed, as estimating daily adjustments or interpolating monthly correction factors to a daily basis present a new level of complexity.

Before undertaking any data quality control (QC) and homogeneity tests it is necessary to develop, as far as possible, complete metadata or station history information. Metadata are a key piece of information about the data that might report on how, where, when and by whom the data were recorded, collected and transferred. Preferably, good metadata should record all of those potential causes altering the quality and homogeneity of any time series, as these pieces of information about the data are a valuable and essential guide for adequately basing any of the QC and homogeneity assessments chosen for the raw climate instrumental data. Thus, the development of high-quality and adjusted datasets (both on a monthly and a daily basis) constitutes a multi-task activity that incorporates several actions going from the data location and recovery to the calculation and application of the adjustments. In brief, to generate high-quality datasets implies:

- Searches in archives and libraries in order to locate data and metadata
- The recovery of the records and their imaging in order to preserve them in less fragile media than hard-copies
- The digitization of the imaged data in order to enable their treatment
- To iteratively check the quality of the digitized data in order to detect, label and assess suspicious values and outliers, leading to rejection, validation or their substitution after consulting the original sources
- To test the homogeneity of the data in order to detect breakpoints and establish necessary adjustments
- The estimation of the adjustments and their application in order to get homogenized data.

This technical document is published as a WMO World Climate Data and Monitoring Programme (WCDMP) series. It describes the procedures followed by Brunet *et al.* 2006a in the development of the Spanish Daily Adjusted Temperature Series (SDATS), and as a case study it provides good practices for NMHSs and scientists to create high-quality, adjusted and reliable long instrumental climate datasets.

It reviews the array of procedures adopted in order to develop the new daily adjusted dataset, the SDATS dataset, which is composed of the 22 longest and most reliable Spanish daily maximum (T_{max}), minimum (T_{min}) and the derived mean (T_{mean}) temperature records. It uses as a specific example, detailed information on the processing of the Madrid record, in order to give an extended illustration of how the whole procedure has been carried out for this station. The document is structured into ten (10) sections including the introduction. A short background on currently developed and available temperature datasets at different spatial and time scales is provided in section 2. Details of the selected temperature network, data and metadata collection and the sources used are introduced in section 3, together with a specific assessment of the sources employed for recovering Madrid's data. Quality controls applied to daily maximum and minimum temperature series are discussed in section 4, together with an assessment of the results for the entire network and, particularly, for Madrid. Section 5 describes the whole approach to produce daily adjusted temperature series, with a special emphasis on documenting Madrid's homogenization. It is divided into three subsections: In the first, we will show the procedures adopted for minimizing the bias induced by temporal changes in thermometric exposures from the raw T_{max} and T_{min} monthly data, which affects most of the records as it was a common and contemporary bias across the entire network. In the second subsection, the homogeneity test chosen and the results reached are shown and discussed by, first, describing the selection of the groups of reference/candidate stations; through, second, defining the detection pattern of the inhomogeneities found in the data; and by, third, showing the estimated correction scheme for adjusting time series on a monthly basis. Lastly, the third subsection is focused on describing the interpolation method used for estimating daily adjustments. Section 6 gives the approach for creating the regional temperature series for Spain, the Spanish Temperature Series (STS), and provides an assessment of long-term trends of daily T_{mean}, T_{max} and T_{min} time series included in the SDATS dataset. Section 7 reviews and compares the procedures employed by other groups in order to develop daily adjusted datasets. Section 8 gives a glossary of the principal terms for developing high-quality and homogenized datasets and, finally, section 9 lists the reference used in this guidance.

Brief background on the temperature datasets developed at different time and spatial scales

On a global scale and monthly basis, gridded datasets such as CRUTEM3v (Jones *et al*. 2001; Brohan *et al*., 2006), HadCRUT3v (Rayner *et al.,* 2003, 2006; Jones and Moberg, 2003) or the Global Historical Climatology Network (GHCN, Peterson and Vose, 1997) have been extensively used as a source of information for exploring temperature variability (Houghton *et al*. 2001), which enabled the documentation and analysis of long-term temperature change at the largest spatial scales (i.e. Jones and Moberg, 2003). These gridded datasets led to an improved knowledge of temperature variability and change on continental to global scale and they have been the ideal basis for climate research at these largest spatial scales. However, large-scale datasets always present some deficiencies, mainly related to their low spatial resolution, the varying data quality among the records compiled or the decaying number of records going back in time, which makes them less reliable during the earliest parts of the instrumental period. The reconstruction of climate variability has also progressed via the national and supra-national efforts made in data compilation. At these smaller scales, the datasets can achieve better data quality and homogeneity, particularly where data homogenization is based on a higher density of stations and detailed metadata.

On regional and national scales, as well as on monthly time scales, datasets of higher spatial resolution have been developed over different regions of the world: Manley (1974) compiled the Central England Temperature (CET) series, Vincent and Gullet (1999) the Canadian Historical Temperature Dataset, Böhm *et al*. (2001) and Auer *et al.* (2007) the HISTALP dataset, Brunetti *et al*. (2006) the Italian temperature and precipitation dataset and Begert *et al.* (2005) the MeteoSwiss dataset. These and other datasets have enabled an improvement of our knowledge on long-term changes concerning the mean state of local to regional thermal climate (Folland, Karl *et al*., 2001).

On a daily scale, several global and regional databases have been recently developed and are readily available to the research community: the GCOS Surface Network (GSN, Peterson *et al.,* 1997) and the Global Daily Climatology Network (GDCNv1.0, NCDC, 2002). At this time scale, the sparse spatial and temporal resolution of the large scale datasets is still poorer than in the case of the monthly datasets. In this regard, fewer records are available (i.e. 6 stations over mainland Spain at the GCOS/GSN dataset) and they are relatively short time series (i.e. mainly covering the 20th century). At regional to continental scales, others, like the Australian Daily Adjusted Temperature dataset (Trewin, 1999), the European Climate Assessment & Dataset (ECA&D; Klein Tank *et al.,* 2002), the Canadian Daily Temperature dataset (Vincent *et al*., 2002), the EMULATE temperature and precipitation dataset for Europe (Moberg *et al.* 2006), the longest daily CET series (Parker *et al*., 1992) or the SDATS dataset (Brunet *et al.* 2006a) have enabled documentation of changes in temperature extremes on a continental, regional and national scales over different periods.

Therefore, the lack of accessible data in many regions has been and is continuing to hinder our further knowledge on long-term changes in the extreme state of climate, as well as our understanding of whether the observed changes in the mean are affecting the variance and the extreme tails of the temperature distribution (Folland, Karl *et al*., 2001). It is worth mentioning here the activities of the joint World Meteorological Organization (WMO) Commission for Climatology (CCl) / World Climate Research Programme (WCRP) project on Climate Variability and Predictability (CLIVAR) / Joint WMO-Intergovernmental Oceanographic Commission of the United National Educational, Scientific and Cultural Organization (UNESCO) Technical Commission for Oceanography and Marine Meteorology (JCOMM) *Expert Team on Climate Change Detection and Indices* (ETCCDI). The group's aims are to fill in the obdurate lack of daily data and document changes in the occurrence of climatic extremes over regions with sparse data coverage. Regional workshops have been organized and a set of climate change indices developed which focus on extremes (see http://www.clivar.org/organization/etccdi/etccdi.php, Alexander *et al.,* 2006 and references therein). In spite of this, much more effort has to be made both by NMHSs and scientists for developing higher spatial and temporal-resolution quality controlled and adjusted datasets, both on monthly, daily and sub-daily scales. Such datasets constitute an essential requirement for the analysis and monitoring of climate variability and change.

Network details and data and metadata sources for SDATS

Meteorological data have been regularly recorded by individuals, scientific institutions or official meteorological services since the beginning of the instrumental era and they are currently being taken by different kinds of institutions (NMHSs, environmental agencies, private companies, etc.). Regardless of who were the institutions in charge of recording and gathering the data and their related metadata, thousands of millions of individual observations have been registered and archived throughout the instrumental period. Unfortunately, most of these data still remain in hard-copy and in fragile media kept in a wide variety of local, national and international libraries and archives.

The most common sources for locating and retrieving earlier climate instrumental data are longstanding scientific institutions (i.e. National Science Academies), local/central historical archives of the NMHSs and other meteorological-related institutions, local/national/colonial/ international archives, local/national/international libraries and local/national newspaper archives. For more recent recorded data and metadata, the primary source are the institutions responsible for data registering and archiving (NMHSs) either for data holdings at the central national archive and library or at the meteorological sub-regional centres and site archives.

Regardless the institutions and places where the data are kept, another key issue is exploring the type of source where the data were transcribed. Data can be hand-written and kept in the original daily weather reports (observatory reports on daily or monthly basis), or the data can have been transformed to some extent by the staff at the observing site or by the national climatological branch (i.e. calculations of averages or accumulated sums) and held in different sorts of printed meteorological collections (monthly bulletins, annual reports, monographs, etc.). The first step in the ongoing process of developing high-quality climate datasets has to start with intensive searches in the institutions and in the documentary sources where the data could be collected and archived most likely, followed by the recovery of the data (i.e. imaging and storing them), together with an assessment of the potential quality of the source where the data are held. These tasks will initiate the process of getting reliable climate data, and only once the following steps (digitization, quality controlling, homogeneity testing and homogenization of climate time series) have been made, the scientific community and other end users can confidently employ the data in their studies.

In the framework of the European Community (EC)-funded project EMULATE (**E**uropean and North Atlantic daily to **MUL**tidecadal clim**ATE** variability[1]), the authors initiated the process of colleting the longest and most reliable twenty-two Spanish time-series of raw daily T_{max} and T_{min} from different institutions and documentary sources. One aim was to develop a new daily adjusted temperature dataset over mainland Spain, which could confidently be used in the analysis of long-term temperature change and variability over this region.

The rationale for selecting the network was based on various criteria including temporal and spatial coverage, climatic representativeness, long-term continuity of data and potential data quality at highly monitored sites (synoptic or first order stations). Stations with the longest, continuous and most reliable records were chosen in the first place. They had to extend back to the second half of the 19th century or at least to cover the whole 20th century, as one of the objectives of EMULATE was to relate variations and trends in atmospheric circulation patterns to prominent extremes in temperature and precipitation for the period 1850-2003, http://www.cru.uea.ac.uk/cru/projects/emulate/objectives.pdf. Then, stations had to be at well-spaced locations across mainland Spain. Related to this principle, stations also had to be representative of the different climatic regions of Spain. Besides, stations had to be still in use and likely to continue so for the

[1] http://www.cru.uea.ac.uk/cru/projects/emulate

foreseeable future. The last principle that guided the network selection is the data quality which led us to choose those stations catalogued at present as the first order (synoptic) stations of the Spanish official meteorological network and hence they are presumably highly monitored and well-quality controlled by the Servicio de Desarrollos Climatológicos (SDC, Climatological Branch) of the Instituto Nacional de Meteorología (INM, Spanish Meteorological Office).

According to these principles, the 22 stations depicted in Figure 1 and Table I were selected. In Table I, we provide geographical details of the current location of the stations and the potential lengths of record for each one.

Table I. The selected Spanish temperature network. Name of station, current geographical location (geographical coordinates and elevation) and lengths of record

LOCATION	LONG	LAT	ALT (m)	LENGTH
ALBACETE	01º 51' 47" W	38º 57' 08" N	699	1893-2005
ALICANTE	00º 29' 40" W	38º 22' 00" N	81.5	1893-2005
BADAJOZ	06º 49' 45" W	38º 53' 00" N	185	1864-2005
BARCELONA	02º 10' 36" E	41º 25' 05" N	420	1885-2005
BURGOS	03º 36' 57'' W	42º 21' 22" N	881	1870-2005
CADIZ	06º 12' 37" W	36º 27' 55" N	30	1850-2005
CIUDAD REAL	03º 55' 11" W	38º 59' 22" N	627	1893-2005
GRANADA	03º 37' 52" W	37º 08' 10" N	685	1893-2005
HUELVA	06º 54' 35" W	37º 16' 48" N	19	1903-2005
HUESCA	00º 19' 35" W	42º 05' 00'' N	541	1861-2005
LA CORUÑA	08º 25' 10'' W	43º 22' 02'' N	67	1882-2005
MADRID	03º 40' 41" W	40º 24' 40" N	679	1853-2005
MALAGA	04º 28' 57" W	36º 39' 57" N	6	1893-2005
MURCIA	01º 07' 14" W	37º 58' 59" N	57	1863-2005
PAMPLONA	01º 38' 21'' W	42º 46' 06" N	452	1880-2005
SALAMANCA	05º 29' 41" W	40º 56' 50" N	789	1893-2005
SAN SEBASTIAN	02º 02' 22'' W	43º 18' 24'' N	251	1893-2005
SEVILLA	05º 53' 47" W	37º 25' 15" N	31	1893-2005
SORIA	02º 29' 01" W	41º 46' 29" N	1083	1893-2005
VALENCIA	00º 22' 52" W	39º 28' 48" N	11	1864-2005
VALLADOLID	04º 44' 35" W	41º 38'40" N	691	1893-2005
ZARAGOZA	01º 00' 29'' W	41º 39' 43" N	245	1887-2005

The location map of the Spanish network employed in this study (Fig. 1) also provides approximate lengths of records and the elevation of each station. From this plot a reasonably well-spaced distribution of stations emerges representing the main physiographic units of Spain:

- In the coastal lowland sectors there are 2 stations over the Northern Spanish Atlantic coast, 5 over the Spanish Mediterranean coast and 2 over South-Western Atlantic coast.
- In the Spanish inland plateau there are 4 stations over the Northern plateau and 4 over the Southern plateau.
- In the Ebro Valley lowlands there are 3 stations and in the Guadalquivir Valley lowlands, 2 stations.

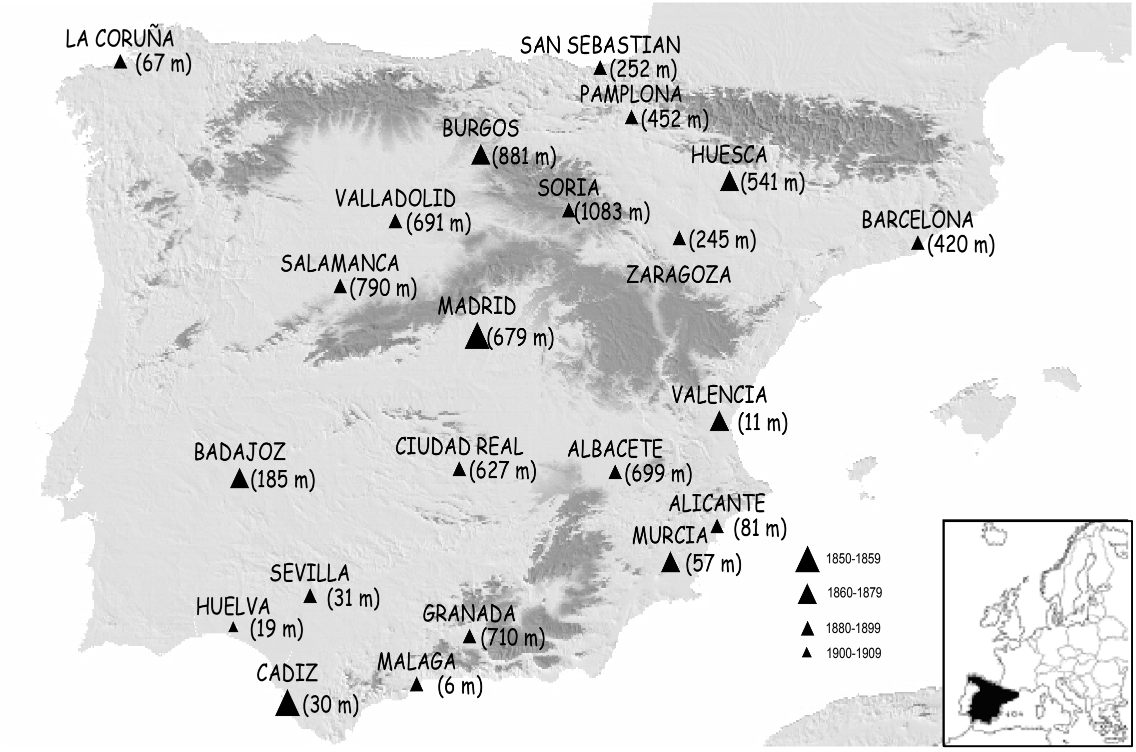


Figure 1. Location map of the 22 Spanish stations used to develop the Spanish Daily Adjusted Temperature Series (SDATS). Names, elevations and approximate lengths of record are shown.

This network essentially covers the entire country and the main Spanish climate types (Oceanic and Mediterranean), sub-types (Atlantic; Mediterranean Continental, Mediterranean Oriental, Mediterranean Southern and Mediterranean Arid or South-Oriental) and variants, according to Martin-Vide and Olcina's (2001) Spanish climate classification:

- Atlantic Galician: 1 station
- Atlantic Littoral Basque: 1 station
- Mediterranean Continental of the Northern Plateau: 4 stations
- Mediterranean Continental of the Ebro Valley: 3 stations
- Mediterranean Continental of the Southern Plateau: 3 stations
- Mediterranean Oriental Catalan: 1 station
- Mediterranean Oriental Valencian: 1 station
- Southern Mediterranean Guadalquivir Valley: 2 stations
- Southern Mediterranean Extremaduran: 1 station
- Southern Mediterranean Littoral: 3 stations
- Mediterranean Arid or South-Eastern variant: 2 stations

An additional and final criterion was considered when choosing the network: the meteorological stations had to be, as far as possible, situated in non-urban areas at least for the second half of the 20th century when a generalized Spanish urban development took place affecting most of the Spanish cities. The authors followed a double strategy in order to avoid potentially biased records related to “urban heat island” (UHI) influences:

First, we compiled, when possible, the records using registered data in small to middle size cities during the 19th century and first half part of the 20th century with those observations taken from the mid-20th century onwards in the nearest non-urban stations situated mainly at airfields and airports. Therefore, we joined the records of Albacete city with Los Llanos Airfield, Badajoz with Talavera la Real Airfield, Burgos with Villafria Airfield, Granada with Armilla Airfield, Malaga with Rompedizo Airfield, Pamplona with Noain Airport, Salamanca with Matacan Airfield, Seville with San Pablo Airfield, Valladolid with Valladolid Airport, Zaragoza with Zaragoza Airport, or with stations located away of urban boundary layer influences in nearby hills (Barcelona with Fabra Observatory and San Sebastian with Igueldo Observatory). See Table IV for further details on the dates of the stations’ compositions.

Second, for those records that could not be compiled with time series taken in nearby rural or non-urban stations due to the absence or low quality and continuity of the data in these nearby monitoring sites, the authors opted for minimizing any artificial trend, especially those related to UHI effects, present in the data by detecting and correcting them using the statistics and factors emerging from the application of the homogeneity test chosen, as documented in the section on homogeneity.

Raw daily T_{max} and T_{min} data were collected from a wide variety of sources, although the bulk of these data (~ 80%) were obtained in digital (~ 48%) and hard-copy (~ 32%) form from INM. The remainder (~ 20%) were recovered in digital (~ 14%) and hard-copy (~ 6%) form from other sources (see legend of Table II and Table III). Table II gives geographical details and lengths of record for every station, with their principal sources. Table III provides information on data and metadata holders and documentary sources. INM data mainly cover the 20th century, while the other sources mostly provide data covering the second half of the 19th century. Hence, the authors, had to locate and retrieve very old data kept in hard-copy, and faced another time-consuming task, the digitization of about 40% of the total data recovered (~ 2 million-daily values).

Table II. List of stations, geographical details and length of records from specific sources

LOCATION	LONG	LAT	ALT (m)	LENGTH	SOURCES
ALBACETE	01º 51' 17" W	39º 00' 00" N	686	1893-1900	BMD
ALBACETE	01º 51' 47" W	38º 57' 08" N	699	1901-2005	INM
ALICANTE	00º 29' 17" W	38º 21' 00" N	19	1893-1900	BMD
ALICANTE	00º 29' 40" W	38º 22' 00" N	81.5	1901-2005	INM
BADAJOZ	06º 49' 45" W	38º 53' 00" N	185	1864-2005	INM
BARCELONA	02º 09' 00" E	41º 23' 00" N	42	1885-1900	MB_ADVICE
BARCELONA	02º 10' 36" E	41º 25' 05" N	420	1901-2005	INM
BURGOS	03º 42' 00" W	42º 20' 00" N	860	1870-1893	INM
BURGOS	03º 42' 00" W	42º 20' 00" N	860	1894-1898	BMD
BURGOS	03º 36' 57'' W	42º 21' 22" N	881	1899-2005	INM
CADIZ	06º 12' 17" W	36º 27' 55" N	30	1850-1996	IMPROVE
CADIZ	06º 12' 37" W	36º 27' 55" N	30	1997-2005	ROASF
CIUDAD REAL	03º 55' 43" W	38º 59' 21" N	627	1893-1900	BMD
CIUDAD REAL	03º 55' 11" W	38º 59' 22" N	627	1901-2005	INM
GRANADA	00º 21' 00" W	37º 11' 00" N	701	1893-1900	BMD
GRANADA	03º 37' 52" W	37º 08' 10" N	685	1901-2005	INM
HUELVA	06º 54' 35" W	37º 16' 48" N	19	1903-2005	INM
HUESCA	00º 19' 35" W	42º 05' 00'' N	541	1861-2005	INM
LA CORUÑA	08º 24' 23'' W	43º 22' 10'' N	9	1882-1900	DWR
LA CORUÑA	08º 25' 10'' W	43º 22' 02'' N	67	1901-2005	INM
MADRID	03º 41' 15" W	40º 24' 30'' N	655	1853-1854	RSP
MADRID	03º 41' 15" W	40º 24' 30" N	655	1855	RSM
MADRID	03º 41' 15" W	40º 24' 30" N	655	1856-1859	LG
MADRID	03º 41' 15" W	40º 24' 30" N	679	1860-1892	ICM/ROAM
MADRID	03º 40' 41" W	40º 24' 40" N	679	1893-2005	INM
MALAGA	04º 25' 36" W	36º 43' 28'' N	29	1893-1900	BMD
MALAGA	04º 28' 57" W	36º 39' 57" N	6.54	1901-2005	INM
MURCIA	01º 07' 45" W	37º 58' 59" N	66	1863-1950	CMTM
MURCIA	01º 07' 14" W	37º 58' 59" N	57	1951-2005	INM
PAMPLONA	01º 38' 21'' W	42º 46' 06" N	452	1880-2005	INM
SALAMANCA	05º 40' 00" W	40º 58' 00" N	811	1893-1900	BMD
SALAMANCA	05º 39' 41" W	40º 57' 23" N	812	1901-1943	INM
SALAMANCA	05º 29' 41" W	40º 56' 50" N	789	1945-1999	PG
SALAMANCA	05º 29' 41" W	40º 56' 50" N	789	2000-2005	INM
SAN SEBASTIAN	02º 00' 00" W	43º 19' 00" N	23	1893-1900	BMD
SAN SEBASTIAN	02º 02' 22" W	43º 18' 24'' N	252	1916-2005	INM
SEVILLA	05º 59' 37'" W	37º 23' 25" N	30	1893-1900	BMD
SEVILLA	05º 53' 47" W	37º 25' 15" N	31	1901-2005	INM
SORIA	02º 28' 00" W	41º 49' 10" N	1058	1893-1900	BMD
SORIA	02º 29' 01" W	41º 46' 29" N	1083	1901-2005	INM
VALENCIA	00º 21' 00" W	39º 28' 00" N	18	1864-1893	INM
VALENCIA	00º 21' 00" W	39º 28' 00" N	18	1894-1900	BMD
VALENCIA	00º 21' 00" W	39º 28' 00" N	18	1901-1935	INM
VALENCIA	00º 22' 52" W	39º 28' 48" N	11	1937-1999	PG
VALENCIA	00º 22' 52" W	39º 28' 48" N	11	2000-2005	INM
VALLADOLID	04º 43' 00" W	41º 39' 00" N	694	1893-1900	BMD
VALLADOLID	04º 44' 35" W	41º 38'40" N	691	1901-2005	INM
ZARAGOZA	01º 00' 29'' W	41º 39' 43" N	245	1887-2005	INM

Sources' acronyms. INM: Instituto Nacional de Meteorología; DWR: Daily Weather Reports; BMD: Boletín Meteorológico Diario; MB-ADVICE: Mariano Barriendos and EU-project Annual to Decadal Variability In Climate in Europe; ICM/ROAM: Instituto Central Meteorologico/Real Observatorio Astronómico de Madrid; RSM: Ricos Sinobas Manuscript; RSP: Rico Sinobas Paper; LG: La Gaceta de Madrid; PG: Pavel Groisman; CMTM: Centro Meteorológico Territorial en Murcia del INM; IMPROVE: EU-project Improved Understanding of Past Climatic Variability from Early European Instrumental Data; ROASF: Real Observatorio de la Armada en San Fernando (Cádiz, España).

Figure 2 shows the amount of available data with respect to the potential daily mean temperature data for each year of the period 1850-2005, together with the absolute number of records contributing to each year of this period. Two stations are available from the 1850s onwards, six from the 1860s, seven from the 1870s, 11 from the 1880s, 21 from the 1890s and 22 from the 1900s. Besides, different periods of missing data are also evident from the inspection of this Figure. The fraction of missing data for the whole period (1850-2005) is about 7% of the potential daily data; however a higher percentage is evident for the period 1863-1940, in which the percentage of available to potential data is ~ 87%. Furthermore, for the two shorter sub-periods of 1899-1905 and 1932-1939, missing data percentages are higher at 22% and 15% respectively. This remarkable reduction in the available data during both time intervals is mainly related to the political instabilities that Spain experienced between the end of the 19th century and the early 1940s, associated amongst others with events like losing the last overseas Spanish Colonies (Cuba and Philippines) in 1898 and with the Spanish Civil War. Due to these severe political and socio-economical crises, the meteorological operational services in Spain were dramatically disrupted during these times. The lack of data drops down to 1.5% during 1940-2005. Although data gaps are filled in some studies, no attempt to fill in gaps has been performed in the present study.

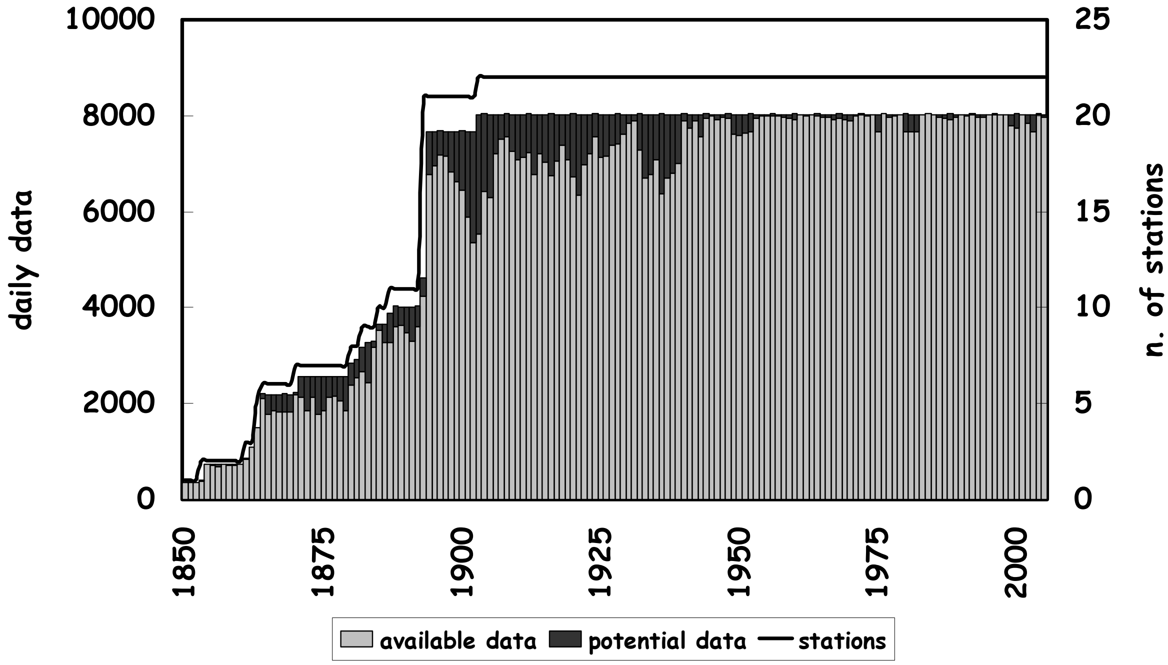


Figure 2. Available versus potential total amount of daily mean temperature data per year and for the whole network (left axis, light and darker grey columns) in relation to the number of meteorological stations (right axis, solid line)

Official meteorological observations in Spain started in 1869, although prior to this date non-official observatories were operating in several Spanish cities. In addition and due to diverse historical circumstances, the majority of the pre-1900 official instrumental data were lost and hence not archived in the central headquarters of INM. However, as was common practice among the nineteenth century meteorological observatories, the data were exchanged both among stations in the same pre-national network and among international networks using data retransmission: telegrams, weekly and monthly reports or annual summaries. It was also common that local newspapers offered daily weather data to their readers, which implies that old newspapers are a source of meteorological data. Nevertheless, as with other indirect sources where the data have been manipulated up to a certain degree, the credibility of this source is lower.

Table III. Data and metadata supplied by data holders and external contributors, together with the documentary sources employed to locate and retrieve daily data and metadata for 1850-2003.

Data holders and external contributors (EC)	**Details on data and metadata recovered and their documentary sources**
INM*, SDC Headquarters, Madrid, Spain	Data: Twenty-one series covering the 20th century in digital and hard-copy, gross errors checking by SDC for the fraction of the digitised records.
CMTM*, Murcia, Spain	Data: One digitised record (Murcia), 1863-1950, gross errors checking by territorial service of INM in Murcia
EC: IMPROVE* project	Data: One digitised record (Cadiz), 1850-1996, gross errors checked (Camuffo and Jones, 2002). Partial metadata on Cadiz's station available in Barriendos *et al*. (2002)
ROASF*, Cadiz, Spain	Data: Cadiz's updating, 1997-2005, gross errors checked by ROASF staff. Metadata: Partial metadata on Cadiz's station available in Gonzalez (1992)
Pavel Groisman, NCDC, Asheville, NC, USA	Data: Two digitised records: Valencia 1937-1999 and Salamanca 1945-1999. NCDC (2002).
EC: MB_ADVICE* project	Data: One digitised record (Barcelona, 1885-1900).
UK-MO National Library and Archive, Bracknell, UK	Data: Daily Weather Reports (DWR): one record (La Coruña, 1882-1900), Boletín Meteorológico Diario (BMD): twelve records, 1893-1900
INM Library & Archive, Madrid, Spain	Data: Madrid station, 1853-1854, in Rico Sinobas (1857). Madrid station, 1860-1889, in ICM (1893). Madrid station, 1890-1892, in ROAM* (1892, 1894). Metadata for the entire network available in: Observatorio de Madrid (1866-1880, 1882, 1884, 1886-1887, 1889, 1891-1892, 1895-1896, 1899-1900): *Resumen de las observaciones meteorológicas efectuadas en la Península y algunas de sus islas adyacentes*. Several editorials: Madrid; ICM (Instituto Central Meteorológico) (1907-1949): *Resumen de las observaciones meteorológicas efectuadas en la Península y algunas de sus islas adyacentes*. Est. Tipo-litográfico de I. Barredo: Madrid, and Almarza *et al*. (1996).
Royal Academy of Medicine, Madrid, Spain	Data: Madrid station, 1855, in RSM*: Observaciones meteorológicas de Madrid:1800-1857, (23 v.) v. 21, (12-8-M-4-23)
Municipal Newspaper Library, Tarragona, Spain	Data: Madrid station, 1856-1859, in La Gaceta (LG) newspaper.
Library of Sociedad Económica de Amigos del País, Badajoz, Spain	Metadata partially available for Badajoz station from Sánchez Pascua (1985)
CMTG (Centro Meteorológico Territorial de Galicia), Climatological Archive, La Coruña, Spain	Metadata available for La Coruña station in Rios Pardo L. 2000. *El observatorio meteorológico (1)*. CMTG, internal report: La Coruña and in "La Ilustración Gallega y Asturiana" newspaper, n 16, June 8, 1881.

* See Table II for meaning of the acronyms employed

To accomplish EMULATE objectives; the authors undertook exhaustive searches in the meteorological sources held in local, national and international archives and libraries. Different documentary repositories were visited in order to locate and retrieve the required 19th century data and their associated metadata, as summarized in Table III. For the data recovered in hard-copy from the documentary sources listed in Table III, the potential quality and credibility of the recovered data varies according to the source employed. From direct documentary sources, such as Rico Sinobas Manuscript (RSM) which contains one year of data (1855) for Madrid station directly transcribed by the observer (M. Rico Sinobas) from his readings in the observatory, to indirect documentary sources, such as the data transcribed and published in newspapers, the sources employed in this study present a wide range of reliability.

Here we provide an assessment of the credibility of the documentary sources used for recovering old data. As listed in Table III, the authors employed the following documentary sources: Daily Weather Reports (DWR), Boletín Meteorológico Diario (BMD, Spanish Daily Weather Report), Rico Sinobas Paper (RSP), RSM, Instituto Central Meteorológico Monograph (ICM, Spanish Meteorological Central Office), Real Observatorio Astronómico de Madrid Reports (ROAM, Royal Astronomical Observatory of Madrid), La Gaceta Newspaper (LG) and El Noticioso newspaper (EN).

The DWR constitutes a daily summary of the 19th century climate data, which were transmitted mainly by telegrams from different meteorological observatories across Europe to the United Kingdom-Meteorological Office (UK-MO), where the data were transcribed and compiled by the UK-MO staff. To being an indirect documentary source, errors of transmission and transcription are possible, although the reliability of this source is high as it was managed by conscientious observers both at the transmission and reception sites. The main problem of this source lies in the lack of data, mainly due to a failure in the delivery of the data.
The RSP is a paper written by the observer in charge of ROAM, M. Rico Sinobas, where he published and analyzed the daily observations for 1853 and 1854 recorded at Madrid Observatory. This is also an indirect documentary source, as the data were transcribed from the original readings, which could incorporate some typographical errors. However, the fact that the author of this paper was the observer of the Madrid observatory makes this source also highly reliable.

The BMD also constitutes a daily summary of the late 19th century Spanish climate data, but this time compiled by the Spanish ICM and containing Spanish data transmitted from individual observatories by their observers to the ICM in Madrid (Spain) where the data were collated, transmitted to other European observatories and published in contemporary Spanish newspapers (i.e. LG) and in monthly or annual reports, bulletins and monographs. It presents the same high degree of reliability as DWR but still suffers the main problem encountered; the lack of data due to failures in the transmission of the daily data from the observatories. Despite this, the completeness of the data is in this case remarkable.

The ICM (1893) publication is a monograph compiled and published by the Spanish Meteorological Office containing 30 years (1860-1889) of daily climate data for Madrid. Typographical errors are quite possible, as it is a printed version of the original readings, although the institution in charge of its publication endorses the accuracy of this monograph.

The ROAM data for Madrid station during 1890-1893 is also held in a highly reliable documentary source of daily climate data, as it constitutes original daily weather reports compiled by ICM/ROAM in 1892 and 1894.

The LG and EN newspapers are likely the most indirect sources of daily climate data employed in this study, and their reliability is thus more constrained. The EN newspaper is the most untrustworthy; about 30% of data during three years of daily observations (1850-1852) were unreliable, as it was found when running the QC of these data. Suspicions were aroused by the presence of repeated values apparently copied from previous day observations (mainly during the weekends using identical values to those recorded for the former Friday). This finding enforced the authors to not employ this documentary source as it was too unreliable. In contrast, the LG newspaper was found to be a more credible source of continuous daily climate data. This official newspaper was publishing, on a daily basis, the observations recorded by the Spanish meteorological network during the second half of the 19th century as transmitted by ICM. Nevertheless, typographical errors are also possible in the process of typewriting the delivered data.

From these sources and others listed in Table III, the authors also compiled metadata for the entire network, in order to better guide the detection of breakpoints in homogeneity of these records. The metadata recovered are far from complete, but a reasonably well-documented history of the stations and records could be recovered to help build a metadata history for the 22 records employed in this study. In this metadata archive we kept complete information on station identifiers, geographical locational data (geographical coordinates, elevation), climate types, subtypes and variants of each station, dates of station relocations related to the different sites that each station has had, measurement units, missing data and data sources. Almost complete metadata on thermometric exposures, instruments, observers and historical circumstances covering the observations are also archived in the metadata database. Finally, partial metadata on the micro- and meso-environments around stations were also recovered and kept in the metadata archive.

In the case of Madrid station, from the potential 113,880 T_{max} and T_{min} daily values to be recovered for the period 1850-2005, the authors retrieved 110,960 values covering the 1854-2005 period. 67% of the data were obtained in digital form from the SDC for the 1893-2005 period, meanwhile the remaining 33% was attained in hard-copy from different data holders and documentary sources covering the 1854-1893 period. The data holders and documentary sources visited for recovering Madrid's temperature are as follows:

- Madrid data from January 1850 to November 1853 was located and recovered from Tarragona Municipal Newspaper Library. The data were kept in the El Noticioso newspaper. As stated, this documentary source showed a very low credibility, as once these data were digitized and quality controlled about 30% of the data were found to be inconsistent. Therefore, these data were not employed in this study.
- From December 1853 to December 1854, the data were located at the INM Library and Archive and recovered from the paper authorized by Rico Sinobas (1857). The source is very credible as this paper was developed by the observer in charge of Madrid station, although being a printed paper some mistakes could be incorporated in the process of transcribing the data.
- For 1855, Madrid's data were located and recovered in the Royal Academy of Medicine (Madrid, Spain) and kept in manuscript form in the collection "Observaciones meteorologicas de Madrid" (Table III). The reliability of this source is the highest as the data were copied by M. Rico Sinobas, the observer in charge of Madrid station.
- For the 1856-1859 sub-period, the authors located and recovered these data from the Municipal Newspaper Library in Tarragona (Spain). The documentary source where the data were kept is in La Gaceta newspaper. The reliability of this newspaper is much greater than in the case of El Noticioso, as it was an official newspaper, presumably well managed, and with a high continuity in its publication. Furthermore, this source, once the data were digitized, did not show the problems with duplicate data.
- From 1860 to 1889, the data were found in the INM Library and Archive. They were recovered from the ICM publication of 1893, which is a 30 years summary of daily climate observations recorded at Madrid and published by the Spanish Central Meteorological Office. Therefore, transcription of the data from the original readings could introduce some typographic mistakes, but the organization in charge of its publication endorses the reliability of this source.
- For the 1890-1893 period the data were also located in the INM Library and Archive and recovered from ROAM reports of 1892 and 1894. Thus, the reliability of this source is also very high, as they are internal annual summaries elaborated by the Madrid's station staff.
- Finally, for the 1893-2005 period the data were obtained in digital form from the SDC. Thus these data have been previously quality controlled from this service's staff looking for gross errors and then the reliability of both the source and data are high.

Raw data quality control

Although some gross error checks were taken by the SDC on the fraction of the digitized daily records retrieved from INM or by Barriendos *et al*. (2002) in a simpler data corroboration with the Cadiz record, a more complete data quality assessment on the entire network has been carried out here to identify and label the suspicious, errant values and outliers that could remain both in the supplied readily-available digitized records and in the data recovered and digitized by the authors.

Careful analysis of data quality, particularly of daily data, is a key and essential activity before undertaking any homogeneity assessment or any long-term climate change analysis with these data. A well-defined data quality control (QC) can avoid many potential errors that could compromise the adjustments calculated from any homogeneity tests run on the data. This is also an indispensable step for improving any assessment with respect to changes in climatic extremes. Furthermore, data quality tests largely enable improvement not only of the quality of the raw data but also the completeness of the dataset, as some fraction of flagged values can be filled in after examining the original sources.

For these reasons, the 22 Spanish raw T_{max} and T_{min} time series were subjected to diverse QC tests, in order to isolate and flag potentially errant values introduced when recording, manipulating, formatting, transmitting or archiving the data, as well as for ensuring internal consistency and temporal and spatial coherence of the data. In this regard, the following recommended (Aguilar *et al*. 2003) set of data quality tests were undertaken with the raw data:

1) Gross error checks:
   i. Aberrant values (T_{max} and T_{min} values > 50 ºC and < -50 ºC)
   ii. Consistency of calendar dates: no. of days per year and no. of days per month
   iii. Comparison of monthly averages between those calculated from the digitized daily data and those listed in the accessible original sources
2) Tolerance tests:
   i. Four or more successive identical values
   ii. Values beyond ± 4 standard deviation (σ)
3) Internal consistency: (T_{max} < T_{min})
4) Temporal coherency: (values exceeding a 25 ºC difference between consecutive observations)
5) Spatial coherency:
   i. Values exceeding ± 4 σ threshold for the difference between the candidate record and its group of reference time series
   ii. Visual comparisons among neighbouring stations

According to the metadata available, the majority of the weather stations have been moved to another location and setting at least once during their operational history. For this reason, quality tests were undertaken separately for each record to distinguish those time intervals belonging to the different station locations. This enabled the use of more realistic thresholds for detecting/labelling suspicious values. Table IV gives details of the various periods tested for the entire network.

The records digitized by the authors and those coming from the old documentary sources described in section 3 were subject to stricter quality controls for avoiding major mistakes in data digitization. First, from the digitized daily data the monthly averages were calculated and these were compared with the monthly values computed by the authors of the documentary sources employed (test 1.iii). When the monthly average calculated by the authors and that given by the source did differ, then all month days values were checked to determine which daily value or values were wrongly keyed and then substituted accordingly. Second, a more restrictive tolerance test 2.i. (i.e. three consecutive equal values instead of four consecutive equal values) was carried out on the data coming from the old documentary sources.

Table IV. Composition of the records: time intervals for each station, together with the INM local code, for which the described quality control procedures have been applied

NAME	INM CODE	PERIOD	NAME	INM CODE	PERIOD
ALBACETE	8178	1893-1936	MURCIA	7182C	1863-1950
AL/LOS LLANOS	8175	1939-2005	MURCIA	7182A	1951-1967
ALICANTE	8025E	1894-1920	MURCIA	7182	1968-1984
ALICANTE	8025G	1921-1938	MU/GUADALUPE	7181I	1985-2005
ALICANTE	8025	1939-2005	PAMPLONA	9262	1880-1974
BADAJOZ	4478	1864-1954	PA/NOAIN	9263D	1975-2005
BA/TALAVERA	4452	1955-2005	SALAMANCA	2870D	1893-1944
BARCELONA	0201E	1885-1925	SA/MATACAN	2867	1945-2005
BAR/FABRA OB.	0200E	1923-2005	SAN SEBASTIAN	1024D	1893-1900
BURGOS	2327	1870-1943	SS/IGUELDO	1024E	1916-2005
BU/VILLAFRIA	2331	1944-2005	SEVILLA	5787D	1893-1932
CADIZ	5972	1850-2005	SEVILLA	5790	1933-1950
CIUDAD REAL	4121C	1893-1970	SE/SAN PABLO	5783	1951-2005
CIUDAD REAL	4121	1971-2005	SORIA	2030	1893-2005
GRANADA	5515A	1893-1937	VALENCIA	8416A	1863-1932
GR/ARMILLA	5514	1938-2005	VALENCIA	8416	1935-2005
HUELVA	4605	1903-1984	VALLADOLID	2422C	1893-1923
HUELVA	4642E	1984-2005	VALLADOLID	2422F	1924-1940
HUESCA	9901F	1861-1943	VALLADOLID	2422C	1942-1969
HU/MONFLORITE	9898	1944-2005	VALLADOLID	2422G	1970-1973
LA CORUÑA	1387	1882-2005	VALLADOLID AIR	2422	1974-2005
MADRID	3195	1853-2005	ZARAGOZA	9443D	1887-1950
MALAGA	6171	1893-1942	ZARAGOZA AIR	9434	1951-2005
MA/ROMPEDIZO	6155A	1943-2005			

A summary of the results of the QC is given in Table V, meanwhile Table VI shows results according to the different tests applied. From the total amount of daily values only a very small fraction of the entire dataset (0.58%) was flagged as potentially erroneous values.

Table V. Summary of QC results for raw daily data as absolute counts and percentages respect to the total amount of data examined

Total amount of tested values	1981192	
Flagged values	11505	0.58%
Recovered values	8090	0.41%
Not recoverable values	3415	0.17%

In a second step, these values were scrutinized one-by-one in the original sources, when available. As stated, about 62% of daily data were obtained in digital form, so we did not have access to the original sources. The authors were only able to look at the original sources among the 752,853 daily values available (38% of the total dataset). This fact constrained the possibility of recovering a larger part of the flagged values, as the original sources could not be directly examined. However, as most of the records in digital form come from the second half of the 20th century they were previously checked for gross errors by SDC.

Accessible original sources were consulted to ensure that no digitization or data manipulation errors were introduced in the dataset. When this occurred, the true value recovered from its source was accordingly replaced within the dataset. About 70% of the total labelled values could be corrected by examining the original sources (Table V), which made the QC of raw data a valuable exercise. In the cases when the flagged values could not be inspected, expert judgment guided the comparison of these values leaving them unchanged or making them missing values. This was done by comparing adjacent values in the same time series, by using other climate elements (pressure and precipitation) and by considering simultaneous observations for well-correlated, climatically related and nearby stations. When these checks supported the hypothesis of an incorrect value, it was set to missing and excluded from any further analysis. In this study, only 30% of the flagged values were set to missing values.

Table VI. The QC results distributed according to the kind of test applied as absolute counts with percentages in parenthesis

	Gross error checks	Tolerance tests	Internal consistency test	Temporal coherency test	Spatial coherency tests
Total of flagged values	4941 (0.25)	5995 (0.3)	161 (0.008)	192 (0.01)	216 (0.01)

The most frequent kind of errors found was associated with mistyped values when digitizing from available original sources, as aforementioned. Most were highlighted when running the 1.iii test (Table VI) and could be recovered, corrected and incorporated into the dataset. The tolerance tests applied also provided a remarkable number of potentially errant values, mainly related to four successive and identical values, but after examining the sources the bulk of them were validated and remained in the dataset to be considered as true values. The rest of the quality checks were passed with few cases of potential errors in comparison with the other two tests, and most of them could be easily corrected. QC procedures were repeated several times to account for the effect of an outlier's removal on the statistical distribution of data.

In the case of Madrid data, from the recovered 110,960 daily T_{max} and T_{min} values covering the 1850-2005 period, only 111 values (0.10% of the total amount) were labelled as potential errant values during the QC. The 0.06% of the total (73 values) was flagged when running the 1.iii test. Having scrutinized the original data, 64 out of the 73 values were found as incorrect values and could be corrected. The remaining 9 values could not be identified as incorrect daily values as the original monthly averages were wrongly computed in the source and, then, were validated as true values. 21 values (0.02% of the total) corresponded to 4 or more days of consecutive identical values, but all of them were validated as right values when visiting the original data and testing their temporal and internal consistency. 7 values were flagged as exceeding $\pm 4\sigma$, but all of them were validated and kept as true values once they were compared with the highest correlated stations (Soria, Badajoz, Burgos and Valladolid). 4 values exceeded the expected amount of change between consecutive daily observations (test 4), but were validated as well once compared with the corresponding values recorded at those nearby stations. Finally, 6 values were labelled as potential wrong values when running the spatial coherency test 5 i. 4 out of 6 values were correct and then they were validated, whereas 2 were considered wrong, but they could be substituted by the right values having examined the original data.

The homogenization procedures adopted

As has been widely documented in the relevant literature for many years, the majority of long-term climate time series have been influenced by non-climatic factors, mainly related to changes in station locations, local environments, instrumental exposures and instrumentation, observing practices or data processing. All these factors can introduce and have introduced gradual or abrupt breaks in the homogeneity of climate records. To estimate adjustments for these problems is another key pre-requisite before undertaking any climatic analysis, especially regarding any long-term climate variability and change assessment.

In this section, the homogenization procedures applied to the 22 Spanish daily temperature records are explained. Firstly, we describe the empirical approach employed, for the very first time, to minimize the so-called "screen bias" affecting the earliest parts of the temperature records, which is related to changes over time in thermometric exposures. Secondly, we show the application of the Standard Normal Homogeneity Test (SNHT) developed by Alexandersson and Moberg (1997) to detect breakpoints in the data, establish the correction pattern and estimate the corresponding monthly adjustments. Thirdly, we address the scheme adopted for interpolating the estimated monthly adjustment factors to the daily timescale.

The minimization of the "screen bias" in the Spanish temperature network

Prior to the generalized use of Stevenson screen, varied types of exposures and stands were used for protecting thermometers (i.e. north walls, Glaisher and Montsouris' open stands, Wild metallic cylindrical screen, etc.). Exposures differed on a country-by-country basis over the globe, as documented by Parker (1994). According to this study, temperature readings recorded under the older stands are likely biased to a higher or lower degree depending on latitude, observation time during the day and year, and hence this bias has had a different impact on daily extreme (maximum or minimum) temperatures. Several studies have shown and estimated the sign and magnitude of this bias on thermometric records over different areas of the world, highlighting varied impacts on the records associated with the thermometer shelters and exposures employed in the past. For Mediterranean climates it has been documented (i.e. Nicholls *et al.* 1996) that pre-sheltered temperatures tended to have a strong warm bias in T_{max} records compared to current observing practices while T_{min} readings had a small cold bias compared to the modern period. Therefore, the use of the original data in estimating long-term temperature evolution gives negatively-biased results in terms of trend.

For the Spanish meteorological network during the 19th century and early decades of the 20th century, it was common to protect thermometers under open stands. These were mostly the Montsouris or French stand and to a lesser extent the Glaisher stand. These types of open stand were more sensitive to radiation effects than the currently employed Stevenson screen. Therefore, pre-screened observations mainly show unreliable high maximum temperatures in the early parts of the longest Spanish records. However, to address this bias by means of relative homogenization procedures is difficult, as it was a contemporary and common feature of the early surface observational network. Consequently, as all records were affected to a lower or higher degree during a similar period, a relative homogeneity assessment will not work and fail to identify both its magnitude and fluctuations throughout the day and year in the examined records. Therefore, other absolute homogeneity approaches are necessary for coping with this bias. In this regard and within the framework of the Spanish-funded SCREEN[2] research project, paired temperature observations were taken using the old-Montsouris stand and modern-Stevenson screens in the meteorological gardens of La Coruña and Murcia, representative of the Oceanic climate and the Mediterranean Arid or South-eastern climate subtype and of high and low cloud cover levels respectively. In this project, the 19th century Montsouris stands were built and operated according to details given in publications from the period (i.e. Angot, 1903).

[2] "Assessment and minimisation of 'screen bias' incorporated into the longest Spanish air temperature records by time changing thermometric exposures throughout dual temperature observation (SCREEN)"

Figure 3. Picture showing the reproduction of the old Montsouris stand and the new Stevenson screen as replicated in the Meteorological Garden of Murcia (INM, Spain). Dual temperature observations are taken with identical sensors sheltered in a Stevenson screen (back) and in a Montsouris stand (in front). Courtesy of the project SCREEN

Figure 3 shows the replica of the ancient Montsouris stand assembled and installed at the Meteorological Garden of Murcia station illustrating the pre-Stevenson Screen exposures employed in the Spanish network. Dual temperature observations were and are carried out at both meteorological gardens in order to estimate the magnitude and fluctuations of the "warm bias" ("cold bias") of daily maximum (minimum) records in order to develop an analytical procedure for minimizing it from the longest Spanish records.

A preliminary empirical minimization of the "screen bias", before undertaking the homogeneity testing of the records, was carried out when adjusting the raw monthly averages of 20 out of the 22 Spanish daily T_{max} records as discussed in Brunet *et al.* (2006a), as only one year of paired observations were then available. The approach simply consisted of subtracting from the T_{max} monthly raw values the median of the daily differences estimated from simultaneous T_{max} readings recorded under both the old (our reconstructed Montsouris stand) and new (Stevenson screen) exposures. After collecting two complete years of simultaneous daily temperature observations (from July 2003 to June 2005 at La Coruña and from March 2003 to February 2005 at Murcia), we have reassessed both the scheme employed for minimizing the "screen bias" from the longest Spanish records and its application to also adjust this bias in the longest T_{min} records (Brunet *et al.* 2006b).

Here we briefly describe the new procedure followed for a more robust approach to minimizing "screen bias" from 21 out of the 22 Spanish monthly T_{max} and T_{min} raw averages. This time we have applied this scheme to also minimize Cadiz T_{max} and T_{min} records for 1850-1875, as new recovered metadata showed that at this station, thermometric observations were also taken under an open stand during the earliest pre-1875 instrumental period. The records for Malaga have not been adjusted, as our metadata for this station indicated that the thermometers were exposed for an undefined period between the last decades of the 19th century and early 20th century inside a louvered rectangular hut of 2m × 3m × 2m with a door opening to the north, which would have a different influence on the readings to that the induced by open stands. From the two years of simultaneous daily temperature and other related meteorological variables (i.e. daily values of sunshine, cloud coverage, air pressure, and wind speed and also sub-daily observations) recorded at both locations, we first correlated these variables in order to explore the most robust relationships among them. The highest Spearman (Rho) correlation coefficients have been estimated between T_{max} and T_{min} series recorded under Montsouris stands and Stevenson screens (r: > 0.99). Also, the highest correlations have been found between the “screen bias” of maximum temperature (ΔT_{max}) and T_{max} temperatures registered under Montsouris screens at both locations. Series of ΔT_{max} have been estimated as the difference between daily readings registered under both exposures (Montsouris minus Stevenson).

Table VII shows Spearman (Rho) correlation matrix between daily maximum and minimum temperatures taken under a Montsouris stand and a Stevenson screen at the meteorological garden of Murcia, the “screen bias” for maximum (ΔT_{max}) and minimum (ΔT_{min}) and other related daily and sub-daily meteorological variables, meanwhile Table VIII gives similar information but for La Coruña. These correlations show an inverse relationship between both measurements indicative of higher maximum temperatures leading to a higher screen bias. Other significant relationships, but with much lower correlation strength, have been found between ΔT_{max} and sunshine and wind speed (sunshine, wind speed and cloud cover) for Murcia (La Coruña). The relationships between ΔT_{min} and the other variables at both locations show much weaker correlation coefficients both between the T_{min} series recorded under Montsouris stand and among the other related variables.

Based on the highest correlation coefficients and the linear relationship established between temperature observations taken under both exposures, we have developed two linear regression models, one for each location, using Montsouris T_{max} readings to predict Stevenson T_{max} values. Both models explain the 99% of T_{max} series variance for Murcia and the 98.6% for La Coruña and their expressions are:

$$Tc = -0.508 + (0.975\ Tr) \text{ for Murcia and}$$
$$Tc = 0.059 + (0.949\ Tr) \text{ for La Coruña}$$

Where Tr is the raw data measured under Montsouris stands and Tc the corrected temperature as measured under Stevenson screens.
Table VII. Spearman (Rho) correlation matrix between daily maximum and minimum temperatures recorded with a Montsouris stand and a Stevenson screen at the meteorological garden of Murcia, the “screen bias” for maximum (ΔT_{max}) and minimum (ΔT_{min}) and other related daily and sub-daily meteorological variables. Bold (italic) indicates significance at 1% (5%) confidence level.

	T_{max} Stevenson	T_{max} Montsouris	ΔT_{max}	T_{min} Stevenson	T_{min} Montsouris	ΔT_{min}
cloud amount 07	**-0.39**	**-0.39**	**0.18**	**-0.11**	**-0.11**	0.01
cloud amount 13	**-0.43**	**-0.42**	*0.09*	**-0.16**	**-0.16**	0.06
cloud amount 18	**-0.35**	**-0.35**	*0.09*	**-0.15**	**-0.15**	0.01
Daily average wind speed	**0.23**	**0.23**	**-0.14**	**0.27**	**0.26**	0.02
daily sunshine	**0.66**	**0.67**	**-0.37**	**0.39**	**0.39**	*-0.09*
air pressure 00	**-0.11**	**-0.11**	0.05	**-0.21**	**-0.21**	**-0.11**
air pressure 07	**-0.11**	**-0.11**	0.05	**-0.20**	**-0.20**	**-0.10**
air pressure 13	**-0.15**	**-0.16**	0.08	**-0.20**	**-0.20**	*-0.10*
air pressure 18	**-0.21**	**-0.21**	**0.11**	**-0.24**	**-0.24**	-0.08
ΔT_{max}	**-0.52**	**-0.56**	**1**	**-0.48**	**-0.48**	0.07
ΔT_{min}	*-0.09*	*-0.09*	0.07	**-0.19**	**-0.21**	**1**
T_{max} Montsouris	**0.99**	**1**	**-0.56**	**0.87**	**0.86**	*-0.09*
T_{max} Stevenson	**1**	**0.99**	**-0.52**	**0.86**	**0.86**	*-0.09*
T_{min} Montsouris	**0.86**	**0.86**	**-0.48**	**0.99**	**1**	**-0.21**
T_{min} Stevenson	**0.86**	**0.87**	**-0.48**	**1**	**0.99**	**-0.19**

Table VIII. As Table VII but for observations recorded at the meteorological garden of La Coruña

	T_{max} Stevenson	T_{max} Montsouris	ΔT_{max}	T_{min} Stevenson	T_{min} Montsouris	ΔT_{min}
cloud amount 07	**-0.21**	**-0.21**	*0.09*	0.07	0.07	-0.02
cloud amount 13	**-0.25**	**-0.27**	**0.23**	-0.03	-0.03	0.03
cloud amount 18	**-0.22**	**-0.23**	**0.15**	-0.07	-0.07	0.01
Daily average wind speed	**-0.34**	**-0.35**	**0.25**	**-0.21**	**-0.20**	**-0.19**
daily sunshine	**0.37**	**0.40**	**-0.35**	**0.13**	**0.13**	0.00
air pressure 00	-0.05	-0.06	**0.12**	**-0.11**	**-0.11**	-0.06
air pressure 07	-0.05	-0.06	0.07	**-0.11**	**-0.11**	-0.05
air pressure 13	-0.07	-0.07	0.02	**-0.11**	*-0.10*	-0.03
air pressure 18	*-0.08*	*-0.08*	0.00	*-0.10*	*-0.10*	-0.01
ΔT_{max}	**-0.26**	**-0.36**	**1**	**-0.19**	**-0.19**	**-0.14**
ΔT_{min}	**0.24**	**0.25**	**-0.14**	**0.25**	**0.22**	**1**
T_{max} Montsouris	**0.99**	**1**	**-0.36**	**0.87**	**0.87**	**0.25**
T_{max} Stevenson	**1**	**0.99**	**-0.26**	**0.88**	**0.88**	**0.24**
T_{min} Montsouris	**0.88**	**0.87**	**-0.19**	**0.99**	**1**	**0.22**
T_{min} Stevenson	**0.88**	**0.87**	**-0.19**	**1**	**0.99**	**0.25**

The Spanish stations corrected with the Murcia and La Coruña equations are listed in Table IX, which also show the dates of Stevenson screen introduction in each one of the analyzed stations.

Table IX. Dates of Stevenson screen introduction in the Spanish meteorological network, which define the periods of application of the monthly adjustment factors for maximum and minimum temperatures. In bold (italic) adjusted stations using Murcia (La Coruña) estimated monthly factors.

Albacete	4/1915	**Alicante**	1/1909	**Badajoz**	1/1909	**Barcelona**	1/1901
Burgos	1/1905	**Cadiz**	1/1875	**Ciudad Real**	1/1908	**Granada**	1/1909
Huelva	1/1909	**Huesca**	6/1912	*La Coruña*	4/1912	**Madrid**	1/1894
Murcia	1/1913	*Pamplona*	1/1916	**Salamanca**	1/1909	*S. Sebastian*	1/1901
Sevilla	5/1912	**Soria**	1/1914	**Valencia**	1/1901	**Valladolid**	10/1912
Zaragoza	4/1913						

To test the performance of both linear regression models, we have compared the observed monthly differences between readings taken under Stevenson and Montsouris exposures with the predicted values.

Table X. Observed versus predicted monthly differences (ΔT_{max}) of daily maximum temperatures (in ºC) recorded with the Stevenson and Montsouris exposures at the meteorological gardens of La Coruña and Murcia for the two years of paired temperature observations (see text for details)

La Coruña					Murcia				
Years	Months	Montsouris T_{max} averages	Observed ΔT_{max}	Predicted ΔT_{max}	Years	Months	Montsouris T_{max} averages	Observed ΔT_{max}	Predicted ΔT_{max}
2003	7	22.99	-1.19	-1.17	2003	3	21.60	-1.01	-1.05
2003	8	26.32	-1.09	-1.37	2003	4	24.53	-1.11	-1.12
2003	9	23.97	-1.10	-1.23	2003	5	28.23	-1.23	-1.21
2003	10	18.47	-0.79	-0.91	2003	6	34.95	-1.39	-1.38
2003	11	17.03	-0.56	-0.83	2003	7	36.82	-1.33	-1.43
2003	12	14.48	-0.42	-0.68	2003	8	36.91	-1.21	-1.43
2004	1	14.96	-0.60	-0.71	2003	9	31.90	-1.22	-1.30
2004	2	15.51	-1.00	-0.74	2003	10	25.70	-0.99	-1.15
2004	3	15.53	-1.14	-0.74	2003	11	21.17	-0.77	-1.04
2004	4	16.35	-1.06	-0.79	2003	12	18.52	-0.87	-0.97
2004	5	19.59	-1.32	-0.98	2004	1	20.68	-0.76	-1.02
2004	6	23.54	-1.50	-1.21	2004	2	19.15	-1.02	-0.99
2004	7	23.41	-1.43	-1.20	2004	3	20.49	-1.16	-1.02
2004	8	24.13	-1.25	-1.24	2004	4	22.96	-1.35	-1.08
2004	9	23.32	-1.08	-1.19	2004	5	25.82	-1.59	-1.15
2004	10	19.04	-0.68	-0.95	2004	6	33.80	-1.51	-1.35
2004	11	15.87	-0.39	-0.76	2004	7	34.63	-1.59	-1.37
2004	12	14.13	-0.36	-0.66	2004	8	36.29	-1.42	-1.41
2005	1	14.28	-0.57	-0.67	2004	9	32.49	-1.34	-1.32
2005	2	12.50	-0.76	-0.57	2004	10	28.75	-1.20	-1.23
2005	3	17.38	-0.79	-0.85	2004	11	21.17	-0.94	-1.04
2005	4	17.41	-1.11	-0.85	2004	12	17.63	-0.90	-0.95
2005	5	19.59	-1.26	-0.98	2005	1	17.62	-0.96	-0.95
2005	6	23.29	-1.12	-1.19	2005	2	16.91	-1.12	-0.93

Table X shows the monthly averages of daily maximum temperatures recorded under Montsouris stands at both locations, the observed monthly differences (Stevenson minus Montsouris) values and the predicted differences for the two years of paired observations. As can be deduced from this Table, the performance of both linear regression models predicting T_{max} monthly differences between both exposures is highly accurate, with an average difference between the observed and predicted monthly average of 0.1 ºC for Murcia and 0.2 ºC for La Coruña.

For minimum temperature, linear regression models do not provide good adjustments, due to the weak relationships established between the T_{min} values taken under Montsouris stands and the screen bias estimated from T_{min} series (ΔT_{min}), as well as the very weak and constant magnitude of the screen bias across the year, as shown in Table XI. This Table also provides the monthly median differences estimated during the two years of dual daily minimum temperature observations recorded under both exposures and locations together with their 95% confidence intervals. Given these statistically significant differences, we have also opted for minimizing the screen effects from the minimum temperature series by adding the estimated monthly medians to the monthly T_{min} raw averages of the 21 Spanish records. The application of Murcia and La Coruña adjustments to the T_{min} monthly raw averages of the 21 Spanish records is made according to the same association as for T_{max} as in Table IX. The Cadiz T_{min} record is now corrected with Murcia monthly factors during 1850-1875.

As evident from the monthly ΔT_{min} at both experimental meteorological sites shown in Table XI, the “cold bias” induced by open exposures in minimum temperatures, even though being statistically significant, is of a very low magnitude across the year without showing a clear or marked annual cycle as maximum temperature records do.

Table XI. Monthly adjustments (in ºC, with the 95% confidence interval in brackets) estimated from difference time series of daily minimum temperature recorded with the Montsouris and Stevenson exposures at the two Spanish meteorological gardens of La Coruña (north-western Spain) and Murcia (south-eastern Spain) where the paired observations were simultaneously recorded for minimizing "screen bias" of the pre-Stevenson records

Months	La Coruña	Murcia
Jan	0.18 (0.14/0.22)	0.27 (0.21/0.33)
Feb	0.16 (0.10/0.21)	0.19 (0.14/0.24)
Mar	0.17 (0.13/0.21)	0.13 (0.08/0.18)
Apr	0.14 (0.09/0.19)	0.16 (0.10/0.22)
May	0.14 (0.10/0.17)	0.16 (0.10/0.22)
Jun	0.21 (0.17/0.24)	0.21 (0.16/0.26)
Jul	0.17 (0.14/0.20)	0.13 (0.09/0.17)
Aug	0.26 (0.22/0.30)	0.19 (0.14/0.24)
Sep	0.24 (0.20/0.29)	0.11 (0.06/0.17)
Oct	0.20 (0.16/0.23)	0.27 (0.22/0.32)
Nov	0.19 (0.11/0.27)	0.21 (0.15/0.26)
Dec	0.19 (0.12/0.26)	0.28 (0.23/0.33)

The application of the screen correction to Madrid T_{max} and T_{min} records can be seen in Figure 4. It shows the factors applied to the raw data for correcting annual, winter (December, January and February) and summer (June, July and August) raw averages. As can be seen, annual and seasonal factors have reduced annual averages of T_{max} by about -1 ºC; meanwhile for summer averages a larger reduction of about -1.3 ºC was applied, with a more modest correction for winter raw values (about -0.8 ºC). For the equinoctial seasons a similar reduction to those for the annual values was estimated (not shown).

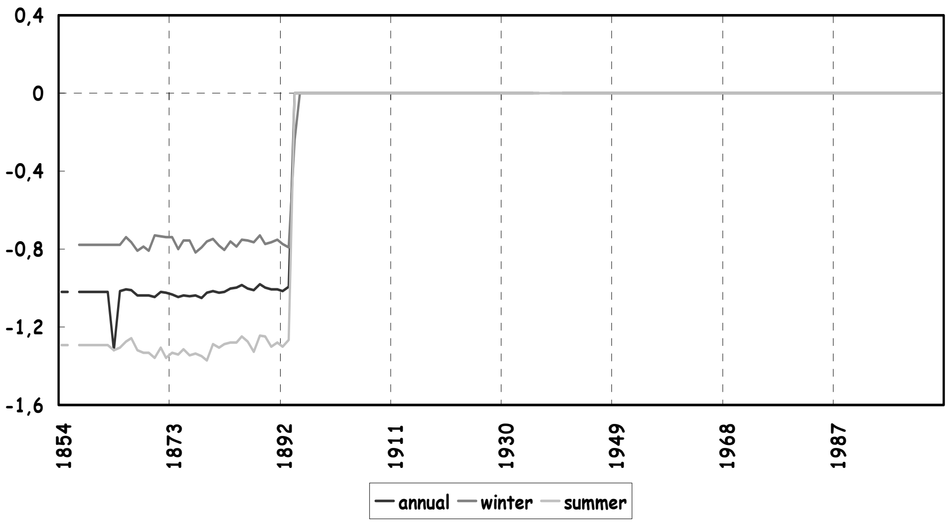


Figure 4. Annual and seasonally (winter and summer) averaged correction factors applied to the corresponding annual and seasonal raw averages of daily maximum temperatures for Madrid station and for the 1854-1893 period when temperatures were recorded under an open stand.

The adjustments made to the Madrid T_{min} records, as stated, consisted in adding to the monthly raw values the estimated median differences at the meteorological garden of Murcia (Table XI).

The application of the SNHT on a monthly basis and results

In order to assess homogeneity of the T_{max} and T_{min} monthly records, once the screen bias adjustments have been applied to these raw time series, the SNHT relative homogeneity approach developed by Alexandersson and Moberg (1997) has been used across the entire network. Detection and correction of inhomogeneities of monthly temperature records has been undertaken following the SNHT application scheme described in Aguilar *et al.* (2002).

Here we discuss both the application and results obtained after using the SNHT on a monthly basis. First, we describe the selection of sets of reference stations for each one of the candidate stations. Second, we present the scheme adopted to detect inhomogeneous periods on an annual and a seasonal scale. And third, we show the correction pattern applied to monthly averages of daily maximum and minimum temperatures within SDATS.

The selection of the groups of reference stations

The selection of the reference stations for undertaking the relative homogeneity assessment of each candidate station was made according to three complementary criteria: first, highly correlated stations; second, geographical and climatic proximity/affinity; and, third, availability of a sufficient number of reference series for any time slice of the candidate record.

Pearson product-moment correlations were estimated among all stations for annual, seasonal and monthly averages of daily T_{mean} calculated from monthly averages of daily T_{max} and T_{min} time series (T_{max} + T_{min} / 2). Coefficients were obtained from first difference series, to avoid the impact of inhomogeneities. A maximum set of 8 to 9 stations were found to be highly correlated with nearby stations across the 20^{th} century.

Table XII. Pearson product-moment correlations (r: in brackets) among annual averages of daily mean temperatures of each candidate station and its group of reference stations

Candidate	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS8	RS9
Albacete (Alb)	Ali (0.90)	Bad (0.85)	Cre (0.92)	Gra (0.87)	Madrid (0.89)	Mur (0.83)	Sev (0.90)	Val (0.84)	
Alicante (Ali)	Alb (0.90)	Bar (0.86)	Gra (0.80)	Huel (0.81)	Mal (0.88)	Mur (0.86)	Sev (0.89)	Val (0,87)	
Badajoz (Bad)	Alb (0.85)	Bur (0.86)	Cad (0.87)	Cre (0.82)	Huel (0.89)	Mad (0.89)	Mur (0.77)	Sal (0.83)	Sev (0.92)
Barcelona (Bar)	Alb (0.84)	Ali (0.86)	Hues (0.85)	Mad (0.84)	Mal (0.82)	Mur (0.80)	Val (0.90)	Zar (0.93)	
Burgos (Bur)	Cre (0.85)	Hues (0.86)	Mad (0.89)	Pam (0.89)	Sal (0.79)	Sor (0.92)	Vall (0.91)	Zar (0,91)	
Cadiz (Cad)	Ali (0.77)	Bad (0.87)	Huel (0.87)	Hues (0.71)	Mad (0.78)	Mal (0.78)	Mur (0,71)	Sev (0,82)	Val (0,73)
Ciudad Real (Cre)	Alb (0.92)	Bad (0.82)	Gra (0.82)	Mad (0.84)	Mur (0.79)	Sal (0.77)	Sev (0.87)	Soria (0.84)	
Granada (Gra)	Alb (0.87)	Bad (0.83)	Cad (0.83)	Cre (0.82)	Mad (0.84)	Mal (0.81)	Mur (77)	Sev (0.83)	
Huelva (Huel)	Ali (0.81)	Bad (0.89)	Cad (0.87)	Cre (0.78)	Gra (0.84)	Mal (0.82)	Mur (0.71)	Sev (0.85)	
Huesca (Hues)	Bar (0.85)	Cad (0.71)	Mad (0.82)	Mur (0.73)	Pam (0.91)	Seb (0.77)	Val (0.73)	Zar (0.90)	
La Coruña (Cor)	Bur (0.90)	Cad (0.84)	Hues (0.87)	Mad (0.90)	Pam (0.87)	Sal (0.79)	Seb (0.87)	Vall (0.92)	
Madrid (Mad)	Bad (0.89)	Bur (0.89)	Cad (0.78)	Hues (0.82)	Mur (0.76)	Sal (0.76)	Sor (0.90)	Val (0.85)	Vall (0.87)
Malaga (Mal)	Alb (0.91)	Ali (0.88)	Bad (0.83)	Cad (0.78)	Cre (0.85)	Gra (0.81)	Huel (0.82)	Mur (0.86)	
Murcia (Mur)	Alb (0.83)	Ali (0.86)	Bad (0.77)	Cad (0.71)	Gra (0.77)	Mad (0.76)	Mal (0.86)	Sev (0.82)	Val (0.88)
Pamplona (Pam)	Bar (0.87)	Bur (0.89)	Hues (0.91)	Mad (0.88)	Seb (0.88)	Sor (0.89)	Vall (0.87)	Zar (0.91)	
Salamanca (Sal)	Bad (0.83)	Bur (0.79)	Cre (0.77)	Hues (0.78)	Mad (0.76)	Soria (0.78)	Vall (0.83)	Zar (0,77)	
San Sebastian (Seb)	Bar (0.83)	Bur (0.86)	Hues (0.77)	Cor (0.87)	Pam (0.88)	Soria (0.83)	Vall (0.83)	Zar (0.84)	
Sevilla (Sev)	Alb (0.90)	Bad (0.92)	Cad (0.82)	Cre (0.87)	Gra (0.83)	Hues (0.85)	Mal (0.84)	Mur (0.82)	
Soria (Sor)	Bur (0.92)	Hues (0.90)	Mad (0.90)	Pam (0.89)	Sal (0.78)	Seb (0.83)	Vall (0.90)	Zar (0.90)	
Valencia (Val)	Alb (0.84)	Ali (0.87)	Bar (0.90)	Cad (0.73)	Gra (0.76)	Mad (0.85)	Mal (0.82)	Mur (0.88)	Zar (0.87)
Valladolid (Vall)	Bad (0.84)	Bur (0.91)	Cre (0.87)	Mad (0.87)	Pam (0.87)	Sal (0.83)	Seb (0.83)	Sor (0.90)	
Zaragoza (Zar)	Bar (0.93)	Bur (0.91)	Hues (0.90)	Mad (0.87)	Pam (0.91)	Soria (0.90)	Val (0.87)	Vall (0.87)	

However, as the number of available time series decay during the second half of the 19th century, as well as the geographical proximity and climatic affinity among stations, the set of reference stations had to be reduced for that period depending on the candidate station. A minimum of three available reference stations was set for the homogeneity assessment with SNHT. For this reason, SNHT could not be undertaken for the following stations and initial sub-periods: Badajoz between 1864-1875, Burgos 1870-1883, Cadiz 1850-1862, Huesca 1861-1882, La Coruña 1882-1885, Madrid 1853-1862 and Valencia 1864-1887. For these records and periods the monthly correction factors that were obtained from SNHT for the subsequent period of each record were used.

Table XII shows correlation coefficients of annually averaged T_{mean} records between each candidate station (right column) and their potential reference stations (columns 2 to 10), calculated using the first difference time series. As evident in this table, correlation coefficients mainly reach r values of 0.8 and 0.9 and in some cases exceed 0.9, although in a few cases the coefficients are lower than 0.75, but higher than 0.70, corresponding to distant (but necessary in the 19th century) reference stations for making the relative homogeneity assessment with SNHT.

For the Madrid station, the group of 9 reference stations with correlations higher than 0.75 shown in Table XII were used for testing homogeneity during the 1893-2005 period. For the period 1863-1869 the group of reference stations for Madrid was: Badajoz, Cadiz, Huesca, Murcia and Valencia. Finally, for the 1870-1892 period, Burgos was added as reference to the former group.

The detection method's pattern of inhomogeneities

After having identified the reference stations, SNHT has been applied to T_{max}, T_{min} and the derived T_{mean} annual and seasonal averages of the 22 meteorological stations, in order to detect potential inhomogeneities in all three variables. At this step, the objective is just detecting breakpoints in the time series, which potentially indicate inhomogeneities reducing the quality of the time series, but not yet adjusting the records. This initial stage of running the SNHT was called the not-guided application of the test, which just aims to detect potential inhomogeneities in time series. The correction pattern emerges from both the inspection of the statistically obtained breakpoints, the analysis of the Q-series (the difference between the candidate series and a weighted average of the reference series) provided by the implementation of the SNHT and also by employing the available metadata in order to associate the statically defined breakpoint with the physical factor to which it could be related to.

Prior to the SNHT application, all 22 stations were regarded as potentially non-homogeneous and after passing through the records, a total of 61 out of the 108 detected breakpoints in homogeneity were validated (2.7 per station on average) on an annual basis. These potential breakpoints in the records were validated through examining the annual and seasonal T_{mean}, T_{max} and T_{min} Q-series. Consequently, for corroborating a breakpoint given by SNHT, but not documented in the available metadata, it had to be at least present in the annual and two seasonal time series of the inspected record, as well as having to be picked by the T_{mean} series and at least by one of the T_{max} or T_{min} series.

For Madrid's records, the not-guided application of the SNHT yielded six breakpoints on an annual basis for T_{max}, T_{min} and T_{mean} series (Table XIII).

Table XIII. Dates and factors of potential inhomogeneities detected and estimated by the SNHT during the not-guided run of the test for Madrid's T_{max}, T_{min} and T_{mean} records for 1854-2005

Time scale	T_{mean} series		T_{max} series		T_{min} series	
	Date	Factor	Date	Factor	Date	Factor
Annual scale	1879 1930 1956	-2.49 1.62 5.52	1891	-8.53	1890 1944	4.87 12.05

Figure 5 illustrates the number of breakpoints identified per year and the number of stations (top panel) and also the number of breakpoints in relation to the available number of records for each year over 1850-2005 (bottom panel). About 25% of the breakpoints detected are located during the politically problematic times in the 1930s, due to the Spanish Civil War, when political instability severely disrupted the Spanish meteorological services. The largest number of breakpoints was detected during the second half of the 19th century (bottom panel) related to the low network density, while during the 20th century no tendency in the number of breakpoints towards a higher/lower frequency of breaks is obvious.

Table XIV shows both the dates of breakpoints in homogeneity (gradual and abrupt) and their causes over the entire network. Individual years indicate a single shift found in the data, while periods show a trend.

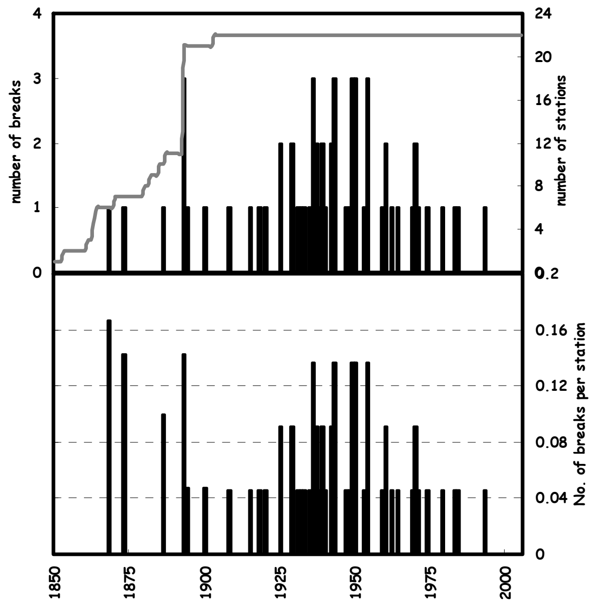


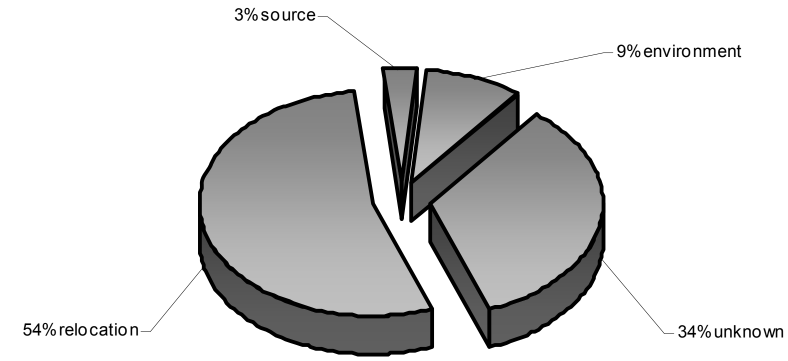
Figure 5. Number of breaks identified per year according to the SNHT applied to annual and seasonal averages of daily T_{mean}, T_{max} and T_{min} time series of SDATS over the 1850-2005 period. Top panel: breaks per year and number of stations contributing per year (grey line). Bottom panel: breaks per year in relation to the number of records available.

Table XIV. The breakpoints (abrupt and gradual) detected by SNHT on an annual basis and their causes for the 22 stations of SDATS

Station	Break	Cause	Break	Cause	Station	Break	Cause	Break	Cause
Albacete	1900	source	1936	relocation	La Coruña	1893	unknown	1915	relocation
Alicante	1920	relocation	1933	relocation		1929	relocation	1930-2003	environment
	1939	relocation	1950	unknown	Malaga	1925	unknown	1931	unknown
	1970	unknown				1942	relocation	1971	unknown
Badajoz	1909-1954	environment	1954	relocation	Murcia	1939	unknown	1953	relocation
Barcelona	1925	relocation				1984	relocation		
Burgos	1905-1943	environment	1943	relocation	Pamplona	1949	relocation	1961	unknown
Cadiz	1964	unknown	1993	relocation		1974	relocation		
Ciudad Real	1938	unknown	1948	unknown	Salamanca	1929	relocation	1936	relocation
	1954	unknown	1962	unknown		1943	relocation		
	1970	relocation	1979	unknown	S. Sebastian	1918	relocation		
Granada	1908	unknown	1937	relocation	Sevilla	1932	relocation	1950	relocation
Huelva	1936	relocation	1950	relocation	Soria	1937	relocation		
	1959	unknown	1983	relocation	Valencia	1873	unknown	1893	unknown
Huesca	1868	unknown	1886	unknown		1935	relocation	1947	relocation
	1894	unknown	1942	relocation	Valladolid	1940	relocation	1969	relocation
Madrid	1893	source	1894-1960	environment	Zaragoza	1906-1949	environment	1949	relocation

From the 61 breaks detected over the entire network, 56 are related to abrupt shifts and only 5 to gradual trends. The gradual trends identified in their respective T_{min} series, likely related to UHI influences, were validated in these 5 cities after consulting the available metadata for these stations related to the urban growth experienced (demographic statistics) during the periods identified by the test.

Figure 6 illustrates the causes of the 61 break points in homogeneity for the entire network (top graph) and the distribution of inhomogeneities found for each station (bottom graph). Changes in location and setting are the main cause of inhomogeneities (about 56% of stations). Station relocations have been common for the longest Spanish temperature records. Stations were moved from one place to another within the same city/town (i.e. from the city centre to outskirts in the distant past and, more recently, from outskirts to airfields and airports far away from an urban influence) and from one setting (roofs) to another (courtyards). These facts obviously induced homogeneity breaks in the analyzed records that have to be corrected.



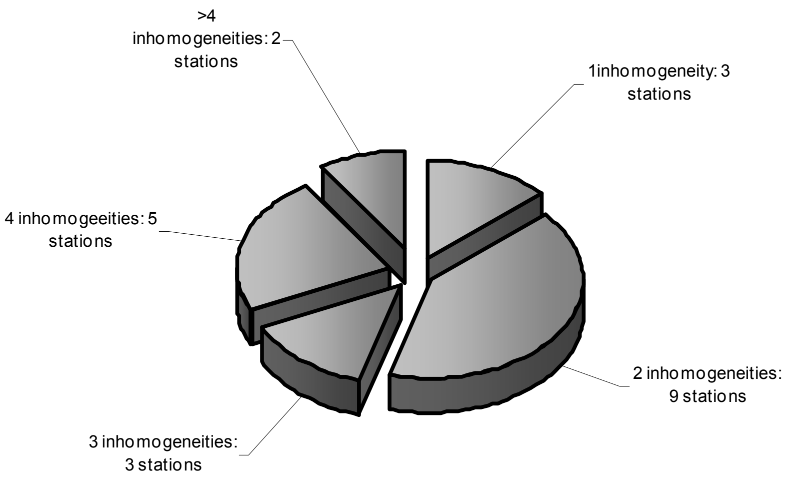


Figure 6. Causes of the inhomogeneities identified (top graph) and frequency distribution of the number of the estimated inhomogeneities (bottom graph) found using SNHT

The topoclimate influence exerted by the urban development had negative effects on five out of the twenty-two stations, as gradual trends were detected (Table XIV and Fig. 6 top graph). For Madrid, the largest Spanish city, an artificial trend was defined between 1894 and 1960 in Madrid's records, when the city had a vigorous expansion (from 489.67 K to 2259.93 K inhabitants). Although this city went on growing after 1960, the urban influence on temperature records has remained steady since then (Almarza, 2000), as described later. The artificial trend present in La Coruña's data has been associated with UHI effects, because since the last documented relocation of the observatory in 1929 to another urban setting the population of this city has increases from

74.13 K to 242.45 K inhabitants in 2002. Zaragoza's records have also been adjusted accounting for an artificial urban trend identified during the 1906-1949 period, when Zaragoza's population grow from 99.12 K in 1900 to 264.26 K inhabitants in 1950. Badajoz had a more modest urban growth during the first half of the 20th century (from 35.04 K in 1909 to 79.29 K inhabitants in 1950) before the switch from the Badajoz urban-station by Talavera la Real airfield-station, the test's results indicated the existence of a gradual trend during the first half of the 20th century. Although Burgos had the lowest rates of urban growth the test indicated a significant artificial trend that was validated associating it with an UHI influence. Finally, the rest of the inhomogeneities identified by SNHT had no explanation in the available metadata archives, which illustrates the incompleteness of the metadata recovered. Summarizing Fig. 6 (bottom graph), for about the half of the network (10 stations) two homogeneity breakpoints were found, while one was detected at four stations, three at two stations, four at another four stations and more than four inhomogeneities at two stations.

In the case of the Madrid station, only one abrupt breakpoint and one gradual trend were validated after taking into account both the SNHT results and the available metadata for this station. From the SNHT not-guided detection of inhomogeneities, both T_{min} and T_{max} series, but not T_{mean} record, showed among others a potential breakpoint in homogeneity around 1890 and 1891 respectively (Table XIII). As documented in Madrid's metadata, in 1893 a change of source and station's setting took place. The observatory was slightly moved, without changing in elevation, to other setting in the same place at El Retiro Park. At the same time the observations were taken by the new Spanish Meteorological Office instead of ROAM. This, together with other factors and varying breakpoint dates among variables determined the validity of this breakpoint. The authors also decided to adjust the warm bias that Madrid's UHI could induce in the Madrid records. This effect has been studied by Almarza (2000) for this city through analyzing the difference time series between the Madrid and Navacerrada records. Navacerrada station is located in the village of the same name in Guadarrama Range at 52 Km NE from Madrid city and 1203 m asl, far away from any urban effect. This study showed the impact of the Madrid's UHI on Madrid's temperature records, as well as proved that the warm bias induced was stabilized from 1958 onwards as the difference time series became stable then. This result showed the need to undertake urban bias minimization on Madrid records from 1894 to 1960 (Table XIV).

The calculation of monthly adjustment factors

The detection pattern identified after undertaking the SNHT exercise for annual and seasonal T_{mean}, T_{max} and T_{min} time series has been applied to the monthly quality controlled T_{mean} values and to the monthly T_{max} and T_{min} pre-adjusted for screen bias data, in order to estimate the required monthly adjustments. Identical breakpoints detected on an annual and a seasonal basis as documented in Table XIV have been reassigned to the 12 months of each record for obtaining for each month and variable the monthly adjustment factors estimated from SNHT.
The frequency distribution of the size of the inhomogeneities for both T_{max} and T_{min} time series is shown in Figure 7. In both cases, moderate correction factors (positive and negative) have been estimated more frequently. The size of about two thirds of the inhomogeneities (66.6%) found for T_{min} records are distributed between the -1.5 and +0.5 ºC intervals. This concentration is even more evident for T_{max}, as more than a half of the inhomogeneities (55.5%) have a smaller size (-0.5 to +0.5ºC).

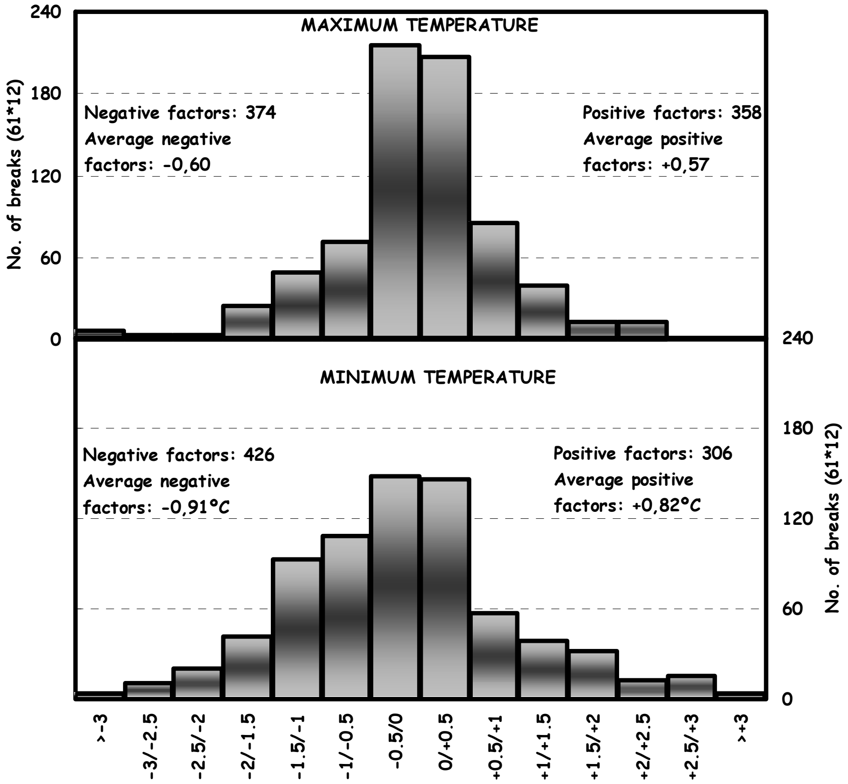


Figure 7. Frequency distribution of break magnitudes referred to the 61 detected and adjusted breaks after using SNHT for homogeneity testing of annual and seasonal averages of daily T_{max} (top plot) and T_{min} (bottom plot) time series for the 1850-2005 period. The figure shows the corresponding breaks as calculated from monthly data.

Figure 7 also makes clear the differences between T_{max} and T_{min} time series taking into account the sign of the inhomogeneities. In the first case, there is a slight predominance of positive values (374) with respect to negative-ones (358), although similar values are obtained on average (-0.63 ºC and +0.57 ºC). For T_{min} records, however, there is a clear difference between positive and negative adjustments. There is a preponderance of negatives values (426) with respect to positive ones (306), which is also evident in the average value calculated (-0.91 ºC and +0.82 ºC).

As an outline for the entire network and period, T_{min} records required larger and more frequent negative adjustments than T_{max} data. These results can be partially explained by the available metadata. As stated, series compilation between urban and rural areas locations has been a common feature for the most of stations, in order to avoid undesirable UHI effect in time series. These compilations have affected T_{min} more than T_{max} series, as usually lower minimum temperatures and slightly higher maximum temperatures are recorded at the new rural sites. In addition, the tendency to change thermometer exposures from courtyard-level to roof-level and *vice versa* could also be another cause of this warm bias present in the minimum temperature data throughout the entire period.

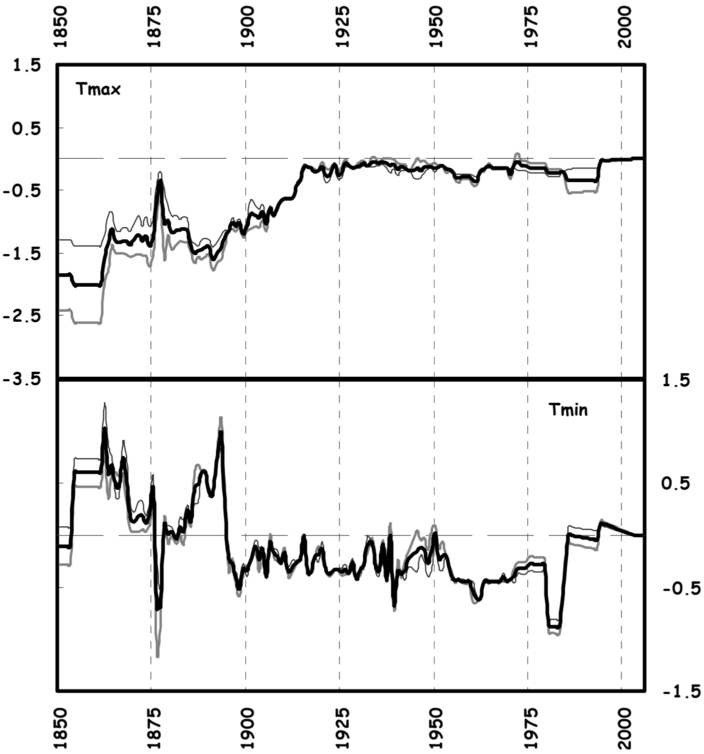


Figure 8. Mean correction factors for T_{max} (top panel) and T_{min} (bottom panel) records expressed as the difference between monthly raw and homogenized values and averaged over all single records. Annual (thick black line), winter half-year (October-March, grey thin line), summer half-year (April-September, grey thick line).

Annual, summer and winter half-year average adjustments, including the systematic error due to the old screens and expressed as the difference between monthly raw and adjusted values, for the 22 T_{max} and T_{min} records are shown in Figure 8 (top and bottom panel respectively).

Several differences between the estimated T_{max} and T_{min} monthly adjustments averaged over all single series can be seen. Pre-1910 adjustments for T_{max} require reduction of the original data by about 1 ºC on average, while the pre-1900 T_{min} records needed lower increases of the raw data (about 0.5 ºC in average). The larger reductions for T_{max} compared to the T_{min} increases during the second half of the 19th century are strongly related to the screen bias minimization scheme. Seasonal differences in adjustments (warm half-year versus cold half-year) can also be seen during that period. The summer half-year required larger reductions than the winter half-year, as the old thermometric stands introduced a higher overestimation of T_{max} readings with respect to the present during this part of the year (Nichols *et al*. 1996). Another difference between T_{max} and T_{min} averaged adjustments across the 20th century is the slightly higher correction factors estimated for $T_{min,}$ as monthly adjustments for T_{max} remain close to zero but for T_{min} they oscillate from -0.5 to 0 ºC.

For the two breaks in homogeneity validated for Madrid records, the abrupt break of 1893 required a reduction of 1.13 ºC for the raw T_{max} and an increase of 1.12 ºC for T_{min} data throughout 1854-1892. The artificial trend validated between 1893 and 1959 required a gradual decrease between 1.14 and 0.02 ºC for the raw T_{max} and an augment between 1.28 and 0.2 ºC for T_{min} data.

The interpolation of monthly adjustment factors on a daily time- scale

As stated in the rationale section of this guidance, homogenizing daily data directly is a complicated task due to the high variability of the daily records in contrast to the more stable monthly, seasonal or annual averaged data. Despite this, the adjustments of daily data are necessary before undertaking any reliable analysis dealing with, for instance, changes in extreme events. From the different and still scarce methodological approaches available to adjust daily data, we have chosen the scheme developed by Vincent *et al.* (2002). This approach attempts to provide a better time-interpolation procedure that preserves monthly averages and does not introduce artificial discontinuities at the beginning and ends of calendar months. This method derives daily adjustments from the calculated monthly correction factors by means of a linear interpolation between midmonth "target" values, which are chosen so that the average of the daily adjustments over a given month equals the monthly correction factors. The "target" values are related to the monthly adjustments by means of a matrix relationship:

$$\mathbf{T} = \mathbf{A}^{-1}\mathbf{M}$$

where **A** is a tridiagonal 12 × 12 matrix, **M** is a 12 × 1 vector of the monthly correction factors and **T** is a 12 × 1 vector consisting of the target values. The "target" values are assigned to the middle day in each month and finally linearly interpolated to get the daily adjustments. This approach has also been employed by Feng *et al*. (2004) to homogenize daily meteorological data for China.
In order to show the impact of our homogenization procedure, the annual averages of daily raw and adjusted temperature records are compared in Figure 9, top and bottom panels respectively for T_{mean} (left), for T_{max} (middle) and for T_{min} time series (right). On these plots, all 22 single annually averaged records together with an estimated regional average are given, each smoothed with a 13-year low-pass Gaussian filter. As evident in the figure, the variability of annual anomalies is reduced by homogenization, particularly for the T_{max} and T_{min} records (Fig. 10, top and bottom panels).

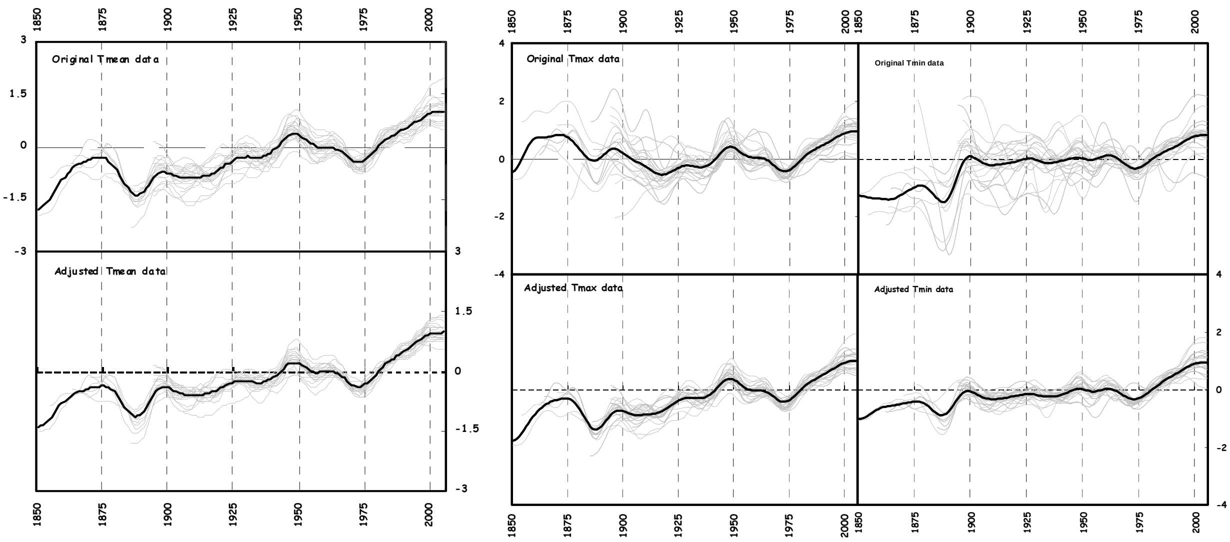


Figure 9. Original (top) and adjusted (bottom) panels annual variations (1850-2005) of the 22 Spanish (thin grey lines) daily mean temperature (left), maximum (middle) and minimum temperatures (right) and their corresponding mean (thick black line) expressed as anomalies from the 1961-1990 baseline period (in ºC) and smoothed with a 13-year Gaussian filter (thick black lines). Notice change in scale for T_{mean} records.

The effect of screen bias minimization on T_{max} records is also evident in the pre-1910 data (Fig. 10, middle bottom panel) when compared with raw data (Fig. 10, middle top panel). The monthly and daily adjustments are listed at http://wwwa.urv.net/centres/Departaments/geografia/clima/adjustments.pdf.

In order to illustrate the adjustment made to Madrid's T_{mean}, T_{max} and T_{min} records, in Figure 11 we show the original versus homogenized time series on annual basis.

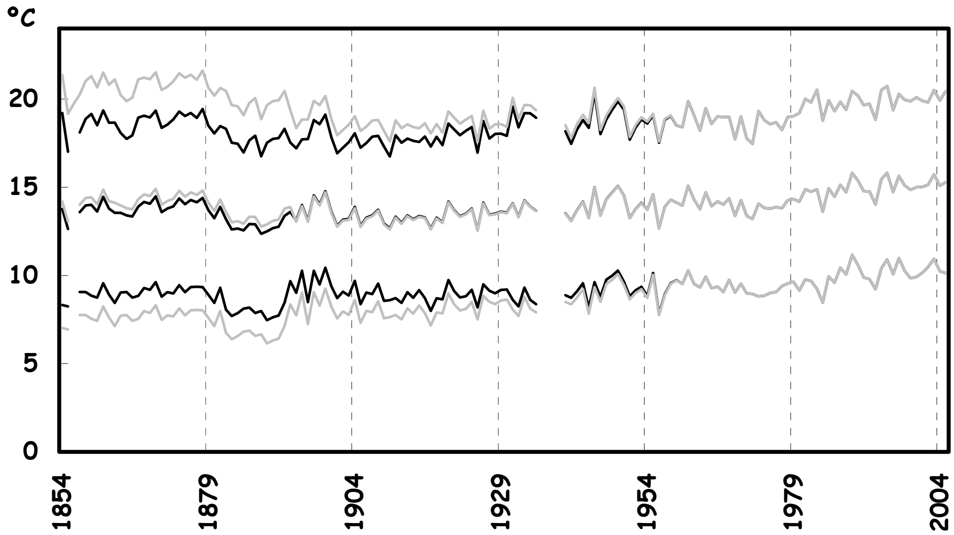


Figure 10. Annually averaged raw (grey lines) and adjusted (black lines) data of the T_{mean} (middle curves) T_{max} (upper curves) and T_{min} (lower curves) time series for Madrid over the 1854-2005 period (in ºC).

As can be seen from Figure 10, the different sign of the adjustments applied to the raw T_{max} and T_{min} records of Madrid has had a slight impact in the correction of the T_{mean} series. For T_{max} series, the shown screen adjustments together with the factors calculated to account for the abrupt shift and the gradual trend detected and validated in 1893 and between 1894 and 1960 respectively have mainly reduced the original data by about 2 ºC on average during 1854-1892 and between 1.14 ºC and 0.02 ºC over the 1894-1959 period. For T_{min} series the adjustments related to the screen minimization have warmed the data by 0.19 ºC, meanwhile the 1893 breakpoint warmed the data by 1.12 ºC with the gradual trend between 1.28 ºC and 0.2 ºC.

The creation of a regional temperature series for Spain: the Spanish temperature series and an initial trend analysis

The 22 daily adjusted records of maximum, minimum and mean temperatures have been combined to obtain regional time series for the period 1850-2005, which gives a representative series of the long-term temperature evolution over mainland Spain, both regarding the mean and the extreme state of Spanish thermal climate. The Spanish Temperature Series (STS), have been generated by averaging daily anomalies from individual records and then adding back the base-period mean (1961-1990), according to the Jones and Hulme (1996) method of separating temperature into its two components (the climatology and the anomaly).

Here we present some results on the variations and trends observed and estimated on an annual basis for the corresponding regional (mainland Spain) averaged time series over the period 1850-2005. In Figure 11 we show annual anomaly values of T_{mean} (upper panel), T_{max} (middle panel) and T_{min} (lower panel) time series, expressed as departures from the 1961-1990 baseline period. The annual anomalies are smoothed with a Gaussian filter of 13-years.

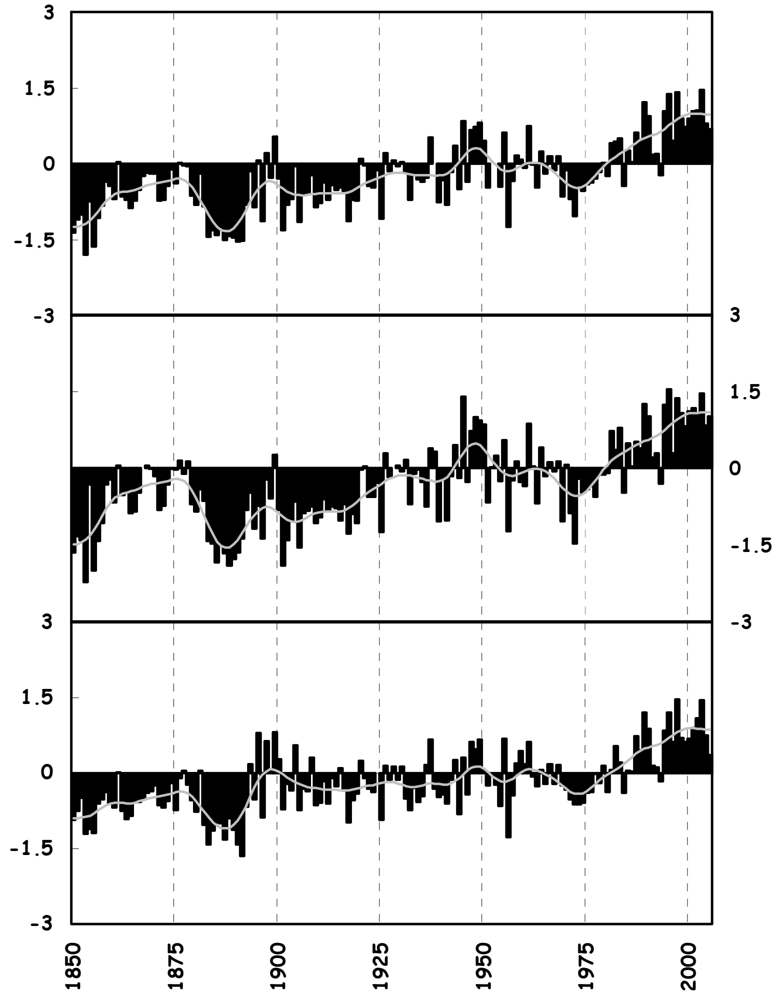


Figure 11. Annual variations (1850-2005) of STS daily mean (upper panel), maximum (middle panel) and minimum (lower panel) temperatures anomaly values (black columns), expressed as departures from the 1961-1990 baseline-period (in ºC) and smoothed with a Gaussian filter of 13-terms (grey lines).

From the inspection of Fig. 11 there is a tendency towards a general Spanish warming over the entire period and for the three daily variables. However, this warming has not taken place in a monotonic or gradual way, as several sub-periods of rising, falling or relatively stable in temperatures can be seen by inspecting the smoothed regional curves. In the three time series there is a strong rise in temperatures observed from 1973 onwards. Also a period of increasing temperatures can be seen between 1901 and 1949 in T_{max} and T_{mean} series, but not in T_{min} series, which retained stable nighttimes temperatures during this episode and indicates that minimum temperatures did not make any contribution to the observed warming. A cold phase centred in the 1970s decade, preceded by an episode of cooling mainly shown in T_{max} and T_{mean}, is also evident in the three time series. Finally, the second half of the 19th century shows a strong cold period centred on the 1880s, where the lowest annual anomaly values of the entire period are evident in the three time series.

A trend analysis over the entire period and several sub-periods of warming/cooling has been performed on an annual and seasonal basis by adapting Sen's (1968) estimator of the slope. The 95% confidence intervals of the trend coefficients have also been calculated from tabulated values (Kendall, 1955). Table XV show annual and seasonal trend coefficients for daily mean, maximum and minimum temperatures together to their corresponding 95% confidence intervals.

Table XV. Annual and seasonal temperature change estimated by a linear trend, and in brackets the associated 95% confidence intervals (in ºC/decade), for daily mean, maximum and minimum temperatures of STS calculated over the entire period and several shorter periods of warming and cooling. Bold (italic) indicates significance at 1% (5%) confidence level.

Periods	1850-2005	1901-2005	1901-1949	1950-1972	1973-2005
Daily mean temperatures					
Annual	**0.10** (0.08/0.12)	**0.13** (0.10/0.16	**0.22** (0.11/0.31)	-0.19(-0.53/0.12)	**0.48** (0.36/0.66)
Winter	**0.10** (0.07/0.14)	**0.14** (0.08/0.20)	0.10 (-0.08/0.32)	0.11 (-0.58/0.68)	0.27 (-0.09/0.56)
Spring	**0.08** (0.05/0.12)	**0.12** (0.06/0.17)	**0.25** (0.06/0.43)	-0.52 (-1.03/0.05)	**0.77** (0.54/0.97)
Summer	**0.09** (0.06/0.11)	**0.13** (0.08/0.18)	**0.23** (0.07/0.38)	-0.29 (-0.71/0.13)	**0.67** (0.41/0.92)
Autumn	**0.10** (0.07/0.13)	**0.12** (0.08/0.17)	**0.26** (0.09/0.42)	-0.08 (-0.57/0.53)	*0.29* (0.02/0.58)
Daily maximum temperatures					
Annual	**0.11** (0.09/0.14)	**0.17** (0.13/0.21)	**0.37** (0.25/0.46)	-0.28 (-0.74/0.16)	**0.51** (0.34/0.66)
Winter	**0.12** (0.09/0.15)	**0.16** (0.10/0.21)	0.18 (-0.02/0.36)	-0.04 (-0.61/0.62)	*0.35* (0.06/0.60)
Spring	**0.11** (0.06/0.15)	**0.17** (0.11/0.23)	**0.37** (0.16/0.60)	-0.62 (-1.38/0.09)	**0.82** (0.53/1.15)
Summer	**0.10** (0.06/0.13)	**0.18** (0.12/0.24)	**0.44** (0.27/0.64)	-0.30 (-0.88/0.17)	**0.73** (0.43/1.04)
Autumn	**0.12** (0.09/0.15)	**0.17** (0.10/0.22)	**0.44** (0.26/0.64)	-0.12 (-0.84/0.70)	0.13 (-0.17/0.47)
Daily minimum temperatures					
Annual	**0.08** (0.06/0.10)	**0.09** (0.06/0.12)	0.08 (-0.02/0.18)	-0.13 (-0.51/0.14)	**0.47** (0.31/0.65)
Winter	**0.09** (0.06/0.13)	**0.12** (0.05/0.19)	0.06 (-0.15/0.24)	0.15 (-0.56/0.78)	0.06 (-0.28/0.62)
Spring	**0.07** (0.04/0.09)	**0.08** (0.03/0.13)	*0.15* (0.01/0.31)	-0.19 (-0.72/0.29	**0.66** (0.46/0.84)
Summer	**0.08** (0.05/0.10)	**0.09** (0.04/0.13)	0.00 (-0.13/0.14)	-0.26 (-0.60/0.08)	**0.62** (0.38/0.93)
Autumn	**0.08** (0.05/0.11)	**0.08** (0.04/0.13)	0.09 (-0.06/0.25)	-0.13 (-0.41/0.33)	**0.43** (0.18/0.77)

Table XV shows a statistically significant warming over the entire period (1850-2005) and over the 20th century, which is picked up by the three regional series both on annual and seasonal scales. Seasonal contributions to annual warming are very similar, although rates of change for winter and autumn are slightly greater for 1850-2005. A somewhat contrasting seasonal contribution to the higher annual warming has been identified over the 1901-2005 period: winter and summer show the greatest contribution, followed by the equinoctial seasons.

Warming in T_{max} is slightly higher than in T_{min} regional series over 1850-2005 and considerably greater during the 20th century. This indicates that daytime temperatures have tended to increase faster than nighttimes temperatures over the period 1850-2005 and particularly over the 1901-2005 period, although in some cases this difference is not statistically significant. However, it is clear over mainland Spain that a larger increase of T_{min} compared to T_{max} has not occurred. This differential diurnal warming at the annual scale has mainly been contributed to by the equinoctial seasons and winter for 1850-2005 and also by spring and autumn together with summer for 1901-2005.

From the three sub-periods identified across the 20th century, only the two warming episodes reach statically significant (at 0.05 level and better) trend coefficients. Annual warming was influenced slightly more by the equinoctial seasons during 1901-1949, followed by similar rates for summer, with winter having the smallest and non-significant contribution. Daytime temperatures contributed more to this increase: T_{max} trends are significant but T_{min} are not (Table XV). A short period of falling temperatures has been identified from 1950 to 1972, in agreement with that recorded at larger global and hemispherical scales [*Jones and Moberg*, 2003]. However, the trends both on an annual and a seasonal basis are non-significant in the three daily temperature series (Table XV). Spring and summer were the seasons with the largest decreases that led to the reduction in annual temperatures. Autumn contributed a little, while winter was the season with only a slight and non-significant positive trend (Table XV).

The latest 1973-2005 episode of accelerated warming is the period that has the highest rates of change among the three sub-periods both on an annual and a seasonal basis and for all daily temperature variables, except for T_{max} in autumns and T_{min} in winters (Table XV). Annual average warming was mainly the result of spring and summer warming; while autumn and winter have contributed less to this warming and, moreover, winter trends do not reach the statistical significance level (0.05) in the T_{mean} series. The larger spring and summer contribution to the increases in annual temperatures in all three Spanish curves highlights the key role played by the warm seasons and should be the focus of further studies on the possible attribution of anthropogenically induced regional warming.

In Table XVI are shown annual and seasonal trends of daily T_{mean}, T_{max} and T_{min} for Madrid calculated over the entire period and several shorter periods of warming and cooling.

Table XVI. Annual and seasonal temperature change estimated by a linear trend, and in brackets the associated 95% confidence intervals (in ºC/decade), for daily mean, maximum and minimum temperatures of Madrid calculated over the entire period and several shorter periods of warming and cooling. Bold (italic) indicates significance at 1% (5%) confidence level.

Periods	1854-2005	1901-2005	1901-1949	1950-1972	1973-2005
Daily mean temperatures					
Annual	**0.10** (0.08/).13)	**0.19** (0.16/0.23)	**0.24** (0.13/0.35)	-0.05 (-0.49/0.34)	**0.44** (0.30/0.62)
Winter	**0.11** (0.08/0.14)	**0.19** (0.14/0.25)	0.04 (-0.14/0.22)	0.14 (-0.31/0.73)	0.13 (-0.14/0.43)
Spring	**0.09** (0.04/0.13)	**0.17** (0.10/0.23)	*0.28* (0.07/0.49)	-0.43 (-0.94/0.23)	**0.71** (0.38/1.04)
Summer	**0.10** (0.06/0.14)	**0.22** (0.16/0.28)	**0.30** (0.11/0.51)	-0.00 (-0.60/0.44)	**0.84** (0.50/1.21)
Autumn	**0.10** (0.06/0.14)	**0.19** (0.14/0.25)	**0.32** (0.16/0.48)	0.11 (-0.55/0.90)	0.13 (-0.28/0.48)
Daily maximum temperatures					
Annual	**0.12** (0.09/0.15)	**0.25** (0.21/0.29)	**0.39** (0.25/0.53)	-0.07 (-0.64/0.31)	**0.44** (0.23/0.62)
Winter	**0.12** (0.08/0.15)	**0.21** (0.16/0.26)	0.09 (-0.08/0.53)	0.08 (-0.44/0.54)	0.12 (-0.05/0.37)
Spring	**0.11** (0.05/0.16)	**0.23** (0.15/0.31)	**0.39** (0.15/0.67)	-0.46 (-1.28/0.23)	**0.75** (0.39/1.16)
Summer	**0.12** (0.07/0.17)	**0.29** (0.22/0.37)	**0.52** (0.24/0.76)	0.03 (-0.60/0.68)	**0.86** (0.48/1.33)
Autumn	**0.11** (0.07/0.16)	**0.25** (0.18/0.33)	**0.48** (0.27/0.72)	0.24 (-0.85/1.08)	-0.02 (-0.44/0.33)
Daily minimum temperatures					
Annual	**0.09** (0.06/0.11)	**0.14** (0.10/0.17)	0.08 (-0.02/0.20)	-0.06 (-0.42/0.31)	**0.48** (0.31/0.64)
Winter	**0.11** (0.07/0.15)	**0.18** (0.10/0.25)	-0.00 (-0.26/0.22)	0.29 (-0.53/0.93)	0.12 (-0.29/0.58)
Spring	**0.07** (0.04/0.11)	**0.11** (0.06/0.17)	0.19 (-0.01/0.37)	-0.20 (-0.76/0.36)	**0.67** (0.39/0.95)
Summer	**0.08** (0.05/0.12)	**0.14** (0.09/0.20)	0.06 (-0.10/0.23)	-0.01 (-0.59/0.38)	**0.84** (0.48/1.11)
Autumn	**0.08** (0.05/0.12)	**0.12** (0.07/0.18)	0.10 (-0.09/0.29)	0.17 (-0.34/0.62)	0.28 (-0.14/0.68)

As can be seen in Table XVI, annual and seasonal trend coefficients for Madrid are quite similar to those estimated over mainland Spain for the entire period and the three 20^{th} century sub-periods. However, Madrid's rates of change throughout the 20^{th} century are slightly higher than over peninsular Spain. As a remarkable aspect of Madrid thermal climate during the recent period of strong warming it can be highlighted that summer and spring daytime temperatures and to lesser extent nighttimes temperatures have been contributing most to the observed annual warming of daily mean temperatures.

A review on other available procedures for developing daily adjusted temperature datasets

In the previous sections, we have described and discussed the complete procedures followed, from data rescue to data homogenization, in order to develop a high-quality and daily adjusted temperature dataset for Spain covering the 1850-2005 period. In this last section, we assess and compare our approach with other procedures adopted for developing several of the daily datasets quoted in section two: the Global Daily Climatology Network (GDCN) dataset (NCDC, 2002), the ECA&D dataset (Klein-Tank *et al*., 2002) and the Australian temperature dataset (Trewin, 1999). Similarities and differences will be highlighted through this comparison exercise in order to better guide the readers in the selection of one or another approach for any specific step in the ongoing process of developing high-quality climate datasets.

The GDCN, developed and updated by the National Climate Data Center (NCDC/NOAA/NESDIS), constitutes a compilation of global climate data into a single data set with a consistent format and composed of daily temperature and precipitation records subjected to a meaningful quality control tests (QC). As no homogenization procedure has been undertaken with these data, we only describe, here, the QC procedures undertaken to these global data and compare with ours.

The GDCN QC procedures run for temperature data and the related metadata can be divided into two phases. The first one mainly comprises simple format checks, in order to identify errors in data format and the existence of values out of range or aberrant values. The second phase is devoted to analyze sets of observations aiming identify potential outliers and incorrect data.

The raw daily T_{max} and T_{min} data are subject to simple format checks (i.e. avoiding invalid characters), duplicate checks (identifying different years or months within the same year/month and different year/month with the same data), bounds checks (i.e. $T_{max} < T_{min}$, or values exceeding world records), streak checks (days with the same value repeated 10 or more consecutive times), gap check (identifying any gap in the frequency distribution of 10 ºC or more by moving both right and left of the median of T_{max} and T_{min} distributions), outlier check (identifying data out of defined ranges by means of a biweight mean and biweight estimate of σ concepts, being the biweight mean defined as the mean estimated out of all the values in a 15 days window centred on the date of the datum and the biweight estimate of σ is the deviation of the biweight mean), manually inspected data (data with known problems recognized without automated inspection) and spatial checks (comparison between monthly averages of T_{max} and T_{min} data to an independently created gridded data set).

The QC undertaken to metadata is also split into two steps: simple format checks (i.e. invalid characters in latitude, longitude, elevation, data source, WMO ID, year, month, day or the existence of the file in the record) and bounds check, as ensuring latitude is between -90.0 and 90.0 degrees, elevation between -408 and 8890 m, year field between 1830 and 2010, or monthly field between 1 and 12).

From this brief description on the QC procedures undertaken to the global records of daily temperatures in the GDCN dataset, several similarities and differences can be appreciated, in addition to the paramount difference in the size and the spatial scale of the GDCN dataset when compared with SDATS. Simple format checks in GDCN, both those undertaken on the data and metadata, are quite similar to those used in SDATS, although in the latter some of the bound checks applied in GDCN were not relevant and then not carried out (i.e. stations' latitude and elevation or values exceeding world records). Tolerance tests carried out within SDATS data are also undertaken in GDCN, although slight variants can be appreciated: a more restrictive threshold in SDATS when looking at repetitions of consecutive identical values has been implemented or the different approach for defining potential outliers. Although the rest of checks employed in GDCN and SDATS are similar, the temporal and spatial coherency tests implemented in SDATS have not been carried out in GDCN. Finally, and as stated before, the main difference between both datasets lies in the fact that GDCN data have not been subjected to any homogeneity test or homogenization procedure as the data are only quality controlled.

The European Climate Assessment & Dataset (ECA&D) contains daily data for several meteorological elements including surface air temperature (T_{max}, T_{min} and T_{mean}) for 42 countries across Europe and the Mediterranean Basin. As in the case of the GDCN dataset, the ECA&D data have been mainly quality controlled, but in contrast to GDCN, the ECA&D data have also been subject to homogeneity testing (Wijngaard *et al*. 2003) in order to condense the records into three categories: useful, doubtful and suspect and, then, to recommend and warn about the use of these flagged time series in climatic analysis.

The QC undertaken to the ECA&D daily series consists of basic control procedures aiming to identify the occurrence of miscoding, like: T_{min} > T_{max}; nonexistent dates; and erroneous outliers, as the series have usually undergone custom QC trial by the supplying institutes. The daily T_{max}, T_{min} and T_{mean} data have to meet a list of criteria, such as T_{max}, T_{min} and T_{mean} must exceed -89.9 ºC and must be less than 60.0 ºC; T_{max} must exceed or equal to T_{min} and T_{mean}, T_{min} must be less or equal to T_{max} and T_{mean}, and T_{mean} must exceed or equal T_{min} and must be less or equal to T_{max}; the daily T_{max}, T_{min} and T_{mean} data must not be repetitive for 5 days; and the daily T_{max} T_{min} and T_{mean} data must be between ± 5 σ of overall average for T_{max} T_{min} and T_{mean} for the inspected day. All daily values labelled by these checks were set to missing values as they were non-correctable mistakes. Finally, the ECA&D time series are coded as "useful" (labelled with 0) for those data that passed QC procedures, as "suspicious" (labelled with 1) when data failed to pass any QC procedure and as "missing" (labelled with 9) when data are absent.

Comparing the QC procedures undertaken for both datasets (ECA&D vs. SDATS) it is apparent that the SDATS gross error checks i and ii, tolerance and internal consistency tests are quite similar in both datasets, despite of small differences in the chosen thresholds (i.e. 4 σ in SDATS instead of 5 σ in ECA&D or four or more consecutive identical values in SDATS instead of five in ECA&D). The main differences between the QC procedures undertaken for both SDATS and ECA&D datasets is that for the latter any temporal and spatial coherency checks have been carried out. Also in the ECA&D dataset gross error checks iii in SDATS were not made.

Finally, for the Australian high-quality daily temperature dataset (Trewin, 1999, henceforth NCCT), a variety of QC tests and, in this case, data correction techniques have been undertaken on the data, which make the Australian exercise more like that undertaken with the SDATS. The Bureau of Meteorology Research Centre has developed several high-quality and adjusted climate datasets at different time scales aiming to monitor long-term trends and climate variability over Australia, which are then transferred to the National Climate Centre (NCC) to maintain and use operationally.

The NCCT data have been subjected to several tests: first, to detect and remove gross single-day errors and, second, to identify and adjust inhomogeneities in the data. The method used in the QC of daily T_{max} and T_{min} were internal consistency tests, such as ensuring that daily maximum temperatures were greater than the minimum temperature on the same day and that the maximum was greater than, and the minimum less than, any available hourly observations on the same day. Besides, T_{max} and T_{min} were checked against pre-defined thresholds depending on station's location as well as against a number of neighbouring stations with similar climates. Furthermore, any data suspected of being accumulated over several days were also removed, as well as the validity of some suspect data tested against the synoptic situation at the time of observation.

These procedures compared to those undertaken in SDATS show mainly similarities rather than differences. Internal consistency checks are the same in both datasets, although in NCCT T_{max} and T_{min} were also checked for consistency against hourly observations. A mix among the SDATS 1 i, 2 ii 5 i tests, which are looking for labelling values exceeding some pre-defined thresholds, has been carried out for NCCT data as well. However, for these data, the thresholds were fixed according to specific ranges for each location and not as fixed thresholds valid for all daily values (i.e. as 1 i. check in SDATS) or depending on the standard deviations defined (4 σ in SDATS). This should return to some extent more labelled as potentially erroneous values due to the finer bounds employed in NCCT. In contrast, NCCT data were not inspected by looking for repetitions of n consecutive identical values or for the gross error iii and the temporal consistency checks carried out in SDATS. However, the NCCT approach of checking the labelled values against the related synoptic situation has not been carried out in SDATS nor in GDCN and in ECA&D datasets, which is likely related to the fact that these later datasets integrate data coming from different networks or sources, meanwhile the NCCT data mainly come from the Australian Bureau of Meteorology (BoM).

As the SDATS, the NCCT data have consistent metadata on site location, instrumentation and observation program stored in the relational database named SitesDb. This will be particularly reliable as it is entered in near-real time by inspectors and managers at the (BoM); however, availability of historical metadata, such as historical changes in observational practices, is more limited. In spite of it, that available in paper form from station history information files are being scanned and incorporated into SitesDb.

The NCCT data have also been homogeneity tested and adjusted. Homogeneity testing was based on comparing each candidate record with a reference series estimated from a weighted mean of highly-correlated nearby time series. Then, a two-phase regression model was employed to identify breakpoints in the difference series between the candidate and reference series. Potential inhomogeneities were visually inspected and the available metadata were used for validating breakpoints in homogeneity. When any breakpoint was determined to be artificial it was corrected by matching the frequency distribution of daily T_{max} and T_{min} on either side of the inhomogeneity, enabling quite diverse adjustments across the distribution. Therefore, the daily temperature records were adjusted for discontinuities at the 5, 10, …, 90, 95 percentile levels, which made temperatures at the higher end of a record's distribution adjusted by different amounts compared to those at the middle or lower end of the distribution.

The homogenization exercise undertaken to the NCCT records show similarities and differences compared with that performed with the SDATS dataset. A similar concept of relative homogenization is behind both procedures, the SNHT scheme applied to the SDATS data and the creation of a difference time series between the candidate record and its group of reference's records employed in NCCT, which makes the homogeneity testing quite similar between both datasets even though there are slightly different statistical techniques employed for detecting breakpoints in the records. A similar strategy for avoiding artificially biased records related to the UHI effects, to limit the number of urban stations in the datasets, has been followed in both datasets, although in SDATS this has also been applied through a trend test for detecting and correcting the records suspected of containing artificial trends while this has not been undertaken in NCCT time series. The largest difference between both procedures lies in the daily adjustment scheme applied to both datasets. In SDATS, a simple daily interpolation technique for avoiding abrupt transitions between months has been applied, while in NCCT rather than making adjustments in mean temperatures different magnitudes of discontinuity were transferred into lower, central and upper classes of the data distributions.

Table XVII, as summary, compare quality control and homogeneity procedures followed at GDCN, ECA&D and NCCT datasets and those undertaken with SDATS.

Table XVII. Comparison among quality control, homogeneity testing and homogenization procedures undertaken with SDATS, GDCN, EAC&D and NCDC daily temperature datasets (see section 4 and 7 for details)

QC procedures							
	Gross error checks	**Tolerance Tests**		**Internal consistency**	**Temporal coherency**	**Spatial coherency**	
SDATS	Values > 50 & < -50 ºC Consistency calendar days Comparison raw/digitised monthly averages	4 consecutive identical values	Values exceeding ± 4 σ	$T_{max} < T_{min}$	25 ºC exceedence consecutive values	Difference time-series candidate vs references exceeding ± 4 σ	Visual comparisons candidate references
GDCN	values < -89.4 & > 57.8 ºC Invalid characters & Consistency calendar days Bounds coordinates, elevation, period, Duplicate checks	10 consecutive identical values	Gap check & outlier check	$T_{max} < T_{min}$	-	-	-
ECA&D	values < -89.9 & > 60 ºC Miscoding & Consistency calendar days	5 consecutive identical values	outlier check	$T_{max} \geq T_{min}$ & T_{mean}, $T_{min} \leq T_{max}$ & T_{mean}, $T_{mean} \geq T_{min}$ & $T_{mean} \leq T_{max}$	-	-	-
NCCT	-	Values exceeding locational depending thresholds		$T_{max} > T_{min}$ & > hourly obs, $T_{min} < T_{max}$ & < hourly obs,	-	comparisons neighbour stations	
Homogeneity testing							
SDATS	SNHT (Alexandersson and Moberg, 1997) application on monthly, seasonal and annual data						
GDCN	-						
ECA&D	SNHT (Alexandersson, 1986), Buishand range test (Buishand, 1982), Pettitt test (Pettitt, 1979) and Von Neumann ratio test (Von Neumann, 1941)						
NCCT	A two-phased regression model based on that used by Solow (1987) and Easterling and Peterson (1995)						
Daily data homogenization							
SDATS	"Screen bias" minimization, SNHT application on a monthly basis & interpolation to daily scale (Vicent *et al*. 2002)						
GDCN	-						
ECA&D	-						
NCCT	Adjustments matching frequency distribution at daily scale (Trewin & Trevitt, 1996)						

Glossary

Artificial Trend: Break in homogeneity of a climate time series affecting any time slice or the entire record related to any gradual bias incorporated in the data from a non-climatic factor.

Breakpoint: Starting point of any inhomogeneity in a record.

Candidate Record: Record to be homogenized

Data Adjustment: Amendment applied to the data in order to improve their homogeneity

Data and Metadata Sources: Places where climatic data and related information about the data (metadata) can be found.

Dataset: Compilation of climate data belonging to a number of stations, which expands over a period of time and contains information for one or several meteorological elements.

Homogeneity Assessment: Procedure undertaken using raw climate time series to detect potential homogeneity breaks in climate records.

Homogeneous Time Series: Climate records that all their values are consistent and only the result of the vagaries of weather and climate.

Homogenization Procedures: set of techniques applied to the raw data in order to make homogeneous records by minimizing or removing artificial biases.

High-quality and Homogeneous Datasets: Groups of climate data in which a set of procedures aiming to ensure data quality and homogeneity have been undertaken on the raw data.

Internal Consistency Test: Part of a quality control procedure run on the raw climate data aiming to ensure the tested values are coherent.

Metadata: Information about the data that aims to document how, when, where and who have recorded meteorological observations.

Potential Inhomogeneity: Break in homogeneity of any climate record detected by any homogeneity test that has to be validated as physically plausible.

Quality Control (QC): set of procedures used to detect erroneous meteorological observations.

Raw Gross Error Checks: Part of a QC procedure aiming to label aberrant or impossible values or inconsistencies within recorded data and calendar dates.

Reference Stations: Group of records employed for testing homogeneity of the candidate record.

Regional Temperature Series: Climatic time series calculated from a number N climatic records registered within a region or any space in order to represent regional climate behaviour.

Relative Homogenization Assessment: Procedure for detecting and correcting breaks in homogeneity of a climatic time series by using other high-correlated time series as reference for the candidate station to be homogenized.

Screen Bias: Artificial effect in any temperature series associated with temporal changes in thermometric exposures.

Spatial Coherency Tests: Part of a QC procedure aiming to label suspicious values in a climate time series by comparing the target data against a set of neighbouring and high-correlated records.

Temporal Coherency Test: Part of a QC assessment planning to test consistency within consecutive observations

Tolerance Tests: Part of a QC procedure to label values exceeding pre-defined thresholds.

Urban Bias: Artificial trend present in a climate time series related to urbanization influences.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidelines on climate metadata and homogenization. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization, Geneve.

Aguilar E, Brunet M, Saladié O, Sigró J, López D. 2002. Hacia una aplicación óptima del Standard Normal Homogeneity Test para la homogeneización de series de temperatura. In *La información climática como herramienta de gestión ambiental*, VII Reunión Nacional de Climatología, Grupo de Climatología de la AGE, Cuadrat JM, Vicente SM, SAZ MA. (eds.). Universidad de Zaragoza: Zaragoza; 17-33.

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661–675.

Alexandersson H, Moberg A. 1997. Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology* **17:** 25-34.

Almarza C. 2000. Respuesta al calentamiento global de la serie de temperatura media anual de Madrid. *Actas de la II Asamblea Hispano-Lusa de Geodesia y Geofísica, Lagos (Portugal)*, Sociedad de Geodesia y Geofísica, Madrid.

Angot A. 1903. Instructions météorologiques, 4ème edition. Gauthier-Villars, Imp-Libraire: Paris.

Auer I, Böhm R, Jurkovic A, Orlik A, Potzmann R, Schöner W, Ungersböck M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones PD, Efthymiadis D, Mestre, O, Moisselin J-M, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajic-Capka M, Zaninovic K, Majstorovic Z, Szalai S, Szentimrey T, Mercalli L. 2007. HISTALP – Historical Instrumental Climatological Surface Time Series of the Greater Alpine Region. *International Journal of Climatology*, **27** (**1**),: 17-46.

Barriendos M, Martin-Vide J, Peña JC, Rodriguez R. 2002.Daily Meteorological Observations in Cádiz – San Fernando. Analysis of the Documentary Sources and the Instrumental Data Content (1786–1996), *Climatic Change*, **53**, 151-170.

Begert M, Schlegel T, Kirchhofer W. 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology* **25:** 65-80.

Böhm R, Auer I, Brunetti M, Maugeri M, Nanni T, Schöner W. 2001. Regional temperature variability in the European Alps: 1760 1998 from homogenized instrumental time series. *International Journal of Climatology* **21:** 1779-1801.

Brohan P, Kennedy JJ, Haris I., Tett SFB, Jones PD. 2006. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophysical Research,* **111**, D12106, doi: 10.1029/2005JD006548

Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, Moberg A, Walther A, Lister D, López D, Almarza C. 2006a. The development of a new daily adjusted temperature dataset for Spain (1850-2003), *International Journal of Climatology,* **26**, 1777-1802, DOI: 10.1002/joc.1338

Brunet M, Jones P, Sigró J, Saladié O, Aguilar E, Moberg A, Della-Marta PD, Walther A, Lister D, López D. 2006b. Temporal and spatial temperature variability and change over Spain during 1850-2005, *Journal of Geophysical Research-Atmosphere,* **112***, D12117*, doi:10.1029/2006JD008249,

Brunetti M, Maugeri M, Monti F, Nanni T. 2006. Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *International Journal of Climatology*, **26**: 345-381.

Buishand TA. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**: 11–27.

Camuffo D, Jones PD. 2002. *Improved understanding of past climatic variability from early daily European instrumental sources*. Kluwer Academia Publishers: Dordrecht/Boston/London.

Easterling DR, Peterson TC. 1995. A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol*., **15**, 369-377.

Feng S, Hu Q, Qian W. 2004. Quality control of daily meteorological data in China, 1951-2000: a new dataset. *International Journal of Climatology* **24:** 853-870.

Folland CK, Karl TR, Christy JR, Clarke RA, Gruza GV, Jouzel J, Mann ME, Oerlemans J, Salinger MJ, Wang SW. 2001. Observed climate variability and change. In *Climate Change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of Intergovernmental Panel on Climate Change*, Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA. (eds.). Cambridge University Press: Cambridge, 99-182.

González FJ. 1992. *El observatorio de San Fernando (1831-1924)*. Serv. Publ. Ministerio de Defensa: Madrid.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA. (eds.) 2001. IPCC, 2001: Climate Change 2001: The scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 881pp.

ICM (Instituto Central Meteorológico) 1893. *Treinta años de observaciones meteorológicas: exposición y resumen de las efectuadas en el Observatorio de Madrid desde el 1º de enero de 1860 al 31 de diciembre de 1889*. Establecimiento Tipográfico de los Sucesores de Cuesta: Madrid.

Jones PD, Moberg A. 2003. A hemispheric and large-scale surface air temperature variations: an extensive revision and update to 2001. *Journal of Climate* **16:** 206-223.

Jones PD, Osborn TJ, Briffa KR, Folland CK, Horton B, Alexander LV, Parker DE, Rayner NA. 2001. Adjusting for sampling density in grid-box land and ocean surface temperature time series. *J. Geophys. Res.* **106**, 3371-3380

Jones PD, Hulme M. 1996. Calculating regional climatic time series for temperature and precipitation: methods and illustrations. *International Journal of Climatology,* **16**: 361-377.

Kendall MG. 1955. *Rank correlation methods*. Charles Griffin and C: London, UK.

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 G, Tzanakou M, Salía S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, Van Encelen AFV, Forland E, Mietus 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.

Manley G. 1974. Central England Temperatures: monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society* **100:** 389-405.

Martín-Vide J, Olcina J. 2001. *Climas y tiempos de España.* Alianza Editorial: Madrid.

Moberg A, Jones P.D, 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, Almarza C, Auer I, Barriendos M, Begert M, Bergström H, Böhm R, Butler CJ, Caesar J, Drebs A, Founda D, Gerstengarbe FW, Micela G, Maugeri M, Österle H, Pandzic K, Petrakis M, Srnec 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 analysed for the period 1901-2000, *Journal of Geophysical Research-Atmospheres*, **111**, in press, doi: 10.1029/2006JD007103.

NCDC (2002): Data documentation for data set 9101, Global Daily Climatology Network, version 1.0, 26 pp. (available online http://www.ncdc.noaa.gov/oa/climate/ climateinventories.html

Nicholls N, Tapp R, Burrows K, Richards D. 1996. Historical thermometer exposures in Australia. *International Journal of Climatology* **16:** 705-710.

Parker DE. 1994. Effects of changing exposure of thermometers at land stations. *International Journal of Climatology* **14:** 1-31.

Parker DE, Legg TP, Folland CK. 1992. A new daily Central England Temperature series, 1772-1991. *International Journal of Climatology* **12:** 317-42.

Peterson TC, Daan H, Jones PD. 1997. Initial selection of a GCOS Surface Network. *Bull. Amer. Meteorol. Soc*. **78**: 2837-2849.

Peterson TC, Vose R. 1997. An overview of the Global Historical Climatology Network temperature data base. *Bulletin of the American Meteorological Society* **78:** 2869-2872.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Böhm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review, *Int. J. Climatol*., **18**, 1493-1517.

Pettitt AN. 1979. A non-parametric approach to the change-point detection. *Applied Statistics* **28**: 126–135.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003: Globally complete analyses of sea surface temperature, sea ice and night marine air temperature, 1871-2000. *J. Geophysical Research* **108**, 4407, doi:10.1029/2002JD002670

Rayner NA, Brohan P, Parker DE, Folland CK, Kennedy JJ, Vanicek M, Ansell T, Tett SFB. 2006. Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J. Climate*, **19**, 446-469

Rico Sinobas M. 1857. *Resumen de los trabajos meteorológicos correspondientes al año 1854 verificados en el Real Observatorio de Madrid*. Imprenta de Eusebio Aguado: Madrid.

Sen P K. 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, **63**: 1379-1389.

Solow A. 1987. Testing for climatic change: an application of the two-phase regression model. *J. Climate Appl. Meteorol.*, **26**, 1401-1405.

Tan LS, Burton S, Crouthamel R, van Engelen A, Hutchinson R, Nicodemus L, Peterson TC, Rahimzadeh F. 2004. Guidelines on Climate Data Rescue. WMO/TD 1210, WCDMP-55, Geneva: 12 pp.

Trewin BC. 1999. The development of a high quality daily temperature data set for Australia, and implications for the observed frequency of extreme temperatures. In: *Meteorology and oceanography at the new millennium: AMOS 1999*. Proceedings of the Sixth National Australian Meteorological and Oceanographic Society Congress, Canberra, 1999.

Trewin BC, Trevitt ACF. 1996. The development of composite temperature records. *Int. J. Climatol.*, **16**, 1227-1242.

Vincent LA, Gullet DW. 1999. Canadian historical and homogeneous temperature datasets for climate change analyses. *International Journal of Climatology* **19:** 1375-1388.

Vincent LA, Zhang X, Bonsal BR, Hogg WD. 2002. Homogenization of daily temperatures over Canada. *Journal of Climate* **15:** 1322-1334.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* **13**: 367–395.

Wijngaard JB, Klein-Tank AMG, Könen GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series, *International Journal of Climatology*, **23 (6)**: 679-692

Acknowledgements

We thank Carlos Almarza from the Servicio de Desarrollos Climatológicos (Climatological Branch) of the INM (Spanish Met Office) for the provision of much of the raw data. Rob Allan and Tara Ansell helped the authors locate valuable sources of daily Spanish data for the 19th century in the UK Met Office Archives. Manolo Bañón from the Territorial Centre of INM at Murcia assisted in the provision of digitized raw data from Murcia station. ROASF staff helped with updating recent (1996-2003) daily data. Francisco García from the Territorial Centre of INM at La Coruña helped us on locating substantial metadata. Finally, Mariano Barriendos provided daily temperature data for the last third of the 19th century for Barcelona. The work was supported by the EU EMULATE project (EVK2-CT-2002-00161) and the Spanish SCREEN project (CICYT: REN2002-0091/CLI).

Global Climate Observing System - Essential Climate Variables (GCOS ECVs)

Domain	Essential Climate Variables (ECVs)
Atmospheric (over land, sea and ice)	**Surface:** Air temperature, Precipitation, Air pressure, Surface radiation budget, Wind speed and direction, Water vapour. **Upper-air:** Earth radiation budget, Upper-air temperature, Wind speed and direction, Water vapour, Cloud properties. **Composition:** Carbon dioxide, Methane, Ozone, Other long-lived greenhouse gases, Aerosol properties.
Oceanic	**Surface:** Sea-surface temperature, Sea-surface salinity, Sea level, Sea state, Sea ice, Currents, Ocean colour, CO_2 partial pressure. **Sub-surface:** Temperature, Salinity, Currents, Nutrients, Carbon, Ocean tracers, Phytoplankton.
Terrestrial	River discharge, Water use, Ground water, Lake levels, Snow cover, Glaciers and ice caps, Permafrost and seasonally-frozen ground, Albedo, Land cover, Fraction of absorbed photosynthetically active radiation, Leaf area index, Biomass, Fire disturbance.