Coordinator of text: Enric Aguilar

(Contributions by : Enric Aguilar, Inge Auer, Manola Brunet, Thomas C. Peterson, and Jon Wieringa)

Edited by: Paul Llansó

WMO/TD No. 1186

THE WCDMP "GUIDELINES" SERIES

In recognizing the need for National Meteorological Services (NMSs) to improve their climate data and monitoring services, the Commission for Climatology (CCl) and the CCl Management Group placed a high priority on the distribution of guidelines for the NMSs.

Within the World Climate Data and Monitoring Programme, a meeting was held at the kind invitation of Spain (Malaga, 24-26 February 2003) in which a number of experts in the two CCl Open Programme Area Groups (OPAGs) on Climate Data and Monitoring initiated the preparation of guidelines on metadata and data homogenization, observation networks and systems, and data rescue. The participants were either members of an Expert Team of CCl, or were invited experts.

The Guidelines on Climate Metadata and Homogenization are meant to be easy to read and refer to, well illustrated, and not bulky. They provide information and assistance on how to organize and implement climate services, and present processes and technological solutions that attempt to address the special situation and needs of smaller NMSs which have limited resources.

The review of the Guidelines was the first such activity that was done within the CCl OPAG structure, so that all CCl Members were given an opportunity to review and comment, as well as to see the progress being made in the OPAGs. It was drafted by a sub-group of the CCl Expert Team on Metadata for Climate Applications, circulated for contributions and comment among the members of the CCl Expert Team, and posted to the OPAG's web site for review and comment by the members of CCl.

It should be kept in mind that this Technical Document, like the other technical documents published under the WMO WCDMP series, is intended to provide guidance in the form of best practices that can be used by Members. Because of the diversity of NMSs with respect to size and stage of technological development along with the variability of weather types and climate, some practices may not have significant utility for specific Member. However, this document does cover a wide range of guidance that should provide some form of assistance to every Member.

GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION

by Enric Aguilar[1], Inge Auer[2], Manola Brunet[1], Thomas C. Peterson[3] and Jon Wieringa[4].

1: Climate Change Research Group, University Rovira i Virgili, Tarragona, Spain
2: Central Institute for Meteorology and Geodynamics, Vienna, Austria
3: NOAA/National Climatic Data Center, Asheville, North Carolina, USA
4: Department Environmental Sciences, Wageningen University, Wageningen, Netherlands

1 RATIONALE

If we measure rainfall, in order for the data to be useful for future users, we also need to document where and how the measurements were made. Station documentation is information about the data or data about the data: metadata. The word metadata is made up by the addition of the Greek "meta" (beyond) and the Latin "datum" (a given fact). Metadata should reflect how, where, when and by whom information was collected. Ideally, a complete metadata should register all the changes a station has undergone during its lifetime, composing what is called the station history.

Supplementary information about the observations, such as type of instrument or exposure, can provide additional insights into interpreting the observed quantities. Sometimes when the instruments change, the observations will show an artificial increase or decrease. Such a jump in the measured amount is an example of an inhomogeneity and adjustments to these data are often applied to account for the effects of the inhomogeneity. If a long-term time series is homogeneous, then all variability and change is due to the behaviour of the atmosphere.

Every user and provider of climate data has to deal with metadata and homogeneity to some extent. Many climate researchers throughout the world have developed effective approaches for dealing with the many aspects of metadata and homogeneity. The following document is based on their collective experience and should and is intended to offer guidance to the NMHSs on these matters.

2 METADATA

Good metadata are needed to ensure that the final data user has no doubt about the conditions in which data have been recorded, gathered and transmitted, in order to extract accurate conclusions from their analysis. The knowledge of the exact date and time when a thermometer was replaced and the technical characteristics of the new and the old instrument, will surely help to remove the non-climatic fingerprint of this change in that particular temperature record. High quality and homogeneous long-term datasets are needed to assess climate related issues.

Metadata have a key role in the process of creating such datasets, as the knowledge of the station history provides increased confidence in the statistical techniques employed to ensure that the only variations that remain in a climate time series are due to actual climate variability and change. Meteorological data users other than the climatological community, working in fields like agrometeorology, engineering or aeronautics, also benefit from good metadata. These professionals also need to extract the maximum accuracy from the observations, and often compare data taken in different places or times. A complete knowledge of the measuring conditions will help them to achieve this goal.

Meteorological data are influenced by a wide variety of observational practices. Data depend on the instrument, its exposure, recording procedures and many other factors. There is a need to keep a record of all these metadata to make the best possible use of the data. This guide will identify the minimum information that should be known for all types of stations, like, for example, location and measurement units. Additional information will be of great advantage for the data users, as well as for the providers. A full list of metadata, which ideally should be stored, is discussed in this guide and a best practice list is included. Complete metadata describe the history of a station since its establishment to the present and hopefully onwards to the future. Most metadata have to be derived from the station's documentation, both from current and historical documents, while some others can be obtained from the data themselves. In order to provide high quality datasets, it is crucial to maintain comprehensive station documentation and to keep it updated. Establishing the necessary procedures to ensure that all metadata needs are taken into account should be done jointly by the station keepers and the network managers.

A good metadata archive helps the NMHSs in asset management and other administrative procedures, as data existences and observing conditions are kept in good order. It also can be said that good metadata helps society to gain a better understanding of weather and climate related processes, as well as climate change. For all these reasons, WMO has a strong interest in encouraging metadata recording for current measurements, in supporting metadata recovery efforts and encourages NMHSs to not only accomplish the minimum requirements, but also try to meet the best practices recommended in this guide. The following sections list the different metadata items which NMHSs should try to store for every station.

The importance and necessity of metadata can also be understood by quoting one of the GCOS Climate monitoring principles: "The details and history of local conditions, instruments, operating procedures, data processing algorithms and other factors pertinent to interpreting data (i.e. metadata) should be documented and treated with the same care as the data themselves." (WMO 2002).

Sections 2.1 to 2.5 will focus on single stations metadata, meanwhile section 2.6 will touch upon the subject of historical networks metadata. Final section on this chapter, 2.7, will discuss metadata storage and access.

2.1 STATION IDENTIFIERS AND GEOGRAPHICAL DATA

Data can always be associated with some place. The first thing the user has to be informed about is where in the world this place is. To do so, the station has to be identified by names and codes and to be located into the geographical network. It is also important to clearly identify when data started to be collected and by whom. Carefully reporting this information and all the changes it may endure is a minimum metadata requirement.

2.1.1 **Station Identifiers**

- *Name*: station names usually refer to the city or village where the data are collected. If a district or town has several stations, it is important to extend the name, (e.g. Bigtown Airport, or Bigtown University.) to leave no doubt about which station the data belongs to. Network managers should avoid selecting names derived from the metadata themselves (Bigtown Wind Station), names that would be expected to change (e.g. business names) or references to cardinal points (e.g. Bigtown North, etc) which accuracy might be affected by city growth.
- *Aliases*: sometimes stations can be known by more than one name. This can happen when, for example, the city where the station is located renamed after a political change and the former name remains in old databases or when a different name is used to refer to the combination of two neighbouring stations. It is important to identify in the metadata the different names (or aliases) a station may have.
- *WMO Code or station number*: worldwide, WMO identifies meteorological stations whose data are internationally exchanged with a 5-digit code. The WMO code identifies universally a single record, although it might have relocated during its history. Its two first digits give information on the region of the world and the country where the station is located.
- *Station number or code in other networks*: for different reasons, not all the meteorological stations in the world have an assigned WMO code. Many other stations have a national or local code which has nothing to do with the international network, but which identifies the station nationally, regionally or in specific purpose networks, like air pollution or agrometeorology networks. Hence, it will be useful to keep the national code number as well.

- *Opening/Closing dates*: identifies when the station referred by a given name and code start its operational period and closed (if applicable)
- *Type of station*: synoptical, aeronautical, agrometeorological, etc.
- *Station information contact*: metadata should provide details on where to obtain more information about the station (name, address, telephone, fax and e-mail) and identify the institution responsible person for measurements in case of a currently open station or for archiving the data in case of non-operating sites. It is useful to have on record the responsible institution both formally (e.g. Office of Education) and also practically (e.g. some particular school).

Metadata on any modifications in names and codes, type of station, etc. also have to be very carefully documented.

2.1.2 **Geographical Data**

Climate data are associated with geographical locations and the following station attributes in particular need to be managed:

- *Latitude and longitude*: preferably with sufficient accuracy that the station is located within a few hundred meters, e.g. in units of 0.001 degree of latitude. When reporting latitudes and longitudes two very important network-wide decisions have to be made to avoid problems. Firstly, the operator has to decide how to report the fractions of degrees: in minutes and seconds or in decimal form. The second important issue is to clearly distinguish between hemispheres and report differently north and south latitudes and east and west longitudes. This can be done by using letter suffixes or prefixes for North (N), South (S), East (E), and West (W). Alternatively, it is suggested the use of signs instead of letters to differentiate hemispheres. This may help data processing and data representation, as many program codes and standard software packages will only deal with numerical coordinates. Although it is easy to convert from one form to the other, it is very important to adopt a permanent criterion and document it in the metadata to avoid errors that can prevent or compromise data analysis. For longitude, it should be also clearly indicated when the 0 meridian is the Greenwich Meridian.
- *Elevation above Mean Sea Level*: elevation of the ground in the station enclosure above sea level has to be reported with an accuracy of several meters or better. If pressure is observed, the barometer will not usually be located in the enclosure and the elevation of the pressure reference level should be specified separately.
- *Relocations*: when any of the location parameters are changed, because the entire station or some individual instrument are moved – even a short distance - or because more accurate measurements have become available, it is very important to report this in the metadata, including the new location parameters and the exact time of the change.

2.2 LOCAL ENVIRONMENT

Coordinates and elevation are not enough information to document a meteorological station. Data are influenced by factors that act at several scales. Mesoscale (1 km to 30 km) climate is influenced by proximity and size of large water surfaces, urbanized areas and mountain ranges; at toposcale ("local" scale, 300 m. to 2 km) the observations are influenced by terrain slope, forests, crops and other roughness, and by nearby obstacles such as trees or houses (at airports: airplanes); at microscale (less than 300 m) the minimum temperature is affected in "frost hollows", the surface energy exchange is influenced by wetness and thermal conductivity of the earth, and reliable radiation measurements depend on an obstacle-free horizon.

Basic requirements for station environment documentation are:

- Updated mapping in some form of the mesoscale region at ~ 1: 100 000.
- Toposcale map (~ 1: 5000), updated each year, as specified by the WMO Technical Commission for Instrument and Methods of Observation (WMO, 1996) (see *Figure 1*)
- Radiation horizon mapping, updated each year (see *Figure 1*).
- Photos, taken from all points of the compass and enough distance, of the enclosure and of instrument positions outside the enclosure, updated upon significant changes.
- A microscale map of the instrument enclosure, updated when individual instruments are relocated or other significant changes occur. The instrumentation section below lists further necessary microscale information.

2.2.1 **Local land use/land cover:**

At different scales, it is recommended to keep track of several attributes. At the mesoscale (1 km to 30 km) it is important to account in the metadata for:

- Proximity and size of large water surfaces,
- Urbanized areas
- Mountain ranges.

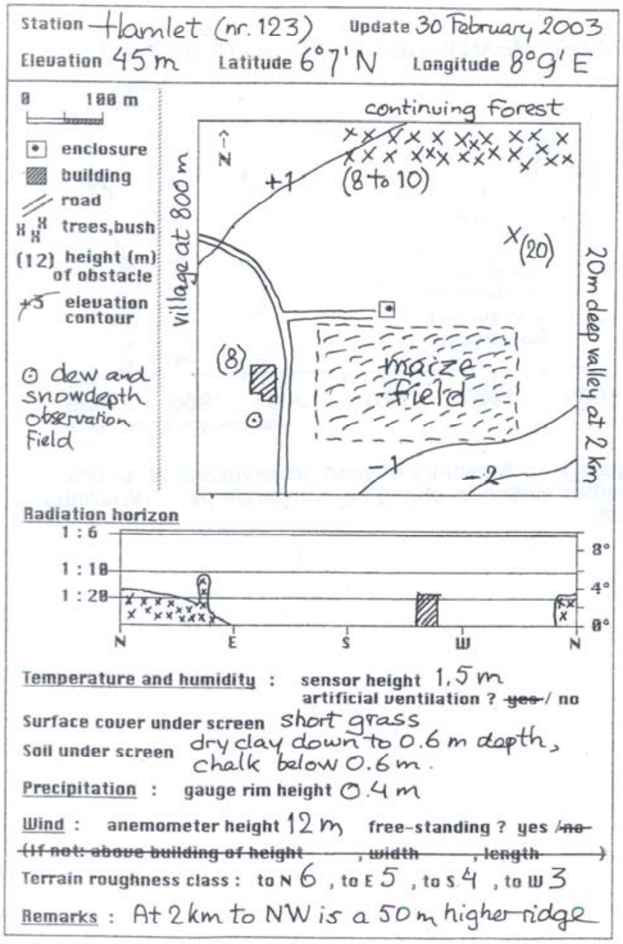
To document the mesoscale situation, good geographical maps (not just road-maps, but showing elevation contour lines and land use) at a scale of 1:100000 or better are a good information source. Such maps are usually in existence and may be even accessible to the public, but may require updating in case of fast growth of a nearby urban area or gradual change to another type of land use. Satellite photos, or acquirable air reconnaissance photography, are worthwhile assets. For the smaller scales, the NMHS should investigate and record the toposcale and microscale environment of its stations.

At toposcale ("local" scale, 100 m to 2 km) observations are influenced by:

- Terrain slope, both steepness and direction
- Forests, crops and other roughness
- Nearby obstacles such as trees or houses (at airports: airplanes)
- Proximity to irrigation.

At the toposcale only the location of the entire enclosure and the land use/land cover of its nearest surroundings must be shown.

The mapping example in *Figure 1* (1: 5000, WMO, 1996) shows what kind of terrain features must be marked. Because trees grow, and buildings arise or are turn down, it is advisable to repeat such toposcale mapping every few years (and to archive old maps). The terrain description as a whole should be sufficient to assign azimuth of 30º to 45º width surface roughness using Davenport's effective roughness classification (see ***Table 1***). Analysis of sunshine duration or global radiation requires information about the radiometer horizon, which can also be depicted on the CIMO-template.



Station Hamlet (nr. 123) Update 30 February 2003

Elevation 45 m Latitude 6°7' N Longitude 8°9' E

Temperature and humidity : sensor height 1.5 m

artificial ventilation ? ~~yes~~ / no

Surface cover under screen short grass

Soil under screen dry clay down to 0.6 m depth, chalk below 0.6 m.

Precipitation : gauge rim height 0.4 m

Wind : anemometer height 12 m free-standing ? yes / ~~no~~

~~(If not: above building of height , width , length)~~

Terrain roughness class : to N 6 , to E 5 , to S 4 , to W 3

Remarks : At 2 km to NW is a 50 m higher ridge

Figure 1: CIMO-Guide Template for documenting metadata at the toposcale.
The example shows an imaginary station (Hamlet).

Table 1: Davenport classification of effective terrain roughness

No.	Class Name	Roughness length (m)	Landscape description
1	Sea	0.0002	Open water, featureless flat plain, fetch > 3 km
2	Smooth	0.005	Obstacle-free land with negligible vegetation, marsh, ridge-free ice
3	Open	0.03	Flat open grass, tundra, airport runway, isolated obstacles separated by >50 obstacle heights H;
4	Roughly Open	0.10	Low crops or plant cover, occasional obstacles separated by ≥ 20 H
5	Rough	0.25	Crops of varying height, scattered obstacles with separation x ≈ 12-15 H if

			porous (shelterbelts) and x ≈ 8-12 H if solid (buildings)
6	Very Rough	0.5	Intensively cultivated landscape with large farms, orchards, bush land, x ≈ 8 H; low well-spaced buildings and no high trees (x ≈ 3-7 H)
7	Skimming	1.0	Full similar-height obstacle cover with interspaces ≈ H, e.g. mature forests, densely-built town area
8	Chaotic	≥2	Irregular distribution of very large elements: high-rise city centre, big irregular forest with large clearings

To report in the maps or other formats such as digital databases, the station's environment land use and cover, the following list gives a useful classification that covers most cases.

- *Artificial surfaces*: continuous urban cover; discontinuous urban cover; industrial and commercial areas; transportation infrastructures; harbour areas; airports; mines, dumps and areas under construction; artificial green areas (non agricultural).
- *Agricultural surfaces*: non irrigated crops; irrigated crops; rice fields and other inundated crops; grasslands; mixed crops; agricultural-forest systems
- *Natural vegetation and open areas*: deciduous forests; evergreen forests; mixed forest; shrub vegetation; mixed shrub and forest; natural grasslands and prairies
- *Wetlands*: swamp areas; peat lands; marshes; inter tidal flat areas
- *Water bodies*: rivers and other natural water courses; artificial water courses; lakes and lagoons; dams; estuaries; seas and oceans

2.2.2 **Instrument exposure**

An instrument's exposure is affected by several microscale factors (less than 100 m) like obstacles and ground cover. It is important to register in the metadata:

- *Obstacles*: usually, the instruments are located in an enclosure of a few hundred square meters. That enclosure may be surrounded by a wall, or high trees, or the observer's house or other buildings may be near. For example, wind in particular is very obstacle-sensitive, a 10m high tree still giving measurable shelter at a few hundred meters distance (see ***Figure 2***). For correction of such shelter effects, the metadata required is the surrounding effective roughness, at least from Davenport-class estimates (see *Table 1*). Alternatively, surrounding roughness can be objectively derived from azimuth-dependent analysis of observed gustiness data (Verkaik, 2000), if we have metadata on the wind instrumentation response.
- *Ground Cover*: the type of ground cover underlying the meteorological station affects the data measurement because different surfaces have different properties (e.g. roughness, albedo, thermal capacity) and the surface energy exchange is influenced by wetness and

thermal conductivity of the earth**.** For these reasons, it is important to note in the metadata if the instruments are mounted over pavement, vegetation, etc, as well as soil composition (e. g. pavement, sand, chalk, clay) down to at least 1 m.

Obstacles and ground cover affect the representativity of the station measurements for the regional state of the atmosphere which often has to be improved by well-developed corrections. To develop such corrections, the first step is to know the station layout. Photography and video-recording can document this if the pictures are taken at sufficient distance - not just a close-up of the thermometer screen, but pictures taken at about 20 meters distance from six or eight directions, showing screen and enclosure in their surroundings. Pictures and/or videos should be dated, and new pictures are recommended when the surroundings change, or when any instrument is relocated.

Similar photos should document the exposure of all instruments not placed in the enclosure. The microscale exposure information must be completed with a map of the enclosure (about 1: 500, dated), showing individual instrument locations and also obstacles (with their height) and ground cover, vegetation or pavement.

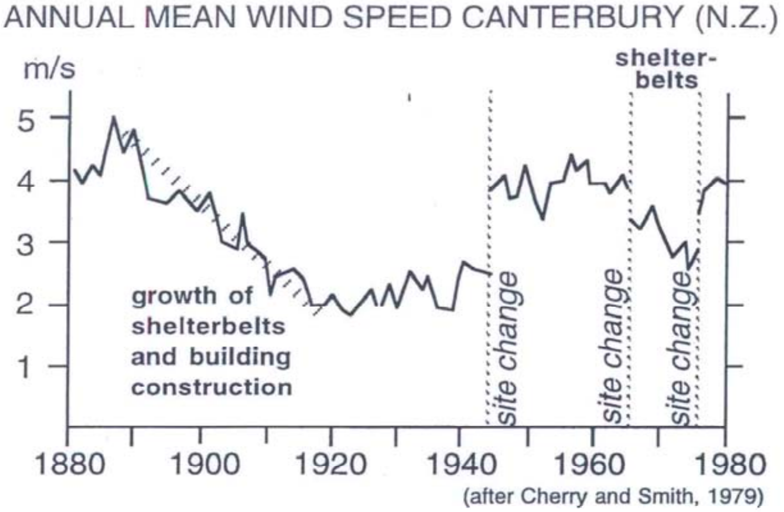


Figure 2: A hundred years long wind data series from an official rural station (Canterbury, New Zealand) spoiled by gradual changes in toposcale surroundings (Source: J.Wieringa, 23rd AMS Conf. on Agricultural and Forest Meteorology, Albuquerque 1998, p.9-12).

2.3 INSTRUMENTATION

2.3.1 **Type of instruments**

It is very important to document the kind of instrument the measurements are taken with. Good metadata should register the following items:

- Instrument manufacturer
- Model of instrument, with size and identification
- Output type and sensitivity
- Transducer type (if applicable)
- Response time (if applicable)

Model identification is insufficient, because manufacturer leaflets are often incomplete. Most of these metadata can be adequately documented by a technician upon installation at the station, and thereafter only require updating in case of change or replacement of instruments.

Depending on each meteorological element, some additional instrument features are very important:

- *Temperature and humidity*: screen (type and size) and ventilation
- *Wind direction*: time and method of azimuth alignment.
- *Wind speed*: response time of anemometer and recording chain, and how these where determined
- *Precipitation*: gauge rim diameter, rim height above ground, presence of overflow storage, presence of a nipher screen or other airflow-modifying feature, presence of heating or other means to deal with solid precipitation
- *Global radiation*: wavelength range transmitted by the dome.
- *Sunshine:* thresholds for automatic sunshine recorders.
- .
- *Evaporation*: any coverage applied to evaporation pan.

Changes in instruments can have a big impact on data. For example, Russian temperatures tend to have fewer values below -40ºC prior to 1900. It seems freezing of mercury in the employed thermometers influences them. When there is a change in instrumentation it is crucial to note in the metadata the exact date of the replacement, because these modifications do have an impact on data homogeneity. Instrument renewal documentation is generally centralized for economic reasons, but a copy of the appropriate documents should be additionally held at the station archive.

2.3.2 Instrument mounting and sheltering

Changes in instrument mounting and sheltering can also have a huge impact on data homogeneity so need to be carefully described. Some relevant aspects on instruments mounting should be made available in the metadata. This includes:

- Height above surface
- Description of shelter, if applicable
- Degree of interference from other instruments and objects, such as an artificial heat source or a ventilator.

Description of the instrument mounting should be provided by the station observer or manager. Blockage of radiation by a mast, the length of an anemometer boom and the degree of airflow freedom around vane and anemometer, the distance of the thermometer screen to possible heat sources, should all be noted. Moreover the soil composition and wetness should be known, because soil thermal conductivity may make a difference of several degrees in nocturnal temperatures at screen height.

As addressed in previous sections, images help to identify possible problems. In the case that several instruments are mounted on a single mast, a picture of the array should complete the metadata.

2.3.3 Data recording and transmission

When a meteorological element is measured with an instrument, data have to be recorded and usually transmitted to the data management section of the organization for checking and archival. Knowing the particular procedures involved may help the user identify potential data constraints and/or problems. Metadata should include:

- *Type of recording*: information about scale units and resolution, range of recorded variations, response time and/or sampling time, averaging period if applicable. For example, it makes a difference if the official maximum and minimum temperatures are 5-second averages or 5-minute averages. Shorter averaging times imply most of the times recording higher maximum and lower minimum temperatures, as the instrument will account for short-term fluctuations.
- *Signal transport*: signal type, type and location of any signal modification unit, length and type of cables (if applicable).

2.4 OBSERVING PRACTICES

Meteorological operational procedures, including observational practices, schedules, or data conversion algorithms have varied in the past and are likely to continue to vary in the future. These changes can incorporate undesirable effects into a climate time-series and reduce its quality or break its homogeneity, invalidating it for use in studies of the detection and modelling of climate variability and change, including anthropogenic forcing. Keeping and archiving information on these changes is another vital metadata issue that needs to be properly documented. Each meteorological station should store, as best practice, information on the observers, the observed elements, the observing times, the instruments maintenance routines and the corrections applied to the measurements by observers. Most station observers record special or uncommon matters (checks, problems) conscientiously. Describing routine matters, however, is often forgotten because "everybody knows that such and such is done in this and that way". In ten years the procedure may have changed and nobody will remember any more what the "old" procedure was. At least all routine procedures (and any change in them) should be documented. A list of items to include in the metadata follows: A detailed list of these items is set out below.

2.4.1 **Observer**

For a given station, it is important to know if the observer is always the same person, or if he/she is different people doing some other specified job. In the latter case the reliability is usually also variable, and it should be explicitly known which person is responsible for the logbook and other metadata archiving. Regular documentation in a logbook is necessary for occasional maintenance matters such as calibrations, results of station inspections and instrument checks, instrument replacements, malfunctions (also those due to excessive rainfall, or frost, or lightning), painting or non-routine cleaning of instruments and instrument shelters, mowing enclosure vegetation, and so on. In particular, if an observation is omitted, or made at the wrong time or in an uncommon way, this should be entered in the logbook, as records with identified errors are greatly preferable.

2.4.2 **Observed elements**

Meteorological stations should keep a single list with the current meteorological elements observed/measured directly and those calculated indirectly from the observations at the station.

2.4.3 **Observing times**

Other crucial metadata information is obtained by documenting the observing times used at the station. Times and number of observations vary between stations and over the years at a particular station, and these changes can be the cause of a break in the homogeneity of a climate time-series. So it is vital that stations clearly identify the time of daily observations currently recorded and also if observations are omitted on holidays or weekends. In addition, particular care should be taken when a daylight-saving time scheme is used in a given country because that may imply the modification of observing times. Observers should indicate the operational period of daylight-saving time and the time when the observations have been taken.

2.4.4 **Routine Maintenance**

Another relevant issue in metadata is keeping information on routine maintenance at the station. Aspects like controlling the accuracy of instruments, replacing disposable items and housekeeping maintenance should be documented as recommended below:

- *Checking instruments*: routine checks of instruments to ensure their maintenance and correct functioning are usually performed by regular inspections at every station. Dates of inspections and results of the instruments checks should be kept. It is particularly important to document instrument sensitivity shift, if detected, and results of calibrations and recalibrations, both for conventional instruments and for automatic weather stations (AWS), upper air instruments and remote sensing platforms and sensors. Also occasional malfunctions due to e.g. frost or to storm damage should be logged.
- *Replacing disposable items*: a maintenance schedule includes replacing disposable items in autographic instruments, psychrometer wick or pan water, among others. Dates of chart change and other disposable material should be kept in this section of the station's metadata.
- *Housekeeping:* maintaining observing sites in good order requires operations such as cutting grass and cleaning of exposed instrument surfaces, like the domes of radiation meters. All these should be documented by logging the times when these works have been performed.

2.4.5 Corrections made by observer

Several elements are corrected *in situ* by observers or calculated from the observations, like the reduction to sea level of air pressure or when a humidity parameter is derived from the psychrometric readout. Any corrections of the original instrument readout which are made by the observer before logging the data, or which are calculated by the apparatus prior to data entry in the database, should be archived. For example, observations of relative humidity may be logged as dew point temperatures, and in this case backup copies of the used monograms, conversion tables or computer programs should be kept. So, which elements have been corrected and the procedures algorithms used for calculating derived quantities from the observations should be reported, including dates of changing such procedures. Comprehensive references on bibliographic sources employed in carrying out these procedures should be retained.

2.5 DATA PROCESSING

It is very important to keep information on how the data have been processed, validated and transmitted to the regional or central office from every single station. The knowledge of the data processing, the quality control procedures and the homogeneity tests applied to the data are vital to ensure both the accuracy of the observations made and the validity of the resulting time-series. A complete and accurate metadata record should retain, as a minimum requirement, information on the units used, the special codes employed, the corrections made to the data, the procedures of quality control applied, the adjustments made to ensure its homogeneity, and the data estimated and filled in, after applying the interpolation procedure selected. Logging and reporting of observations is a task shared by individual stations and regional or central offices, particularly when deriving and validating time aggregated data, such as monthly averages. Some of this information should be recorded at the stations and some by central offices. But in all cases, a complete metadata on the procedures applied to the observations and time-series should be retained, as it is indicated below:

2.5.1 **Units**

To keep and provide information on the units employed when observing, archiving and transmitting each one of the measured elements is an essential metadata requirement that should be accomplished by every station. The units employed for quantifying a measured element have varied in the past and still differ today on a country-by-country basis (e.g., air temperature has been and is expressed in Celsius, Kelvin, Fahrenheit or Reamur degrees, according to the country or the historical period). Confusions with units can cause data misinterpretations. A good and clear example on how not keeping the metadata of units can alter and hinder both the temporal evolution of a time series and the application of any quality control to the data is provided in ***Figure 3***, and described in its caption.

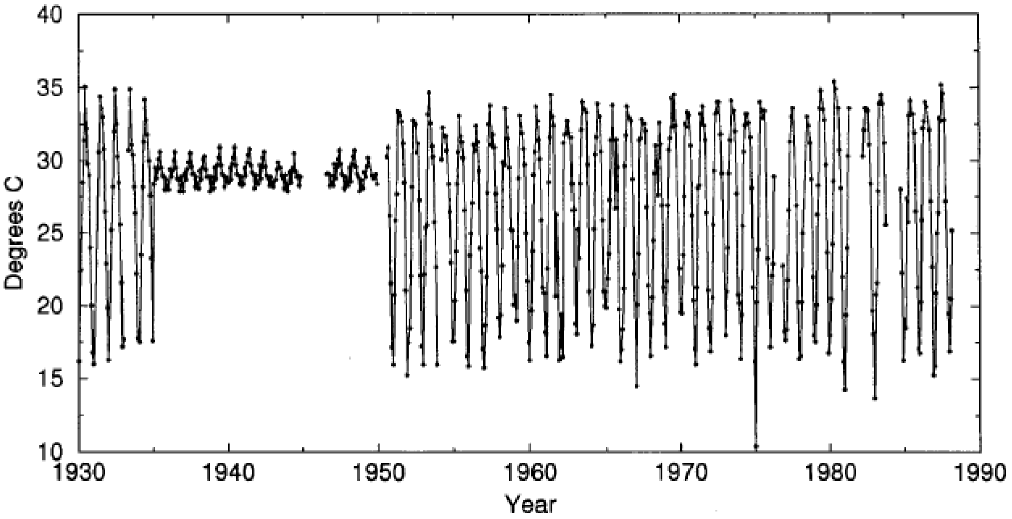


Figure 3. Mean temperature data for Bilma, Niger. For a period during the 1930s and 1940s, many stations in French West Africa reported in Kelvin. Since the mean temperature was approximate 300 K in a region with 30°C temperatures, someone 'corrected' the Kelvin temperatures by dividing by 10. Many QC checks would identify the January 'corrected' Kelvins as erroneous because they average over 10°C warmer than the mean of the other January data (28° vs. 17°), but accept most of the May data as valid because they average only 2°C warmer than the mean of the other May data (32° vs. 30°). The approach outlined in Peterson, T.C. et al, (1998) identify such problems by examining the time series as a whole using the SCUSUM test, developed to identify changes in variance. Source: Peterson, T. C. et al., 1998: Global historical climatology network (GHCN) quality control of monthly temperature data, International Journal of Climatology, 18, 1169-1179.

2.5.2 **Special codes**

Meteorological stations usually include special codes in the data, to report special situations, like missing data, wrong value, trace precipitation, non-precipitation, or accumulated precipitation, among others. In the process of data transmission, these kinds of codes should be properly identified and retained in the metadata base. As an example, it is common in many countries to code trace precipitation with "-3" or some other negative value. In absence of metadata the user may reject a value below zero for this element, and by doing so, producing a bad estimate of rainy days.

2.5.3 **Calculations**

Calculations other than those made on-site by the observers, such as time averaging (daily, monthly and so on) of elements, can also be performed at stations or at regional and central meteorological offices. Stations and NMHS should guarantee that the algorithms or formulae used to correct and convert elements are adequately documented. For instance, it should be a minimum requirement to provide information on which formulae have been employed to reduce air pressure to sea level and to detail how the elements derived from observations have been estimated. Finally, information on formulae employed to calculate daily averages and to insert monthly totals and averages of meteorological elements, should also be retained.

2.5.4 **Quality Control**

Quality Control (QC) procedures are applied to detect and identify the errors made in the process of recording, manipulating, formatting, transmitting and archiving data. Therefore, knowledge of the applied procedures will allow assessment of the validity of the observations and improve the data usage. GCOS Climate Monitoring Principles recommend to regularly assess as part of routine operations the quality control and homogeneity operations (WMO, 2002).

As a minimum requirement, a yes/no answer is recommended to indicate whether any QC has been applied or not. If the answer is positive, it would be a good practice to document the degree of QC applied to the data (e.g., subjected to logical filters only, compared for internal coherency in a sequence of observations, for spatial consistency among suitable neighbouring stations, for coherency with its climatological values and limits, etc.) and provide details on the employed techniques and their application:

- *Gross error checking*: report what kind of logical filters have been utilised to detect and flag obviously erroneous values (e. g., anomalous values, shift in commas, negative precipitation, etc.).
- *Tolerance tests:* document which tolerance tests to flag those values considered outliers to their own climate-defined upper/lower limits have been applied. Provide the percentage of values flagged and information on the approximate climate limits established for each inspected element.
- *Internal consistency check:* indicate whether data have undergone inspection for coherency between associated elements within each record (e.g., maximum temperature < minimum temperature; or in psychrometric measurements, the dry-bulb temperature ≤ wet-bulb temperature).

- *Temporal coherency:* inform if any test has been performed to detect whether the observed values are consistent with the amount of change that might be expected in an element in any time interval and to assess the sign shift from one observation to the next.
- *Spatial coherency*: notify if any test is used to determine if every observation is consistent with those taken at the same time in neighbouring stations affected by similar climatic influences.

Any available details about the exact techniques applied will be a great help for the future data user if provided, as well as information on the data that fail the tests and the period which the tests have been run for.

2.5.5 **Homogeneity Adjustments**

A later section of this guidance provides an extensive insight on homogenization. Homogeneity testing is performed to ensure that time fluctuations in the data are only due to the vagaries of weather and climate. Temporal homogeneity of a climate record is essential in climatological research, particularly when data are used to validate climate models, satellite estimates or to assess climate change and its associated environmental and socio-economics impacts. Therefore, it would be essential to report whether any kind of homogeneity testing has been applied to the data. Again, a minimum required practice would be to report if any homogenization technique has been applied or not. If the station or the regional and central Climatological Sections of an NMHS have implemented some of the existing approaches for homogenizing data, then it would be a best practice to report as much of the following information as possible:

- Which elements have been tested for homogeneity
- During which periods
- On which time scale (daily, monthly, seasonally or yearly)
- Name and reference of the applied test or short test description if no reference is known.
- Number of homogeneous/inhomogeneous records found within a network after applying the test (how many time-series are homogeneous, how many have one inhomogeneity, two inhomogeneities, and so on).
- Number of inhomogeneities found in each single time-series (free of inhomogeneities, one, two, three inhomogeneities and so on).
- Length of the inhomogeneous sections found in each time-series and/or time of breakpoints.
- Annual variation of the number of inhomogeneities in each record (number of cases per month).
- Size of the inhomogeneities detected and the correction factors used to adjust them.
- Causes of the detected inhomogeneities in every time-series (abrupt shifts: relocations, change of instruments/sheltering, change of time of observations, change

of observers, change of time of observations; gradual spurious warming/cooling trends: like those related to urban effects, and land use/land cover change impacts).

2.5.6 **Data recovery**

When observations are subjected to some kind of validation, quality control or homogeneity test, a variable amount of values are flagged as missing, suspicious or inhomogeneous. Some of them are detected and corrected on-site before transmission and others are flagged at the regional and central Climatological Sections of a NMHS. The amended data should be correctly documented, as a minimum requirement. It is vital to know if for any of these reasons some data have been modified from their original values or missing data have been filled in. It is recommended, as a minimum metadata requirement to report if any quality control procedure has been applied to data or not. If so, it would be a good practice to report some additional information, such as:

- Percentage of data filled with estimates in a time-series
- Fraction of missing data allowed in calculating monthly averages of the element from daily values
- Algorithms used for calculating in time interpolations schemes
- Algorithms employed and neighbouring stations used (number of stations, names and location details) for calculations in spatial interpolation schemes.
- Period of data for which the interpolation scheme has been performed.

2.6 GENERAL HISTORICAL NETWORK INFORMATION.

General information concerning the whole network of a country or regions within a country often seems to be trivial, and sometimes is not documented because it is thought to be well known. However, such information is at least as important as the individual station information. For historical metadata this topic becomes more and more important.

Since the end of the 19th century the NMHSs have become aware of the necessity to harmonize their observations and measurements as well as the regulations for the publication of data. Recommendations and resolutions passed at International Meteorological Conferences of Directors were introduced into the networks. Some changes happened suddenly, other processes took decades. Practices and regulations for observations and measurements have been changing since the beginning of measurements.

Due to the history of some countries the responsibilities of network management for larger regions may have changed several times. This could cause prominent inhomogeneities within the time series, which are not very easy to detect by statistical homogeneity testing as larger regions are affected at the same time (see Chapter 3 on homogenization). As an example, consider the station Pula, which is now managed by the Croatian Hydrometeorological Service. Its turbulent history started with the K.K. Central-Anstalt für Meteorologie und Erdmagnetismus. From 1918 until 1930 it was managed by the Ufficio Centrale in Rome, from 1931 until 1941 it belonged to the Federal Republic of Yugoslavia. During World War II it was occupied by the Germans and after 1945 it belonged to the Socialistic Republic of Yugoslavia. Since 1991 it has been part of the network of Croatia (The first part of the precipitation series is shown later in Figure 4).

Metadata are dramatically necessary to describe and interpret long-term data collections. For questions in regard to the long-lasting time-series required, for example, to place climate anomalies and extreme events into historical perspective, information about the topics addressed in the following sections is of great importance.

2.6.1 **Changes in sheltering and exposure**

At the beginning of measurements, thermometers were generally not adequately sheltered to minimize the radiation error. Old station histories reveal stations without any sheltering and/or thermometers sheltered in metal or other less than suitable screens. For each network, it is of great importance to know when the current sheltering has been introduced into the network and what was used previously. Normally, the introduction of a new type of screen into the network is not a sudden change but part of a gradual installation program lasting for years or decades. Hence, it should be highly recommended to retrieve and archive information on types of screens utilized in the past times (names and well-documented descriptions on their structures and shapes) and during which periods they were used.

A similar process can be described for changes in the exposure of instruments, which can affect long-term series significantly. For example, for precipitation gauges and thermometers there has been an evolution from higher to lower elevated installations of instruments in some regions, often a source for significant inhomogeneities in long-term series. Parallel measurements, as those shown in ***Figure 4***, can demonstrate this. So, it is also highly desirable that NMHS recover information on time changing exposure of instruments (elevation and dates of each change).

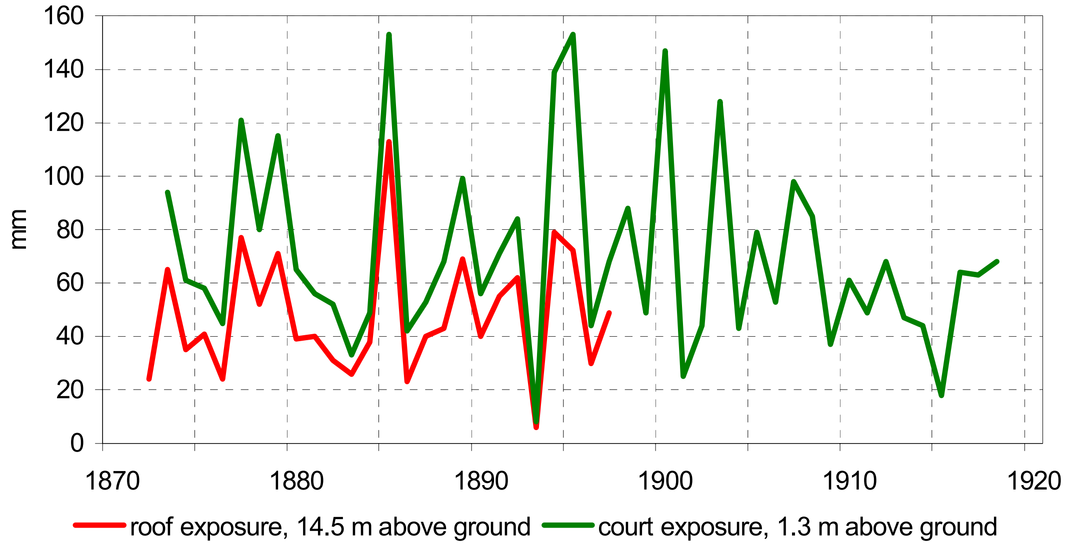


Figure 4: Precipitation series for April at Pula Monte Zaro (Croatia). Measurements were taken on the roof of the building of the K.K. Hydrographical Office from July 1871 and also in the courtyard from 1873. The period of parallel measurements lasts from 1873 until 1897. The mean precipitation amount of April at the roof exposure shows a deficiency of 35% compared to courtyard exposure. Data sources: Jahrbücher der K. K. Central-Anstalt für Meteorologie und Erdmagnetismus 1871-1915, Wien, Beiträge zur Hydrographie Österreichs, X. Heft, Lieferung II, Wien and Archivio del Ufficio Centrale di Meteorologia e Geofisica Italiano, Roma.

2.6.2 **Changes in mean calculations, observation hours and daylight saving times**

At the International Meteorological Conferences of Directors in Vienna in 1873 the first suggestions were made to unify observation hours in order to get daily and monthly means as consistent as possible. Before that time the recording and reporting of climate observations was not well coordinated. Observation hours have been objects of changes. ***Figure 5*** shows the impact of this on the calculation of monthly mean values. Due to this undesirable effect that can reduce the quality and homogeneity of long-term climate records, it is also recommended that NMHS recover so complete as possible information on temporal changes in the number and times of sub-daily observations.

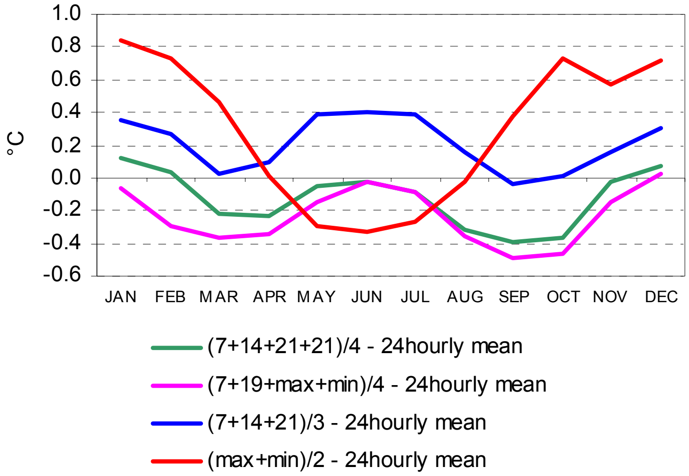


Figure 5: Evolution through the year of the difference between different ways of calculating daily mean temperature and 24-hourly observations average for the inner-alpine station Puchberg in Austria, 1987-1996. Data source: Central Institute for Meteorology and Geodynamics, Vienna, Austria.

As indicated in section 2.4.3, many countries introduce a daylight saving times scheme, which also can impact climate time-series. For this reason, NMHS should clearly document for every year when daylight saving times regulation came into legal force and how this problem has been treated in regard to observation hours.

2.6.3 **Units of observed elements and data accuracy**

In section 2.5.1 we discussed how changing units can affect data quality. Historical data might been archived in units which are no longer in common use (e.g. cloud estimations in quarters, air pressure and vapour pressure in mm or English inches, temperature in Reamur degrees, geographical altitudes in Toises, precipitation in Parisian Lines, foot lengths, etc.). It should be clearly stated when a change to new units took place.

Another matter of concern when analyzing historical and network data is accuracy of measurements. As measuring instruments have evolved and observing practices have changed over time, the accuracy of measurements has been improved. This has to be accounted for, as different degrees of precision could be incorporated in the same time-series.

2.6.4 **Urbanization and land-use changes**

Today the global population is significantly greater than it was in the past, and will continue to increase for many decades to come. It is expected that by the year 2025 almost two thirds of world's population will be living in urban areas. Urbanization often leads to less green space within a town or city and increased use of concrete and steel, more vehicles and industries, higher concentrations of pollutants etc. This results in a built-up of heat commonly known as the "urban heat island". Growing population numbers and changes in land-use can show an impact on our meteorological series.

Urbanization effect in a meteorological time series (see ***Figure 6***) does not cause a sudden break in its homogeneity if the local environment remains unchanged, but instead a gradual trend. However, this trend cannot be regarded as a trend for a city as a whole; it is strongly influenced by the local surroundings. The change in urbanization over time may be smaller for a station which originally was established in a densely built-up area or in an urban park than for a station originally installed in a rural or only light urbanized environment that has experienced growth. It is of greatest importance to collect all available information about building density on local and regional scales, as well as on historical population statistics for the country. Moreover, network managers should document other major land-use changes, like swamp reclamation.

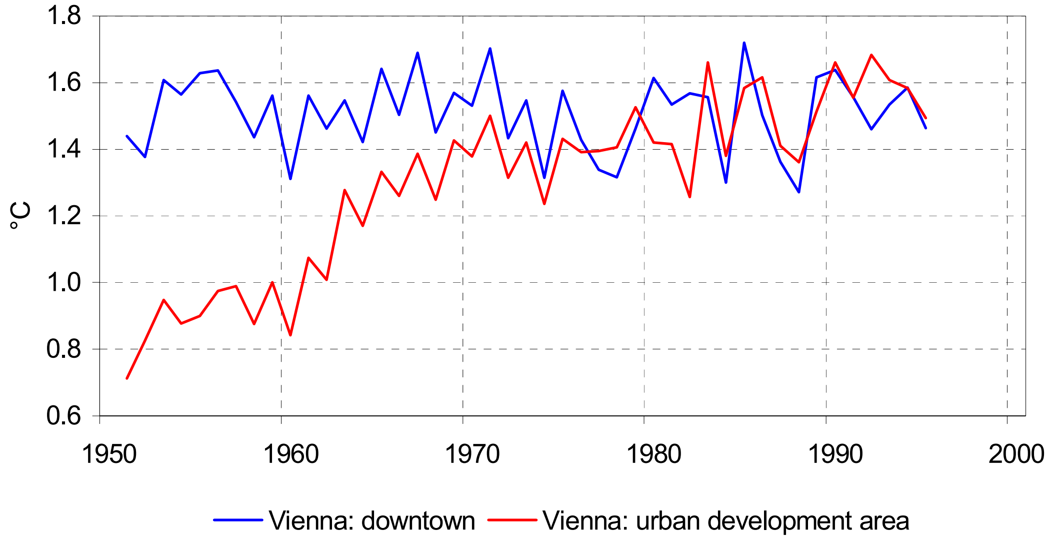


Figure 6: Time series of annual mean urban temperature excess (relative to rural mean 1951 to 1995) based on height reduced temperature records. The station in the densely built-up area shows a stable temperature excess against the rural surroundings, whereas the trend of temperature excess at the station in the urban development area is 0.18 °C per decade. Data source: Böhm, R.: Urban bias in temperature series – a case study for the city of Vienna. Climatic Change 38: 113-128, 1998.

2.6.5 **Introduction of Automatic Weather Stations or new types of instruments**

The number of Automatic Weather Stations has been increasing worldwide and will do so for the foreseeable future. This is a network development which necessitates careful documentation, even if the station is replaced at an identical location. Metadata should register not only the moment of introduction of the automatic system, but also the software employed along with its updates or substitutions. If the conventional station and the AWS co-exist at the same location, metadata should specify if data from the former is employed to fill data failures in the automatic system.

Parallel measurements (see ***Figure*** *7*) are very important to be able to maintain them for so long as possible, as they can document the effect of the introduction of the new system in data. Extra details will be further explained in chapter 3.

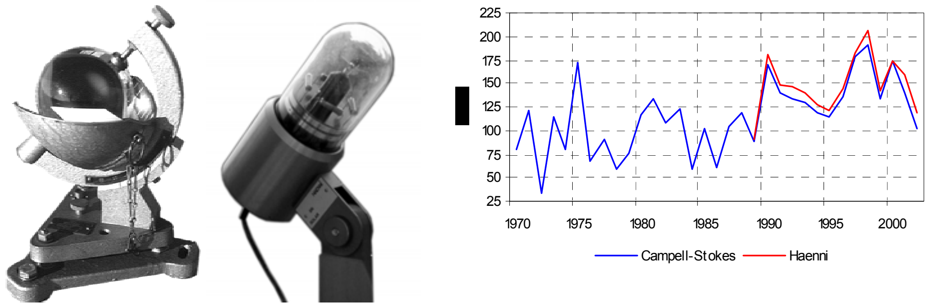


***Figure** 7*: **Left**: two types of instruments to record sunshine duration: Campbell-Stokes sunshine autograph and Haenni Solar system of automatic weather stations. **Right**: Consequences: time series of hours of bright sunshine for February in Graz-University (366m asl.) since 1970: Campell-Stokes sunshine autograph (blue curve). With the introduction of the automatic weather station a Haenni solar system (red curve) was installed next to the Campbell-Stokes, which in February systematically records an excess of sunshine. Sunshine has increased since 1970: for the unchanged Campbell Stokes 2.13 hours per year, continuing the series with Haenni solar since 1989 the trend would be biased with an excess of 0.45 hours per year. Data source: Central Institute for Meteorology and Geodyamics, Vienna, Austria.

2.6.6 **Changes in quality control, homogenization and data recovery procedures.**

As discussed in section 2.5, a complete metadata should include information on the QC, homogenization and recovery efforts a dataset has undergone. From a network point of view, it is recommended to report the changes experimented by these procedures.

Altering QC procedures at one point can have an important impact on data. For example, prior to 1999, radiosonde temperature data processed in the U.S. were truncated at -90°C. After that time, temperatures as cold as -93°C started appearing. While this change was appropriate as it allowed valid cold temperatures to exist in the record, it did create a cold bias in analysis looking at changes in temperature in the tropical tropopause.

In the present times, the growing use of powerful software based on Geographical Information Systems and/or statistical packages allows a better definition of the valid range of data. Metadata should reflect when these tools were firstly implemented.

Starting a data recovery effort will hopefully decrease the amount of missing data, having also an impact on data use. Finally, changes in homogenization procedures need also to be registered in the metadata file.

2.7 METADATA STORAGE AND ACCESS

As discussed in previous sections, metadata are of crucial importance to interpreting measurements and observations. Although many metadata are hand-written in native languages the NMHSs are encouraged to develop a flexible, extensible system to manage climatological station metadata. An optimal solution will be a database, which allows the digitization of metadata, and for flexible usage and accessibility of the digitized metadata for various applications. Although some elements like pictures and maps are difficult to key and code, the data user needs to be able to see metadata in the form of a time series of changing events.

Before entering metadata into a storage system, it is important to QC metadata itself, maybe by checking the accuracy of a sample of entries. After storage, some QC should be made, to ensure anomalous values have not entered into the database.

The more information stored within a metadata-base, the more applications will be possible, first of all for the data provider but also for the data user. The complete history of a station starts with its first installation (including all details about identity, location, instrumentation, observing practices, data programs and station management including a map display and supporting document images) and should continuously be kept until its final closing. During this time many internal (e.g. instrumental change) and external changes (e.g. land-use change) are influencing factors to the observational data and each of the changes has its own finite period of validity. Note that for any given station details may change independently of the others. All this has to be documented and stored to compile the Station History which should include all the facts that affect the data, including network wide changes.

Although up to now there is no final conclusion about the best system to use, there have been good experiences with a system based on a normalized relational database. The optimal solution for the NMHSs will be a database capable of being linked with observational data and transmitted along with them. A set of various tables, each of them representing a different metadata field, can be combined and linked to allow the selection of specific information for further applications. Separate tables can also be maintained for network-wide events, so they can be easily linked to single station metadata when necessary. The list of possibilities for the application of metadata knowledge for internal administrative and scientific tasks of a NMHS may be long. Here, only a few examples can be mentioned.

For the purpose of quality control of “suspect” observational data: a link to metadata allows knowledge about the local situation of the station to be included in the considerations of data acceptance or rejection. Stored information about instruments will allow to select all stations where a special instrument is in use or to determine stations with a specific exposure of instruments, e.g. rain gauges that for particular reasons could not be installed within the norm-height. Should storage in a relational database not be possible, it is always advised to store metadata in some digital form, like spreadsheets or flat text files.

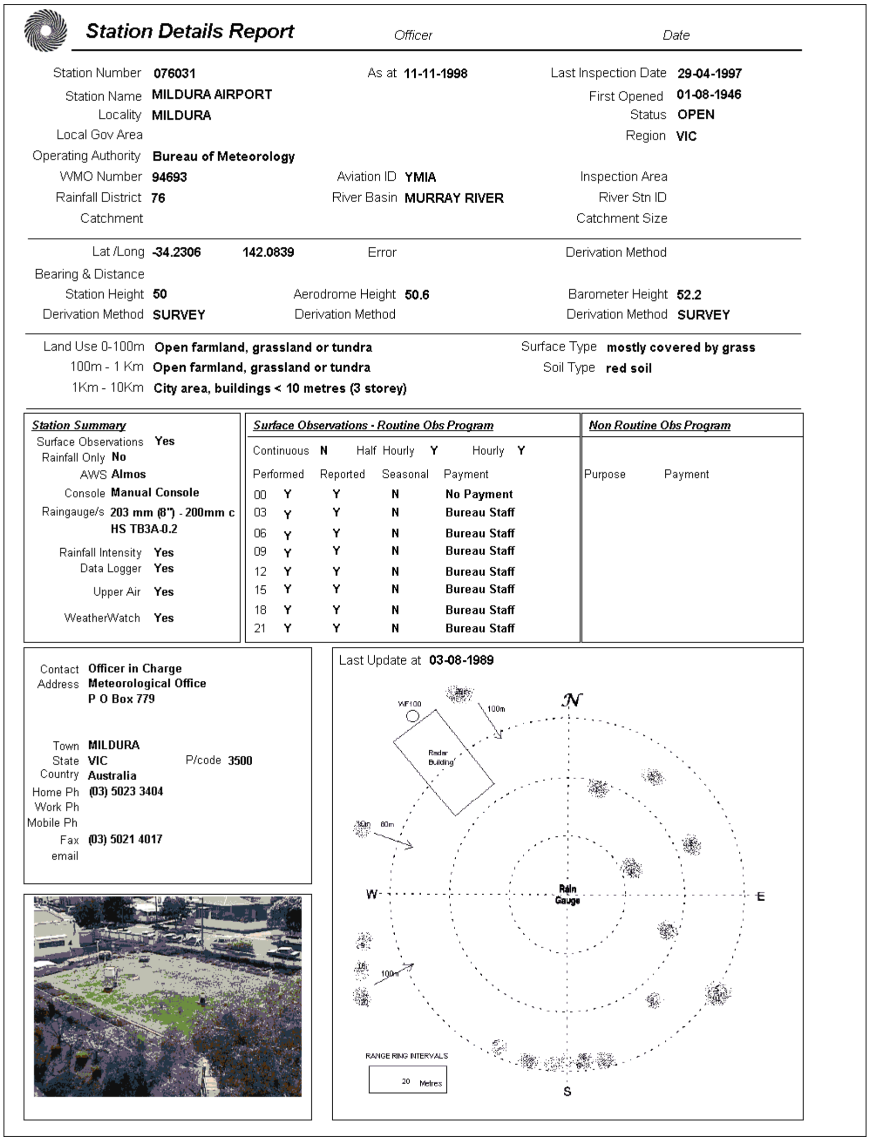
Relational databases require adequate software and some degree of expertise. The lack of those should never be an obstacle to maintain a metadata database, as very simple solutions like spreadsheets, tables in text processor or even flat text files can be useful to create and maintain a digital metadata set. Additionally, modern database software can easily import data from other formats, so any step taken to digitize metadata will be worthwhile in the future. *Table 2* shows an example of location metadata for Spanish stations. This table is useful for understanding changes in location in long-term stations (defined by the Internal Code and Alias fields), often composed by data from different locations (National Code, Name Alias, Lon, Lat, Ele), which participate in the record for a defined period of time (Start, End) and belong to a determinate pre-defined climatic type (Climate). This table can be constructed as part of a relational database or, simply introduced as lines in a text file. Additional files/tables can store other information (instruments, exposure, etc.) using a key field to link both, for example the National Code.

Table 2: Location Metadata for Spanish Stations. Source: Climate Change Research Group, Tarragona, Spain.

Internal Code	Alias	National Code	Name	Start	End	Lon	Lat	Ele	Climate
CO001	MADRID	3195Z	MADRID-1	1860	1893	034041	402440	667	5
CO001	MADRID	3195	MADRID RETIRO	1894	2001	034041	402440	667	5

Internal Code	Alias	National Code	Name	Start	End	Lon	Lat	Ele	Climate
CO002	GERONA	0370	GERONA	1916	1973	-024930	415836	94	4
CO002	GERONA	0367	AEROPUERTO GERONA-COSTA BRAVA	1974	2001	-024537	415405	129	1
CO003	PONTEVEDRA	1484w	PONTEVEDRA 'INSTITUTO'-BIS	1881	1929	083859	422550	19	6
CO003	PONTEVEDRA	1484	PONTEVEDRA 'INSTITUTO'	1930	1949	083859	422550	19	6
CO003	PONTEVEDRA	1485	SALCEDO	1950	2001	083827	422438	40	6

Figure 8 shows an example of a metadata sheet for a single station. This format and metadata lists such that on Table 2 are complementary and can be easily derived one from another with state-of-the-art database programs.



Station Details Report *Officer* *Date*

Station Number **076031** As at **11-11-1998** Last Inspection Date **29-04-1997**
Station Name **MILDURA AIRPORT** First Opened **01-08-1946**
Locality **MILDURA** Status **OPEN**
Local Gov Area Region **VIC**
Operating Authority **Bureau of Meteorology**
WMO Number **94693** Aviation ID **YMIA** Inspection Area
Rainfall District **76** River Basin **MURRAY RIVER** River Stn ID
Catchment Catchment Size

Lat /Long **-34.2306** **142.0839** Error Derivation Method
Bearing & Distance
Station Height **50** Aerodrome Height **50.6** Barometer Height **52.2**
Derivation Method **SURVEY** Derivation Method Derivation Method **SURVEY**

Land Use 0-100m **Open farmland, grassland or tundra** Surface Type **mostly covered by grass**
100m - 1 Km **Open farmland, grassland or tundra** Soil Type **red soil**
1Km - 10Km **City area, buildings < 10 metres (3 storey)**

Station Summary
Surface Observations **Yes**
Rainfall Only **No**
AWS **Almos**
Console **Manual Console**
Raingauge/s **203 mm (8") - 200mm c**
HS TB3A-0.2
Rainfall Intensity **Yes**
Data Logger **Yes**
Upper Air **Yes**
WeatherWatch **Yes**

Surface Observations - Routine Obs Program

Continuous **N** Half Hourly **Y** Hourly **Y**

	Performed	Reported	Seasonal	Payment
00	Y	Y	N	No Payment
03	Y	Y	N	Bureau Staff
06	Y	Y	N	Bureau Staff
09	Y	Y	N	Bureau Staff
12	Y	Y	N	Bureau Staff
15	Y	Y	N	Bureau Staff
18	Y	Y	N	Bureau Staff
21	Y	Y	N	Bureau Staff

Non Routine Obs Program

Purpose Payment

Contact **Officer in Charge**
Address **Meteorological Office**
P O Box 779
Town **MILDURA**
State **VIC** P/code **3500**
Country **Australia**
Home Ph **(03) 5023 3404**
Work Ph
Mobile Ph
Fax **(03) 5021 4017**
email

Last Update at **03-08-1989**

Figure 8: Metadata information for Moldura Station, Australia. Source: Australian Bureau of Meteorology.

In previous sections we described what metadata should be collected. *Table 3* identifies both the minimum requirements for metadata collection and the best practices. Although it is sometimes difficult and time consuming to store or recover metadata, WMO encourages NMHSs and observers to keep them as complete and up to date as possible. By ensuring that the minimum requirements are at least met, and by trying to move further towards achieving the best practices, society will benefit from a better understanding of weather and climate. Metadata helps in the selection of stations and data products and their use for further scientific studies. For these and other reasons, metadata often need to be accessed by external users. Even though it is advisable to make most of the metadata available to external users, there might be some restrictions due to national and international regulations, NMHS data policies and privacy issues. Metadata made available to the general public should be at least those items identified in ***Table 3*** as minimum requirements.

Although some efforts have been undertaken to create worldwide metadata sets (e.g. those undertaken by the GCOS or in the Comprehensive Aerological Reference Dataset (CARDS) project), metadata are usually stored locally at the NMHSs headquarters and, in some occasions, at the same stations. Metadata gathering is a time-consuming matter for researchers, and for this reason, a central depository, perhaps located at World Data Centers, would be of interest to the scientific community.

Table 3: Metadata elements to be stored for a meteorological station. ***Bold italic*** items are minimum requirements; other items are best practices.

CATEGORY	METADATA TYPE	BRIEF EXPLANATION
STATION IDENTIFIERS	***Local Code*** ***WMO Code*** ***Name and aliases*** Active/Closed Beginning/End Date Type of Station Responsible Organization Manual/AWS Time zone Networks	Clearly identify the station and whom responsibility it is. It is very important to do so by reporting all the different codes, as some times WMO codes are not used locally and national codes are not known abroad. It is also useful to know which networks a station is included in.
GEOGRAPHICAL DATA	***Latitude*** ***Longitude*** ***Elevation*** ***Dates of relocation*** Topographical Information Method of deriving lat/long Resolution of lat/long	Geographical coordinates and exact dates of relocations along with other topographical details. Care must be taken in differencing N/S latitudes and E/W latitudes as well as with reporting fractions of degree (minutes and seconds or thousandths of degree)
LOCAL ENVIRONMENT	Local land use/land cover Instruments exposure Soil type Site condition Photographs Site plans Skyline diagrams	Document the station environment and instruments exposure: obstacles, e.g. land use, population growth, obstacles, exposure site land cover, etc.

STATION INSTRUMENTATION AND MAINTENANCE	Type of instruments Instrument comparisons Start/end dates of instruments Condition of instruments Instrument Sheltering and Mounting Type of recording Calibration results Special Maintenance/Faults Modifications Barometer height	Report the characteristics of the instruments in use and their sheltering, accuracy, calibration and maintenance; indicate how data are transmitted. Carefully note any changes in instrumentation.
OBSERVING PRACTICES	Observer information Observer level of training List of observed elements Observing times Units used Observation instructions Routine maintenance operations Disposable items replacement Corrections made by observer	Keep documented what elements are observed and when, with special care to the enforcement of daylight saving times; report the exact moment of maintenance operations and any corrections made to data
DATA PROCESSING	***Units*** ***Special codes*** ***Calculations*** ***Algorithms*** ***QC applied? (yes/no)*** Other details on QC ***Homogenization applied? (yes/no)*** Other details on homogenization ***Data recovery effort? (yes/no)*** Other details on data recovery Treatment of redundant data	Report units in use and give conversion factors if they don't belong to the metric system. Indicate special codes used and their meaning; mention in the metadata any amendment made to the recorded data: calculations, corrections, qc, homogenization and data interpolation. Report criteria for missing data, and if more than one instrument for the same element, which is considered the primary instrument.
HISTORICAL EVENTS	Changes in the social, political and institutional environment Daylight savings dates	Add to metadata any significant changes in the station context that may affect data collection
COMMUNICATION	Signal transport/data transmission General correspondence	General correspondence such as e-mail between station operators and observers can include potentially valuable information about the quality of observations.

2.8 Metadata Sources

Sections 2.1 to 2.6 focused on which metadata should be registered and section 2.7 touched upon the issue of storage and availability. This last section of the chapter will discuss the most common sources to obtain and gather metadata.

- *Meteorological office (or other responsible institutions)*: if good metadata practices as those recommended in this guide are observed, the primary source for metadata should be the institution responsible for data recording and archiving.
- *Meteorological site*: as indicated in previous sections, mapping and photographing the meteorological site provides a great amount of information about the recording conditions
- *Data*: the data themselves are an excellent source of metadata, as metadata elements like station codes, location, observing times, missing data, etc. can be often derived from data
- *Scientific institutions*: many scientific institutions other than the meteorological offices maintain metadata sets which can complete other information.
- *Observers and experts:* interviewing current and former observers and/or local experts can add useful information
- *Archives, libraries, newspaper archives, etc*: consulting those sources may help to recover historical information for a single station or an entire network which has been lost. They can also provide data for population and urban growth or historical events.
- *Instrument manufacturers*: they can provide the most accurate information on technical issues related to instruments performance

All those sources can contribute to set up or complete a metadata set to meet the requirements suggested at this guidance.

3 HOMOGENEITY

3.1 BACKGROUND ON HOMOGENEITY

Climate data can provide a great deal of information about the atmospheric environment that impacts almost all aspects of human endeavour. For example, these data have been used to determine where to build homes by calculating the return periods of large floods, whether the length of the frost-free growing season in a region is increasing or decreasing, and the potential variability in demand for heating fuels. However, for these and other long-term climate analyses –particularly climate change analyses– to be accurate, the climate data used must be as homogeneous as possible. A homogeneous climate time series is defined as one where variations are caused only by variations in climate.

Unfortunately, most long-term climatological time series have been affected by a number of non-climatic factors that make these data unrepresentative of the actual climate variation occurring over time. These factors include changes in: instruments, observing practices, station locations, formulae used to calculate means, and station environment. Some changes cause sharp discontinuities while other changes, particularly change in the environment around the station, can cause gradual biases in the data. All of these inhomogeneities can bias a time series and lead to misinterpretations of the studied climate. It is important, therefore, to remove the inhomogeneities or at least determine the possible error they may cause.

For example, if a weather observing station is moved from a hill top location to the valley floor 300 meters lower in elevation, analysis of the temperature data will show in most cases an abrupt warming at that station. But this observed warming will not be primarily due to climate change. Detecting a problem such as this is easy if the change in elevation is large, but it is much more difficult for small changes in elevation. Also, consider a station located in the garden of a competent and conscientious observer for 50 years. The instruments are maintained in good repair and the observer accurately records the temperature in his or her garden. But what if 50 years ago the observer planted a tree west of the garden? This tree slowly grows up and shades the observing site during the late afternoon when the daily maximum temperature is observed. While the data accurate represent the temperature in the garden, the tree has caused maximum temperatures in the garden to have cooled relative to the climate of the region. Detecting gradual homogeneity problems such as this is very difficult.

Many researchers have put a great deal of effort into developing ways to identify non-climatic inhomogeneities and then adjust the data to compensate for the biases these inhomogeneities produce. Several techniques have been developed to address a variety of factors that impact climate data homogenization such as the type of element (temperature versus precipitation), spatial and temporal variability depending on the part of the world where the stations are located, length and completeness of the data, availability of metadata (see previous section on Metadata), and station density. Each team has developed a different philosophy regarding data adjustments since their requirements and missions have been quite different.

For example, the best technique for a dense regional rain gauge network in the humid extra-tropics might not work well for a sparse subtropical semi-arid network. Metadata in the form of station history documentation that details instrumentation, locations, observing practices, etc. may be digital or in paper archives or not available at all. What level of confidence is to be required before making an adjustment in time series? Numerically, one can use a 95% or 99% confidence level generated by a test statistic, but how would metadata weigh into such analysis? Some of these decisions are made based on the specific goal, for example, if one is trying to produce a homogeneous version of a far flung network a different approach might be required than if the goal was selecting and adjusting only the best stations from a dense network to produce a homogeneous subset. Some decisions are made based on years of experience with the specific data and metadata involved. But other decisions are, of necessity, based on the resources available: e.g., careful analysis of station history documentation can be very labour intensive.

Since the densest networks and sources of most meteorological metadata are usually country specific, many countries are preparing to address inhomogeneities in their data. The following sections will provide guidance on how to deal with inhomogeneity problems depending on the aforementioned circumstances.

3.2 KEEPING THE RECORD HOMOGENEOUS AND ADJUSTING CHANGES WITH DIRECT TECHNIQUES

The previous section described how inhomogeneities impact climate data. The best way to avoid them is to keep the record homogeneous. Changes in and around a meteorological station will lead to inhomogeneities in the data. So it is very important to prevent those changes from happening, at least in long-term stations, suitable for climate analysis. WMO has already identified a number of surface and upper-air stations to conform the Global Climate Observing System (GCOS) network. It is highly advisable to maintain in those locations the observing practices and instruments as unchanged as possible (GCOS, 2002). Ideally, the NMHSs should regard as priority climatological stations or Reference Climate Stations enough sites to fully represent the different climatic varieties of their countries' territory. WMO offers guidance on how to select climate reference stations (WMO, 1986) and the WCDMP has developed guidance for climate observation networks and systems.

In these climatologically important stations, it is advisable to limit changes to the unavoidable minimum. Sometimes, just by keeping in mind how important homogeneity and unchanged measurement conditions are for climate analysis, biases can be prevented. Going back to the growing trees example, a person responsibly concerned with data quality issues would decide not to plant them, anticipating the problems this will bring in the future. Nevertheless, some changes are unavoidable (and sometimes they are even desirable) and will happen even in the more carefully run stations. It is easy to think of many reasons why a meteorological station may need to be relocated or to understand how when an observer who has been in charge of a meteorological site for many years retires, the person who replaces him or her might introduce a bias on the data. It is evident that even the best instruments have an operational life, and sometimes will need to be recalibrated and eventually replaced. It is obvious that many records are not devoted only to climate research but also support aviation, navigation, agriculture, insurance, tourism, etc. as well and they can benefit from state-of-the-art equipment, like an AWS or the last radiosonde model. When these or other needed changes are about to take place, it is the responsibility of the institution in charge of the station to anticipate them and limit their impact on data homogeneity.

First of all, it is recommended to carefully register in the station history a metadata entry describing any change (see *Table 3* in Metadata section), along with the exact time when it happened. As well, it is very important to keep parallel measurements for a long enough period of time, covering several years. Simultaneously recording with the new and the old conditions, is the best option to derive correction factors and adjust data for homogeneity. In some events, like the introduction of an automatic weather station, it might be possible and very advisable to maintain the old conventional station indefinitely and devote its measurements to climate purposes only. It is important to understand that climate research acknowledges the use of new technologies, but obtains even greater benefits from unchanged conditions for long periods of time.

When parallel measurements are not available or dependable, the old conditions can be reproduced to some extent by undertaking some suitable extra efforts. Although it is very difficult to find and use original instruments or replicate the ancient building density, etc., good results can be achieved through two different ways. The first one is to reproduce experimentally the old conditions and make paired measurements (see ***Figure 9***). The second one is to model the old conditions and compare them with present day data (e.g. model recording properties of formerly used radiosondes). Both approaches are very difficult, time consuming and have an elevated cost, although they can offer excellent results and, as will be stressed in section 3.3.2, sometimes constitute almost the only available option.

Figure 9: Experimental reproduction of historical measurements conditions. Parallel measurements are taken with identical sensors

sheltered in a Stevenson Screen (left) and in a *Montsouris Screen* (right, behind meteorologist), widely used in Spain in the 19th century and the early 20th century. Photograph courtesy of Manuel Bañón, INM-Murcia (Spain) and project SCREEN.

3.3 INDIRECT HOMOGENEITY ASSESSMENT FOR MONTHLY, SEASONAL AND ANNUAL DATA.

Although the best way to ensure homogeneity is to keep the record homogeneous through appropriate management of the observations site and associated equipment, this is very difficult to achieve. Besides, because it is almost impossible to be 100% sure about the quality of past data, a homogeneity assessment is always recommended. There is not one single best technique to be recommended. However, the four steps listed below are commonly followed:

1) Metadata Analysis and Quality Control
2) Creation of a reference time series
3) Breakpoint detection
4) Data adjustment

3.3.1 **Metadata and quality control**

In the section on metadata it has been explained how important metadata are for identifying discontinuities in a time series. By putting together all the available metadata and building the station history, we anticipate and preview what problems we may find in the data and when they should appear. Some homogenization approaches only accept discontinuities registered in the metadata. This is indeed a good approach if we believe that our metadata are absolutely complete, from the first to the last observation. When trying to detect inhomogeneities we are looking for the fingerprints of factors other than climate and weather in data. That means there is always a cause for any inhomogeneity. Should metadata be perfect, we could always identify this cause and there would be no need to employ any statistics to find further breakpoints in a time series. Nevertheless, even in the presence of the most carefully documented metadata, it is advisable to compare what the station history says and what data analysis identifies, as a sort of double check.

Another way we may benefit from metadata is to know what kind of QC the data has undergone. QC procedures vary from very simple techniques, such as plotting the data against time (alone or together with neighbouring stations) or identifying data outlying pre-fixed thresholds, to sophisticated analysis that cross validates different meteorological elements at the same station and/or data from different stations. Even in those cases when we are aware that a complete QC has been applied, it is recommended to plot the data before actually starting the homogenization procedures and correct or remove from forthcoming steps obviously wrong data. This is crucial because many homogenization techniques rely on comparing the central value from two different data sections. Failing to remove outlying data enormously complicates statistical detection of any inhomogeneity or, in the best case, alters the value of the correction factor, especially if we are using a parametric test (see ***Figure 10***)

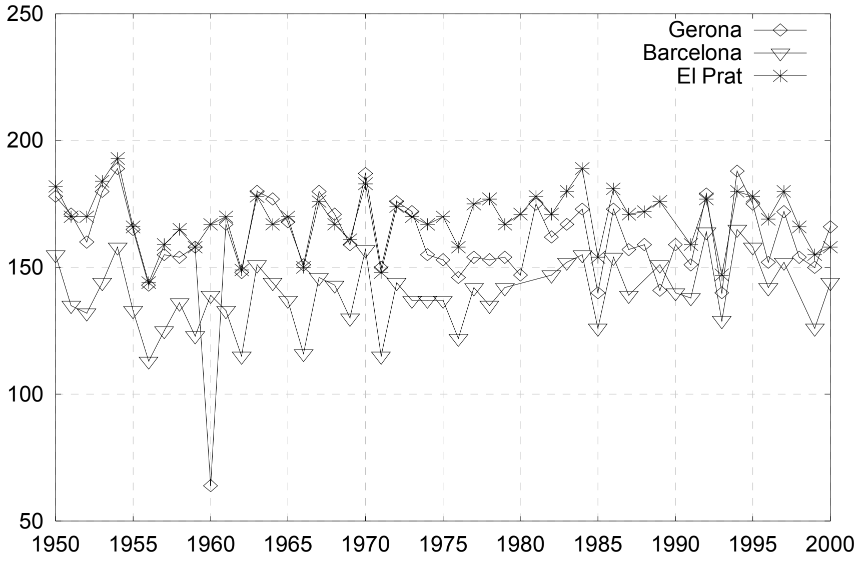


Figure 10: Monthly average of daily maximum temperature for November. Data in 1/10 ºC. Data shows an outlying value in 1960 that must be removed before performing any homogenization analysis. Figure modified from: Aguilar, E., Brunet, M., Saladié, Ò., Sigró, J. y López. D, 2002: Hacia una aplicación óptima del Standard Normal Homogeneity Test para la homogenización de Temperatura, in Cuadrat, J.M., Vicente, S.M. y Saz, M.A. (eds.): La información climática como herramienta de gestión ambiental, VII Reunión de Climatología, Zaragoza. ISBN: 84-95480-69-7

By plotting the data we can also identify other values, like special codes, that must not enter our analysis. For example, it is very common to use figures like "-999" to identify missing data. Failing to cut this code from the analysis will completely ruin it. Plotting the data will help us to decide if the data are free from these kinds of problems and we can go ahead with homogenization or we need to go back and further quality control them. At the same time, data plots can make us aware of obvious inhomogeneities in data.

3.3.2 **Building a reference time series**

Detecting and adjusting inhomogeneities is a hard and difficult task, as on most occasions the magnitude of the inhomogeneities is the same or even smaller than that of true climate-related variations. For this reason, it is advisable to create a reference time series and compare with the station to be homogenized, the so-called candidate station series. A reference time series ideally has ideally to have experienced all of the broad climatic influences of the candidate, but none of its artificial biases. Should the candidate have no inhomogeneities, when the candidate and reference series are compared by differencing (in the case of variables measured on an interval scale, like temperature) or by calculating ratios (for variables measured on a proportional scale, like precipitation), the resulting time series will show neither sudden changes, nor trends, but will oscillate around a constant value. However, if there are one or more inhomogeneities, the difference or ratio time series, will reveal their fingerprint (see ***Figure 11***).

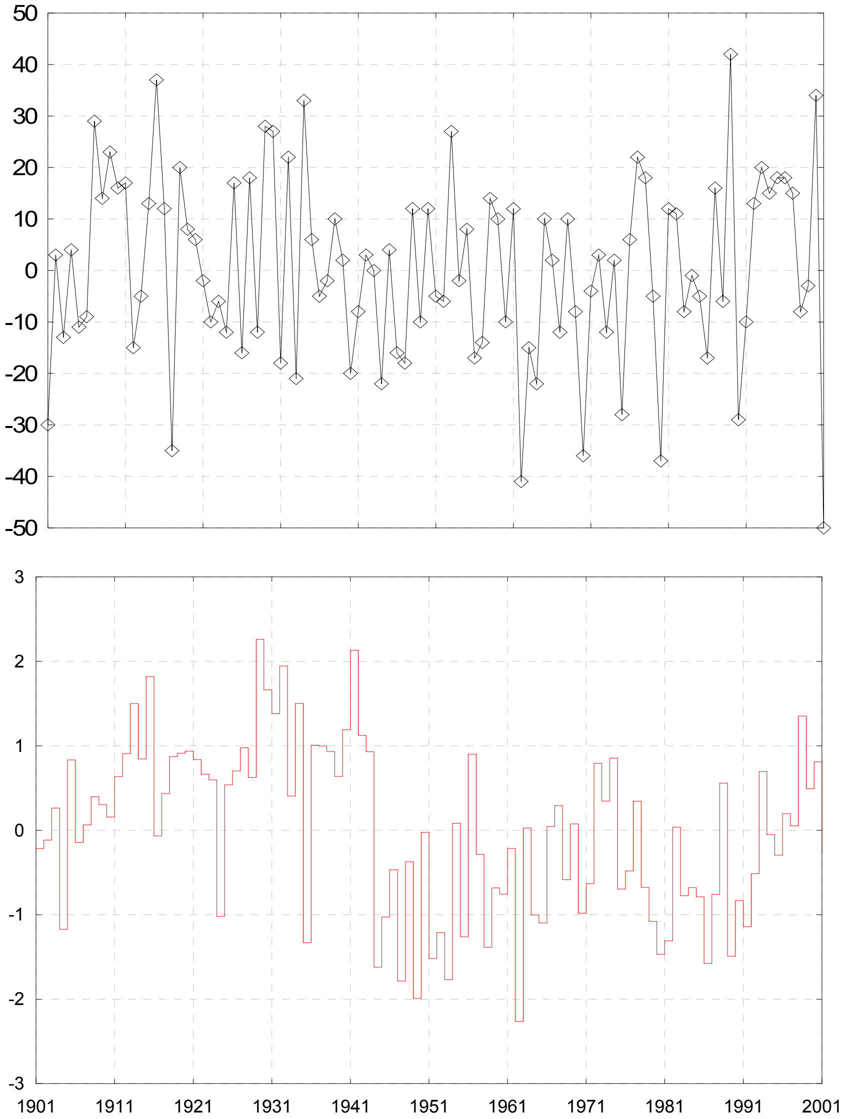


Figure 11: **Top:** Monthly Average of daily minimum temperature for December in Burgos, Spain. Data in 1/10 ºC; **Bottom:** difference between candidate and normalized reference time series calculated following the Standard Normal Homogeneity Test, using 10 neighbouring stations. The difference between candidate and reference time series (bottom) clearly shows an inhomogeneity in 1941, documented in the metadata as a relocation. The original data (top) mask the inhomogeneity. Figure modified from: Aguilar, E., Brunet, M., Saladié, Ò., Sigró, J.

y López. D, 2002: Hacia una aplicación óptima del Standard Normal Homogeneity Test para la homogenización de Temperatura, in Cuadrat, J.M., Vicente, S.M. y Saz, M.A. (eds.): La información climática como herramienta de gestión ambiental, VII Reunión de Climatología, Zaragoza. ISBN: 84-95480-69-7

The most common approach for building a reference time series is to calculate for each year a weighted average of data from neighbouring stations or sections of neighbouring station time series that metadata indicate are homogeneous. Some measure of similarity (usually correlation coefficient) is employed to select the most adequate neighbours and weight them according to their statistical resemblance to the candidate. Several widely used techniques calculate the correlation coefficients between first-differenced time series. A first difference series is made by subtracting year 1's observation from year 2, year 2 from year 3, etc. The correlation then is a measure of the similarity in year-to-year changes, and an inhomogeneity only impacts one observation rather than making all observations after the inhomogeneity artificially warmer or colder. A widely used alternative to integrate different series into an averaged reference is to compare them one by one with the candidate.

Other approaches extract principal components from the whole data network, or use an independent data source thought to be homogeneous. Creating and using reference time series may encounter two major problems, the first one being the lack of data to build them. Thus for satellite data, it will not be possible to use this source for dates prior to the 1970s. Similarly, for some regions of the world, the observational network was, or is, too sparse to find enough neighbouring stations to construct a reliable reference. Just think how difficult it would be to find well-correlated neighbours for remote locations like the Polar Regions or to build reference time series for stations extending back to the early 19th century. It is also true that the difficulties for building a reference time series increase exponentially with the increase in spatial variability of the data. Spatial variability depends on three key factors: the meteorological element we are dealing with; the type of climate; and the time resolution. It is intuitively very understandable that it is easier to create a good reference time series for annual averaged temperatures for a station at the equator than building a reference for August precipitation for a station in the Mediterranean.

Another problem arises when a cause of inhomogeneity impacts a whole data network or the best part of it. For example, until the late 19th or early 20th century, most meteorological stations were not employing the Stevenson Screen. In some countries, the new shelter was introduced about the same time in all the stations, so bias-free data are not available for referencing (see ***Figure 12***).

When encountering these drawbacks, it is very difficult to identify and adjust inhomogeneities through a reference series. If direct homogenization is not possible, the only approach left is the use of techniques based on the statistical analysis of station data, making the distinction between artificial biases and true climatic fluctuations more difficult.

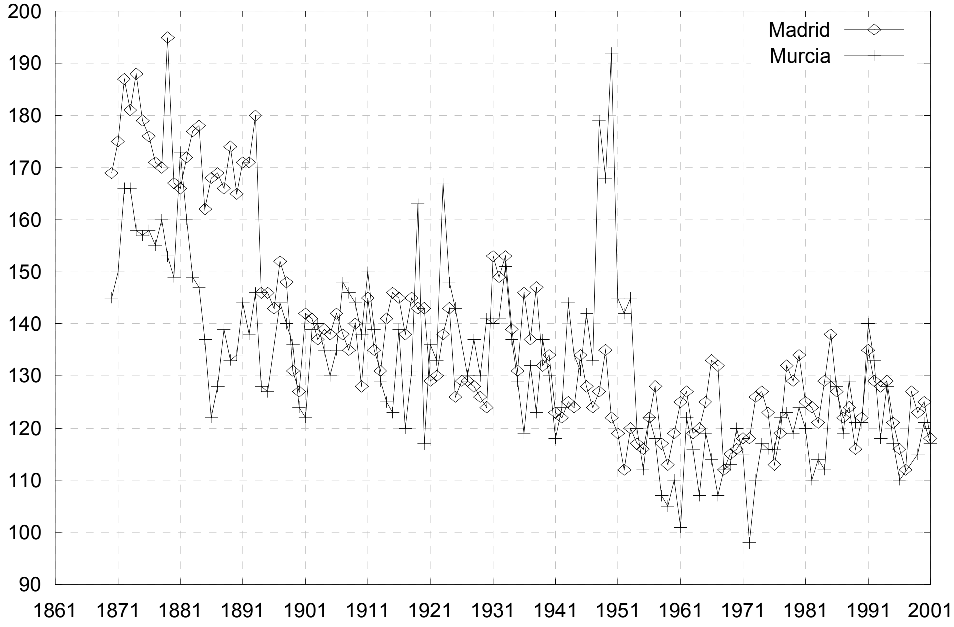


Figure 12: Monthly average of DTR for August for Madrid, Spain and Murcia Spain. Data in 1/10 ºC. 19th century data shows the lack of a Stevenson shelter, which produces higher daytime and lower nighttime temperatures, resulting in an increased DTR. Up to 1884, the effect is clearly present at both stations, so they cannot constitute a valid reference for each other for 1871-1884.

3.3.3 **Breakpoint identification**

The fourth step is to search for breakpoints in the difference or ratio between the candidate and the reference time series (or alternatively in the data when a suitable reference cannot be built), compare them with the available metadata and decide which discontinuities will be indeed regarded as true inhomogeneities. Some methods do not actually search for breakpoints and only use the reference time series to decide if the changes found in the station history produce an effect in the data large enough to require adjustment. This is a good approach only if the metadata are believed to be complete and up to date. Common statistical tests for samples comparison, like the *t-test* or rank-based alternatives (if data normality is in doubt), are adequate to decide when dealing with events that produce a sudden jump, like instrument replacements or relocations. Regression analysis can be used when looking for artificial trends, like those derived from urbanization, gradual change to irrigated crop fields around station or growing trees producing a shadow.

Several other methods are used to search for breakpoints in data. Usually, this is achieved by performing a set of statistical tests between two data samples composed by consecutive data, moving the contact point between them one element at a time. An *n-sized* dataset can be searched for breakpoints by using a fixed size window (comparing data points *1 to 10* vs. *11 to 20; 2 to 11 vs. 12 to 21; …; n-20 to n-11 vs. n-10 to n*) or by using a varying sized window (comparing data points *1 to 10 vs. 11 to n; 1 to 12 vs. 13 to n;…; 1 to n-11 vs. n-10 to n*). Although repeated statistical tests increase the type-I error probability, each test assesses the likelihood of the last element in the first sample to constitute a breakpoint. Some approaches can detect more than one discontinuity at the same test run some others are designed to find only one at a time. As there certainly can be more than one discontinuity in the data, the firstly labelled point is used then to split the time series in two pieces which are searched again for further discontinuities. For this purpose, the *t-test* or similar formulations based in the evaluation of the change in mean have been widely used. Other approaches adjust a regression line to the data before and after the year being tested and evaluate the change in slope. Finally, some techniques are based in the use of rank order change point detection, like the Wilcoxon-Mann-Whitney test. This particular approach is advisable when the normality of the data is in doubt. Normality is usually more difficult to ensure when dealing with precipitation, and is always more easily achieved in year averaged or accumulated quantities than in monthly data. In such cases it can be helpful to apply a data transformation, like cube roots, before homogenizing to achieve normality.

Some techniques only run the test once, trusting the reference to be homogeneous, while others engage in an iterative procedure in which all the stations in the data set are seen consecutively as candidates and references. This is done to produce some preliminary homogenized data, which will be used in the final homogenization process.

When analyzing data from a station, we have to keep in mind that, if we run the selected test on the 12 monthly series, it is possible to find different breakpoints in each one. This is quite understandable, due to the randomness of the time series and because it is obvious that some causes of inhomogeneity can have a larger impact in summer than in winter or the other way around. Same thing may happen when comparing daytime and nighttime temperatures. For this reason, it is good to start the analysis over averaged quantities like annual means, which have also less year-to-year variability and usually allow a better detection. In some cases, like temperature, it is recommended to take seasonally averaged data instead of annual means to account for opposed summer-winter effects.

It also needs to be mentioned that detection tests have less power in detecting breakpoints near the start and end of a series.

3.3.4 **Data adjustment**

Once the breakpoint identification is finished, the next step is to decide which breakpoints are going to be accepted as real inhomogeneities. First, it is always advisable to look for a feasible physical cause in the metadata. When not found, expert judgment will be needed to distinguish between real breakpoints and false positive tests. A good tool for this is to plot the difference or ratio time series, as their visual inspection allows the user to subjectively asses the feasibility of the breakpoints found and, conversely, to account for non-detected problems. Indeed, some skilled scientists do not use any statistical testing to find their breakpoints and rely only in the evaluation of the difference or ratio time series plots. Sometimes, the data are plotted in the form of cumulative sum of anomalies for a better inspection. It has to be said that most of the aforementioned detection techniques could also be applied directly to the station data in the absence of adequate references, but in this case actual fluctuations of climate are likely to be detected along with inhomogeneities, making metadata analysis and subjective decision even more unavoidable.

The last and final step is to adjust the assessed discontinuities. It is always recommended to correct the data to match the conditions of its most recent homogeneous section. By doing so, incoming data in the future will still be homogeneous unless further changes occur in the station. The most usual approach to obtain adjustment factors is to calculate separate averages on the difference or ratio series for the two sections defined by a breakpoint. Then, the obtained means are compared by calculating their ratio or difference and the obtained factor is applied to the inhomogeneous part. . This is appropriate in the case of sudden shifts. When dealing with gradual inhomogeneities, the best approach is to de-trend the inhomogeneous section using the slope calculated on the difference or ratio time series. ***Figure 13*** shows the impact of adjustment in a temperature series.

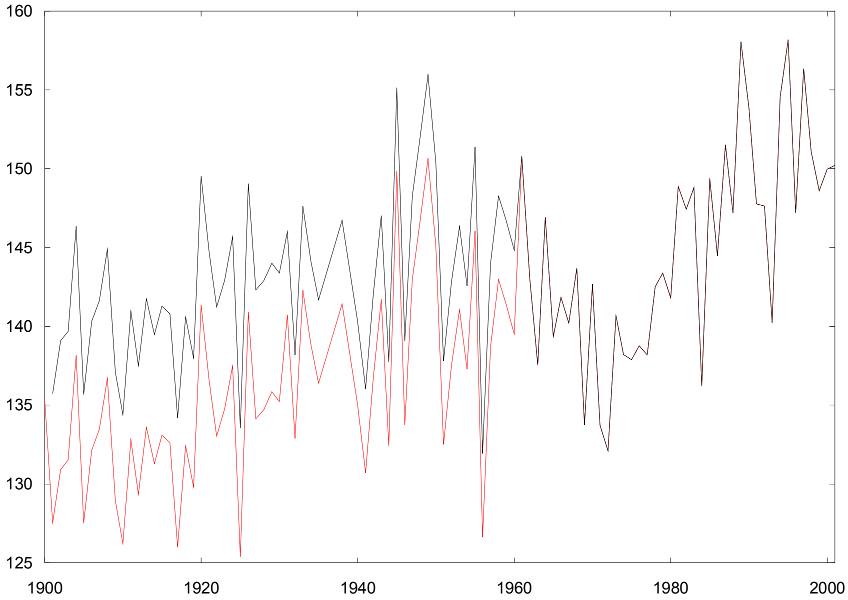


Figure 13: Original (red line) and adjusted (black line) annual averages of daily mean temperature for Madrid, Spain. Data in 1/10ºC. Data was adjusted for sudden shifts in mean and artificial trends using an iterative test which compares the mean value of two different periods over a standardized reference time series, calculated from a number of well-correlated reference stations. Inhomogeneous data (red line) show a much larger trend for the 100 years period, as they contain true climate fluctuations plus artificial biases. Figure modified from Aguilar, E (2002) "Homogenizing the Spanish Temperature Series", personal communication to the 7th *National Climatology Meeting*, Albarracín, Spain.

3.4 RECOMMENDATIONS FOR HOMOGENIZATION OF MONTHLY, SEASONAL AND ANNUAL DATA

Previous sections dealt with the subject of monthly, seasonal and annual data homogenization. This section recaps some aspects that should not be forgotten when deciding between different homogenization methods.

The first issue is the goal of the researcher. In some cases, when trying to identify high quality stations, it may not be necessary to adjust data, but only check the metadata for a first selection of *a priori* good stations and confirm their quality with data plots. On the other hand, when homogenizing a whole network to create a new dataset, using a more sophisticated approach which allows detection and correction will be needed.

Another key issue is metadata availability. If metadata are assumed to be complete and up to date (which is quite difficult to ensure), some researchers may feel inclined to not use a detection technique but rather to directly calculate the adjustment factors. If direct comparison data are available, they should be used to derive the adjustment factors, rather than by comparing the data before and after the inhomogeneity. Nevertheless, even with the best possible metadata, some statistical inhomogeneity detection is advised.

The next key issue is to check if it is possible to build a reliable reference time series. This depends on the data period being homogenized (being more difficult to find reference data for the distant past), the geographical region we are dealing with (being more difficult to build a valid reference series data for remote locations and for climates with a high degree of variability), and on what variable we are assessing (usually needing a denser network for precipitation since the spatial variability is higher). Only if a reference time series cannot be built is it advisable to apply homogenization techniques over the station data.

When deciding how to detect inhomogeneities, it is very important to keep in mind the underlying statistical distribution of the data. If we have any doubts about normality (which is more likely when working with monthly data, and in climates with a high degree of year-to-year variability, again especially for precipitation), it may be adequate to use a rank based test. If we do not, a parametric approach will be better.

Last but not least, we have to evaluate our own skills, resources and time constraints. For example, experimental techniques for direct homogenization can be very expensive to run and often require a long time between starting and the first results. Techniques based on test iteration are powerful, but require high statistical skills and software availability or programming capabilities. Simple subjective techniques, especially those based on reference time series plotting, can be easier to build and, although they benefit from a high level of expertise, always help to understand which problems the data may have.

The final step in homogenization assessment should be evaluation of the results. This is unavoidable and time consuming, no matter what approach has been used. It is very important to understand what adjustment factors are applied to improve the reliability of the time series and to make measurements comparable over all their extent. Sometimes, one might need to apply a technique that has been designed for another set of circumstances (e.g. another climate, meteorological element, network density, etc.) and it is important to analyze how well the homogenization performed. For example, most techniques to homogenize monthly to annual precipitation data have been designed and tested in rainy climates with precipitation throughout the year. Applying them to other climatic conditions may be fine or, conversely, underscore hidden problems. For example, creating a reference time series using ratios can be problematic in the Mediterranean climate when encountering a zero-precipitation summer month, as division by zero will be involved.

To assess corrections, one might compare the adjusted and unadjusted data to independent information, such as data from neighbouring countries, gridded datasets or proxy records such as the date of first flowering of a plant or ice freeze and thaw dates, etc. Another approach is to examine countrywide area averaged time series for adjusted an unadjusted data and see if the homogenization procedure has modified the time series in the direction expected after our knowledge of the station network. For example, in the U.S. there has been a wide spread change from afternoon observations (which have a warm bias) to morning observations (which have a cold bias). The unadjusted network therefore has a cold bias in the time series. The adjusted time series, as one might predict, shows more warming than the unadjusted data set.

If, the homogenization results, after having been analyzed, are accepted as valid, although some single values might remain incorrect, the newly adjusted time series will describe the time variations of the analyzed element better than the original data

Table 4 lists a series of different approaches to homogenization developed and applied by different groups/authors and ***Figure 14*** shows the different steps to be taken for the homogenization of monthly to annual time series.

Table 4: Review of different widely used techniques for inhomogeneity detection and homogenization. Descriptions and references are obtained from THOMAS C. PETERSON et *al*. (1998), except those marked with *

METHOD	DESCRIPTION
BUISHAND RANGE TEST* Busishand, T.A.. 1982. 'Some methods for testing the homogeneity of rainfall records. *Journal of Hidrology* **58**: 11-27.* Wijngaard, J.B., Klein Tank, A.M.G. and Können, G.P. 2003: 'Homogeneity of 20th century European daily temperature and precipitation series'. *Int. J. Climatol*., **23**: 679-692.*	The Buishand range test is defined as $S_k^* = \sum_{i=1}^{k} (Y_i - \bar{Y})$ where *k= 1, ..., n.* When a series is homogeneous the values of will fluctuate around zero, because no systematic deviations of the Yi values respect to their mean will appear. If a break is present in year K, then S_k^* reaches a maximum (negative shift) or a minimum (positive shift) near the year *k* = *K*. The significance of the shift can be tested $R = \left(\max_{0\le k\le n} S_k^* - \min_{0\le k\le n} S_k^* \right) \Big/ s$, where s is the standard deviation of S_k^*. Buishand (1982) gives critical values for the test.
CAUSSINUS-MESTRE TECHIQUE Caussinus, H. and Lyazrhi, F. 1997. 'Choosing a linear model with a random number of change-points and outliers', *Ann. Inst. Stat. Math*., **49**, 761–775. Caussinus, H. and Mestre, O. 1996. 'New mathematical tools and methodologies for relative homogeneity testing', Proceedings of the Seminar for Homogenization of Surface Climatological Data,	The Caussinus-Mestre method simultaneously accounts for the detection of an unknown number of multiple breaks and generating reference series. It is based on the premise that between two breaks, a time series is homogeneous and these homogeneous sections can be used as reference series. Each single series is compared to others within the same climatic area by making series of differences (temperature, pressure) or ratio (precipitation). These difference or

Budapest, 6–12 October, pp. 63–82. Mestre, O. and Caussinus, H., 2001. A Correction Model for Homogenisation of Long Instrumental Data Series, in Brunet, M. and López, D., Detecting and Modelling Regional Climate Change, Springer, pp13-19* Caussinus, H., and Mestre, O. (in press). 'Detection and correction of artificial shifts in climate series'. Submitted to JRRS, series C. *	ratios series are tested for discontinuities. When a detected break remains constant throughout the set of comparisons of a candidate station with its neighbours, the break is attributed to the candidate station time series. Recently, the authors have developed a new technique based in the comparison of several perturbed series instead of comparing a series to an artificial reference. A two factor linear model (time x series) is introduced for all series at any time, and a penalised likelihood procedure to select the best model. To facilitate computation, a stepwise approach is adopted.
CRADDOCK TEST Craddock, J.M. 1979. 'Methods of comparing annual rainfall records for climatic purposes', *Weather*, **34**, 332–346. Auer, I., 1992. Experiences with the Completion and Homogenization of Long-term Precipitation Series in Austria, Centr. Europ. research. initiative, Proj. Gr. Meteorology, Wp. 1, Vienna.	Developed by Craddock (1979), this test requires a homogeneous reference series though sometimes long enough homogeneous sub-periods are sufficient (Boehm, 1992). The Craddock test accumulates the normalized differences between the test series and the homogeneous reference series according to the formula: $s_i = s_{i-1} + a_i*(b_m/a_m) - b_i$ where a is the homogenous reference series, b is the time series to be tested and a_m and b_m are the time series means over the whole period.
EXPERT JUDGEMENT METHODS Jones, P.D., Raper, S.C.B. Bradley, R.S. Diaz, H.F. Kelly, P.M. and Wigley, T.M.L. 1986a. 'Northern Hemisphere Surface Air Temperature Variations: 1851–1984', *J. Climate Appl. Meteorol.*, **25**, 161–179. Rhoades, D.A., and Salinger, M.J. 1993. 'Adjustment of temperature and rainfall records for site changes', *Int. J. Climatol.*, **13**, 899–913.	Judgement by an experienced climatologist has been an important tool in many adjustment methodologies because it can modify the weight given to various inputs based on a myriad of factors too laborious to program. For example, when viewing a graphical display revealing a station time series, a neighbouring station time series and a difference series (candidate-neighbour), a subjective homogeneity assessment can factor in the correlation between the stations, the magnitude of an apparent discontinuity compared to the variance of the station time series, and the quality of the neighbouring station's data along with other information such as the relevance and reliability of the available station metadata. Expert judgement can be particularly helpful in an initial inspection of the stations' data and when the reliability of certain inputs (e.g. metadata) varies.
INSTRUMENTS COMPARISONS Forland, E.J., Allerup, P., Dahlstrom, B., Elomaa, E., Jonsson, T., Madsen, H., Per, J., Rissanen, P., Vedin, H. and Vejen, F., 1996: Manual for Operational Correction of Nordic Precipitation Data, DNMI-Reports 24/96 KLIMA, 66 pp. Nichols, N., Tapp, R., Burrows, K., and Richards, D., 1996. 'Historical thermometer exposures in Australia', *Int. J. Climatol.,* **16**. Quayle, R.G., Easterling, D.R., Karl, T.R. and Huges, P.Y., 1991. 'Effects of recent thermometer changes in the cooperative station network', *Bull. Amer. Met. Soc.*, **72**.	Side by side comparisons are useful to derive the impact of instrument substitutions on data homogeneity. They have been used to assess the difference between shielded and non-shielded rain gauges or Stevenson Screens and other stands. Other approaches are based in statistical comparisons of sets of stations using simultaneously different instruments, for example liquid-glass thermometers and maximum-minimum thermistors. In all cases, the goal is to derive correction factors to subtract the impact of the instrument substitution on data

MULTIPLE ANALYSIS OF SERIES FOR HOMOGENISATION (MASH) . Szentimrey, T. 1996. ‘Statistical procedure for joint homogenization of climatic time series’, Proceedings of the Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, pp. 47–62. Szentimrey, T. 1999: ‘Multiple Analysis of Series for Homogenization (MASH)’, Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.* Szentimrey, T. 2000: ‘Multiple Analysis of Series for Homogenization (MASH). Seasonal application of MASH (SAM), Automatic using of Meta Data’, Proceedings of the Third Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary. *	The MASH does not assume the reference series are homogeneous. Possible break points and shifts can be detected and adjusted through mutual comparisons of series within the same climatic area. The candidate series is chosen from the available time series and the remaining series are considered as reference series. The role of the various series changes step by step in the course of the procedure. Depending on the climatic elements, additive or multiplicative models are applied. A new multiple break points detection procedure has been developed which takes the problem of significance and efficiency into account. This test obtains not only estimated break points and shift values, but the corresponding confidence intervals as well. A special part of the method is appropriate to homogenize the monthly, seasonal and annual series together. The developed version makes possible to use some metadata information – in particular the probable dates of break points – automatically.
MULTIPLE LINEAR REGRESSION Gullett, D.W., Vincent, L. and Malone, L.H. 1991. Homogeneity Testing of Monthly Temperature Series. Application of Multiple-Phase Regression Models with Mathematical Change points, CCC Report No. 91–10. Atmospheric Environment Service, Downsview, Ontario. 47 pp. Vincent, L. 1998. ‘A technique for the identification of inhomogeneities in Canadian temperature series’, *J. Climate*, **11**, 1094–1104.	The technique is based on the application of four regression models to determine whether the tested series is homogeneous, has a trend, a single step, or trends before and/or after a step. The dependent variable is the series of the tested station and the independent variables are the series of a number of surrounding stations. To identify the position of a step, the third model is applied successively for different locations in time, and the one providing the minimum residuals sum of squares represents the most probable position in time of a step in the tested series. The procedure consists of the successive application of the four models (Vincent, 1998).
PETTIT TEST* Pettit, A.N. 1979. ‘A non-parametric approach to the change-point detection’. *Applied Statistics* **28**: 126-135.* Wijngaard, J.B., Klein Tank, A.M.G. and Können, G.P. 2003: ‘Homogeneity of 20th century European daily temperature and precipitation series’. Int. J. Climatol., 23: 679-692.*	This test is a non-parametric rank test. The ranks $r_1, \ldots, r_n$ of a time series $Y_1, \ldots, Y_n$ are used to calculate the statistics $X_k = 2\sum_{i=1}^{k} r_1 - k(n+1)$ where $k = 1, \ldots, n$. If a break occurs in the year E, the statistic is maximal or minima near the year $k = E$., then $X_E = \max_{1 \le k \le n} \lvert X_k \rvert$. Significance tables are provided by Petit (1979)
POTTER’S METHOD Plummer, N., Lin, Z. and Torok, S. 1995. ‘Trends in the diurnal temperature range over Australia since 1951’, *Atmos. Res.*, **37**, 79–86. Potter, K.W. 1981. ‘Illustration of a new test for detecting a shift in mean in precipitation series’, *Mon. Wea. Rev.*, **109**, 2040–2045.	Is a likelihood ratio test between the null hypothesis that the entire series has the same bivariate normal distribution and the alternate hypothesis that the population before the year being tested has a different distribution than the population after the year in question. This bivariate test closely resembles a double mass curve analysis. One part of the test statistic depends on all points on a time series while another part depends only on the points preceding the year in question. The highest value of the test statistic will be in the year preceding a change in the mean of

	the candidate station time series. Potter (1981) applied this technique to ratio series of candidate station's precipitation and a composite reference series
RADIOSONDE DATA* Free, M., Durre, I., Aguilar, E., Seidel, D. Peterson, T.C., Eskridge, R.E., Luers, J.K. Parker, D., Gordon, M., Lanzante, J., Klein, S., Christy, J., Schroeder, S., Soden, B., McMillin, L., and Weatherhead, E., 2001: 'Creating Climate Reference Datasets. CARDS Workshop on Adjusting Radiosonde Temperature Data for Climate Monitoring' , *Bull. Amer. Meteor. Soc*. **83**, 891-899. *	During the past 60 years, radiosondes have been launched around the globe to collect information on vertical profiles of temperature, humidity and other atmospheric variables. The radiosonde record provides more detailed vertical resolution and a longer history than the satellite record. Archived time series of radiosonde measurements can often be plagued by inhomogeneities that compromise the validity of trends calculated from data. Currently, various groups are working to identify and remove these inhomogeneities to make the data more suitable for climate studies. The adjustment of radiosonde data also requires the identification of artificial discontinuities in the data, estimation of the size of these discontinuities, and application of adjustments. But due the special characteristics of upper air data, special strategies have to be adopted. The reference quoted in this table provides insights on different approaches and further references.
RANK-ORDER CHANGE POINT TEST Siegel, S. and Castellan, N. 1988. Nonparametric Statistics for the Behavioural Sciences, McGraw-Hill, New York, 399 pp. Lanzante, J.R. 1996. 'Resistant, robust and nonparametric techniques for the analysis of climate data. Theory and examples, including applications to historical radiosonde station data', *Int. J Climatol*., **16**, 1197–1226.	Using a test based on the ranks of values from a time series has the benefit that it is not particularly adversely affected by outliers. Lanzante (1996) describes such a non-parametric test related to the Wilcoxon-Mann-Whitney test. The test statistic used is computed at each point based on the sum of the ranks of the values from the beginning to the point in question (Siegel and Castellan, 1988). And the maximum value is considered the point of a possible discontinuity.
STANDARD NORMAL HOMOGENEITY TEST Alexandersson, H. and Moberg, A. 1997. 'Homogenization of Swedish temperature data. Part I: A homogeneity test for linear trends', *Int*. J. *Climato*l., **17**, 25–34. Alexandersson, H. 1986. 'A homogeneity test applied to precipitation data', *J. Climat*e, **6**, 661–675. Hanssen-Bauer, I., Forland, E. 1994: 'Homogenizing long Norwegian precipitation series', *J. Climate,* **7**, 1001-1013.	Is a likelihood ratio test. The test is performed on a ratio or difference series between the candidate station and a reference series. First this series is normalized by subtracting the mean and dividing by the S.D. In its simplest form, the SNHT's statistic is the maximum of $Tv = v(\bar{z}_1)^2 + (n-v)(\bar{z}_2)^2$, where $\bar{z}_1$ is the mean for the series from data point 1 to v and $\bar{z}_2$ is them mean of the series from *v+1* to the end, *n*. There are now variations in this test to account for more than one discontinuity, testing for inhomogeneous trends rather than just breaks, and inclusion of change invariance
STOP-TREND METHOD* Kobysheva, N and Naumova, L. 1979: Works of the Main Geophysical Observatory, 425, Saint Petersburg, Russia.*	A non-parametric test. Data are sorted by date and consecutive ranks are assigned. Then, the time series is split into *k*= $n^{0.5}$, where *n* is the number of observations, with size *l=(max-min)/k.*, where *max* and *min* are the maximum and minimum value in the dataset. Inside each interval, if the difference between

	consecutive ranks exceeds a critical level based on the Kolmogorov coordination criterion, the observation corresponding to the first rank is labelled with and "A" and the second with a "B". Once all the intervals have been evaluated, if two adjacent observations in the overall time series are flagged with "A" and "B", the "A" observation defines a breakpoint.
TWO-PHASE REGRESSION Solow, A. 1987. 'Testing for climatic change: an application of the two-phase regression model', J. Climate Appl. Meteorol., 26, 1401–1405. Easterling, D.R. and Peterson, T.C. 1995a. 'A new method for detecting and adjusting for undocumented discontinuities in climatological time series', Int. J. Climatol., 15, 369–377. Easterling, D.R. and Peterson, T.C. 1995b. 'The effect of artificial discontinuities on recent trends in minimum and maximum temperatures', *Atmos. Res.*, **37**, 19–26.	Solow (1987) described a technique for detecting a change in the trend of a time series by identifying the change point in a two-phase regression where the regression lines before and after the year being tested were constrained to meet at that point. Since changes in instruments can cause step changes, Easterling and Peterson (1995a,b) developed a variation on the two-phase regression in which the regression lines were not constrained to meet and where a linear regression is fitted to the part of the (candidate-reference) difference series before the year being tested and another after the year being tested. This test is repeated for all years of the time series (with a minimum of 5 years in each section), and the year with the lowest residual sum of the squares is considered the year of a potential discontinuity.

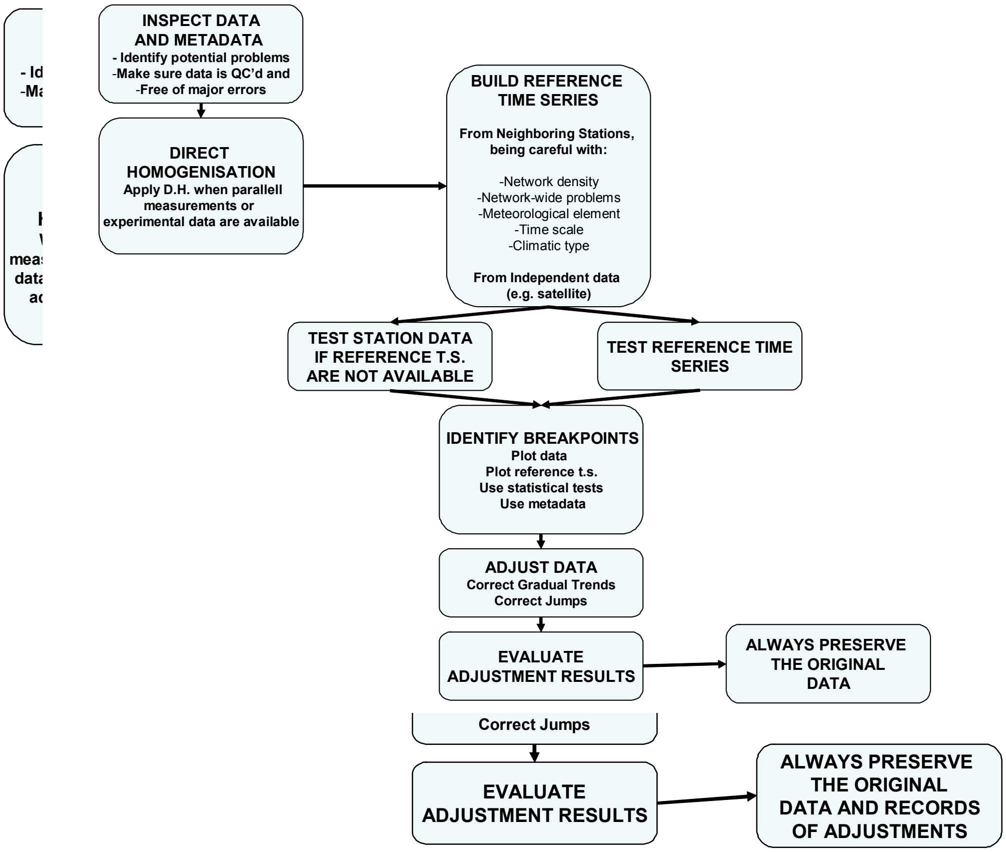


Figure 14: Schematic representation of homogenization procedures for monthly to annual climate records

3.5 HOMOGENEITY OF SUB-MONTHLY DATA

All of the approaches described up to this point work have been applied to annual or monthly data. Unfortunately, sub-monthly data such as daily or hourly observations present a whole new set of problems. Part of the reason for this is that one of the uses of homogeneous daily data is assessing changes in extremes. Extremes, no matter how one defines them, are rare events that often have a unique set of weather conditions creating them. With few extreme data points available for the assessment, determining the proper homogeneity adjustment for these unique conditions can be difficult.

Consider an example of the inhomogeneity in maximum daily temperature observations caused by changing from liquid-in-glass thermometer in a wooden slatted shelter to an electronic thermistor in an aspirated shelter. In large portions of the world, extremely warm days are likely to occur in sunny calm conditions. In these situations, wooden slated shelters tend to overheat as the sun is shining strongly on it and natural ventilation is at a minimum. On the other hand, extremely cold maximum temperatures in many parts of the world are associated with cloudy, rainy, and windy conditions. In such cases, the wooden shelter does not overheat. Therefore, accurate homogeneity adjustments for the cold and warm extremes would likely be quite different from each other and from non-extreme data observations.

This makes dealing with homogeneity of daily and hourly data quite difficult. Many groups around the world are exploring different approaches to solving this problem. The following are some of the approaches that are currently employed.

The first is to perform a general assessment of the homogeneity and remove the worst stations from consideration (Peterson et al, 2002) Figure ***15*** is an example of this. Indices for the number of days when maximum temperature was below the 10th percentile (extremely cold days) were calculated for all the stations in the region. Subjectively assessing this type of figure, the researchers determined that one station clearly had an inhomogeneity (heavy blue line). Therefore, maximum temperature data from this station was not used in area averaged analyses (heavy red line).

Another approach employed by the European Climate Assessment (ECA) project (Wijngaard *et al.*, 2003) to detect the usefulness of daily time series consists in evaluating the results of four homogeneity tests (Standard Normal Homogeneity Test, Busihand range, Pettitt and Von Neumann ratio tests). These tests are applied to carefully chosen testing variables, which are in annual resolution, but contain characteristics of the daily series. Subsequently, the four tests results for a given series are ranked in an overall classification, which provides an indication of the validity of the series for trend and variability analyses.

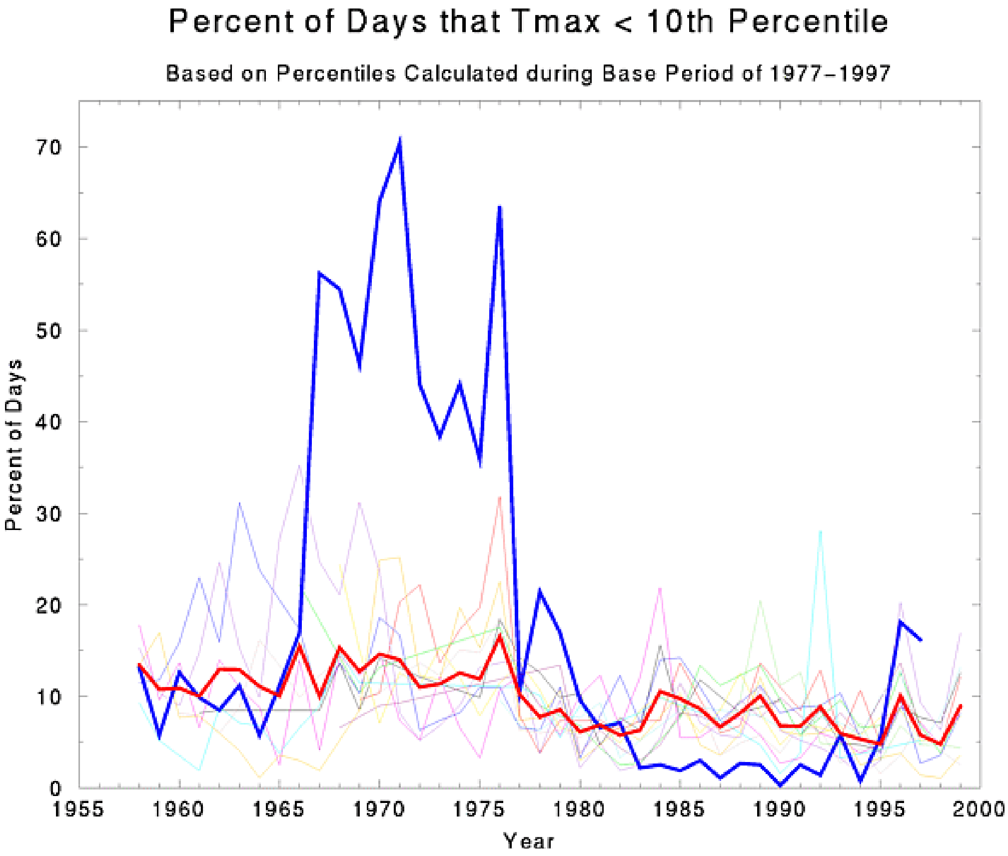


Figure 15: Percent of days where daily maximum temperature is greater than the 10th percentile for the base period 1977-1997 for 16 stations in the Caribbean Region. Subjective assessment of the graph of this index indicates that one station has serious discontinuities in its maximum temperature time series. Peterson, Thomas C., Michael A. Taylor, Rodger Demeritte, Donna L. Duncombe, Selvin Burton, Francisca Thompson, Avalon Porter, Mejia Mercedes, Elba Villegas, Rony Semexant Fils, Albert Klein Tank, Albert Martis, Robert Warner, Antonio Joyette, Willis Mills, Lisa Alexander, and Byron Gleason, 2002: Recent Changes in Climate Extremes in the Caribbean Region. *Journal of Geophysical Research*, **107**(D21), 4601, doi: 10.1029/2002JD002251 (Nov. 16, 2002).

Another approach, which also does not make homogeneity adjustments, is to determine homogeneous sections of a time series. This can be done through careful subjective assessments of the data combined with knowledge of the station history metadata. For example, if a station with data from 1 January 1950 to 31 December 1990 changed instruments on 7 July 1980, and that was the only potential inhomogeneity that could be identified, then the time series could be split into two homogeneous sections, one from 1 January 1950 to 6 July 1980 and the other from 8 July 1980 through 31 December 1990 (assuming the instrumentation making the observations for 7 July 1980 was not known). These two time series could then be treated in various analyses the same way one treats data from two different stations with short periods of record.

The aforementioned techniques try to prevent contaminated data to go into the working data set. Other approaches go one step beyond and try to adjust the sub-monthly data. For example, the Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources (IMPROVE) project (Moberg *et al.*, 2002; Bergrstrom *et al.*, 2002) adjusted daily data using interpolated values from monthly adjustments and essayed creating different adjustments based on weather classifications. So a minimum temperature observation in December would get one adjustment if it was clear and another if it was cloudy or rainy.

This is a step in the right direction as it makes the daily data more homogeneous. But sub-monthly data homogenization has to keep in mind the existence and importance of extreme values. The Australian Bureau of Meteorology homogenized a daily temperature dataset of 103 records, intended for the study of trends in extreme events (Trewin, 2001). Initially gross errors in each record where detected and removed using internal consistency checks, feasibility ranges and neighbour comparisons. Inhomogeneities were identified by comparing each candidate series with a reference series calculated from a weighted mean of highly correlated neighbours. A two-phased regression model (Easterling and Peterson, 1995) was used to detect discontinuities in the difference series which were confirmed visually or with available metadata. Discontinuities were then corrected by matching the frequency distribution of daily temperatures on either side of the inhomogeneity, allowing different magnitudes of adjustment at 5th, 10th, …, 90th and 95th percentile levels (Trewin and Trevitt, 1996). Finally, spatial analysis of the adjusted temperature data was used to identify stations whose data differed substantially from those of its neighbours.

Another group has developed sub-monthly adjustments based on the distribution of the observations. However, in order to get enough observations to make fairly robust adjustments, they divided the distribution only into thirds. So the top, middle, and bottom thirds of the observations would get different adjustments. Again this is a significant step in the right direction, but is unlikely to fully adjust the extremes.

In summary, homogeneity adjustments of daily and hourly data are very difficult. So difficult in fact, that no recommendations regarding adjusting daily or hourly data will be made. However, careful analysis of the data, and particularly derived parameters that one may be using in an analysis (e.g., indices of extremes), can help identify the worst problems. These assessments can be as straightforward as subjective assessments of graphs and station history metadata. It is recommended that, at a minimum, before using daily or hourly data in any long-term climate change analysis, that the data are carefully evaluated for the impacts of inhomogeneities and that portions of time series with homogeneity problems be excluded from the analysis.

3.6 A FINAL STATEMENT: DOCUMENTATION OF THE HOMOGENIZATION PROCEDURES AND DATA PRESERVATION

We have discussed a wide range of approaches to achieve data homogeneity. Homogenization procedures have the ultimate goal of adjusting all the data to the measuring conditions of the most recent observation. Complete success is almost impossible but, through the application of good homogenization practices, the results can be very good and allow a better analysis and understanding of climate. Nevertheless, there is not yet a definitive approach, and possibly there never will be. So even the most carefully homogenized data can be reviewed and their adjustments improved. For this reason it is very important to always preserve the original data.

When data are quality controlled and/or homogenized, it is always important to document the procedures applied in the metadata. When transmitting a dataset it is desirable to provide both the original and the adjusted data. This will be very helpful for other users who may need to apply a different homogenization approach to meet their particular requirements.

To recap, homogenization is a difficult but unavoidable task. By properly adjusting a station or dataset we gain a better understanding of climate and especially of climate variability and change.

4 FINAL REMARKS

Metadata are important for putting observations into proper perspective, for understanding the biases that might be inherent in the observations and the changes in the biases over time. Therefore, metadata should be as complete as possible, as up to date as possible, and as readily available as possible (which generally means storing in digital form rather than relying solely on paper archives).

For putting current observations into an accurate historical perspective for climate change studies, the homogeneity of the data needs to be assessed. More specifically, inhomogeneities should be identified and dealt with. There are many approaches to dealing with inhomogeneities in the climate record. The best methods depend on many factors such as the level of detail in the metadata, the human and computational resources available, and the parameters being assessed.

5 GLOSSARY

Breakpoint: in a time series, starting point of an inhomogeneity.

Candidate Station: station to be homogenized

CIMO: Commission on Instruments and Methods of Observation.

Data set: collection of data belonging to one or several stations of a region of the (or the entire) globe, expanding for a period of time and containing information for one or several meteorological parameters.

Data adjustment: correction applied to data to improve their homogeneity and to make all the observation comparable to the last available data.

Difference and ratio time series: difference or quotient between a candidate station and its reference time series.

DTR: diurnal temperature range, calculated as difference between daily maximum and daily minimum temperature

Exposure: particular conditions in which meteorological instruments are exposed and/or sheltered to register atmospheric conditions.

GCOS: Global Climate Observing System (IOC/WMO/ISSU).

Geographical Data: metadata that indicate where the station is located.

Homogeneous Time Series: climate data time series where variability and change responds only to true climate variability and change and not to other biases.

Homogenization: procedure of making a time series homogeneous by using a technique to remove artificial bias.

Instruments: devices intended to measure a meteorological element.

IPCC: Intergovernmental Panel on Climate Change (WMO/UNEP)

Land use and Land cover: type of utilisation of the ground and characteristics of the materials that compose it, with emphasis here in the area surrounding the meteorological enclosure

Maintenance: routine and non-routine operations performed on a meteorological station to keep it in good condition.

Metadata set: collection of metadata usually referred to a data set.

Metadata: Data about the data, necessary to correctly understand and use meteorological data.

NMHS: National Meteorological Hydrological Services.

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

Parallel measurement: simultaneously measured paired data collected to assess the effect of an inhomogeneity.

Relational database: collection of data stored with the aid of computer software capable of linking different data through key common fields.

Reference Time Series: an independent, homogeneous time series used to assess homogeneity of a candidate station, usually derived from neighbouring stations

Spreadsheet: computer software commonly used for numerical analysis purposes and capable of storing data in a tabular fashion

Shelter: refuge where meteorological instruments are stored to adequately record atmospheric conditions following WMO standards

Station Identifiers: metadata that indicate which station data belong to.

Urbanization: degree of building up an area dedicated to an urban use, which induces an artificial and gradual bias on meteorological records

6 BASIC REFERENCES

Arnfield, J.D., Duchon, C. E., Baker, C. B., Quayle, R. G., Heim Jr., R. R., Robbins, K. D., Janis, M. J., and Horvitz. May, A. H. 2000. 'U.S. Climate Reference Network, Part 4: Metadata.' Prepr. 12th AMS Conference on Applied Climatology. (Asheville, NC., USA): 30-33

Bergström, H., and Moberg, A. 2002: 'Daily air temperature and pressure series for Uppsala' (1722-1998), *Climatic Change*, **53**: 213-252

Easterling, D.R. and Peterson, T.C. 1995. 'A new method for detecting undocumented discontinuities in climatological time series'. Int. J. Climatol., 15: 369-377.

Jones, P. D., Raper, S.C.B., Santer, B., Cherry, B.S.G., Goodess, C., Kelly, P. M. Wigley, T.M.L., Bradeley, R. s. and Diaz, H. F. 1985: A grid point surface air temperature data set for the Northern Hemisphere, US Dept. of Energy, DOE/EV/10098-2, Washington, D.C.

Linacre, E. 1992. "Climate data and resources, a reference and guide" Publ. Routledge, U.K., 366 PP., ISBN 0-415-05702-7, 0-415-05703-5 PBK.

Moberg, A., Bergström, H., Ruiz Krigsman, J and Svanered, O. 2002: Daily air temperature and pressure series for Stockholm (1756-1998), *Climatic Change*, **53**: 171-212.

Peterson, T.C., Easterling, D.R., Karl, T.R., 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, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. e and Parker, D. 1998. 'Homogeneity adjustments of in situ atmospheric climate data: a review', *Int. J. Climatol*., **18**, 1493-1517.

Szalai, S., Szentimrey, T. and Szinell, C. (eds.) 1998: Proceedings of the Second Seminar for homogenization of surface climatological data, Hungarian Meteorological Service, WMO and NCTD, Budapest.

Trewin, B.C. and Trevitt, A.C.F. 1996. 'The development of composite temperature records'. *Int. J. Climatol*., **16**: 1227-1242.

Trewin, B.C. 2001. Extreme temperature events in Australia. PhD Thesis, School of Earth Sciences, University of Melbourne, Australia.

Verkaik, J.W. 2000. 'Evaluation of two gustiness models for exposure correction calculations', J. Applied Meteor., **39**: 1613-1626.

Wijngaard J.B., Klein Tank A.M.G., and Können G.P. 2003. 'Homogeneity of 20th century European daily temperature and precipitation series.' Int. J. Climatol., 23 : 679-692.

WMO. 1986. Guidelines on the selection of Reference Climatological Stations (RCSs) from the existing Climatological Station Network. WMO/TD-No. 130, World Meteorological Organization, Geneva.

WMO. 1996. Guide to Meteorological Instruments and Methods of Observation (6th edition), WMO-No.8., World Meteorological Organization, Geneva.

WMO. 2002. WMO Technical Document 1125, GCOS-76.

GCOS. 2002. Guide to the GCOS Surface and Upper-Air Networks: GSN and GUAN (Version 1.0), GCOS Report No. 73, World Meteorological Organization, Geneva.

Ackowledgements. We would like to acknowledge the following people who reviewed the text and made useful comments (alphabetically by institution): Dean Collins, Neville Nichols and Neil Plummer (Bureau of Meteorology, Australia); Anders Moberg, Phil Jones (Climatic Research Unit, University of East Anglia, UK); Volker Vent-Schmidt (Deutscher Wetterdienst); Chris Folland (Hadley Centre, United Kingdom); Tamasz Szentimrey (Hungarian Meteorological Institute); Carlos Alamrza and José Antonio López (Instituto Nacional de Meteorología, Spain); Jitze van der Meulen and Janet Wijngaard (Koninklijk Nederlands Meteorologisch Instituut, The

Neetherlands); Elena Akentjeva and Nina Kobysheva (Main Geophysical Observatory, Russia); Øyvind Nordli (Meteorologisk Institutt, Norway); Jeff Arnfield (National Climatic Data Center, USA); Lourdes Tibig (Philipine Atmospheric Geophysical and Astronomical Services Administration).