Chapter 1

Basics of Climatological and Meteorological Observations for GIS Applications

Weather and climate data are spatially distributed. Geographical information technologies can therefore provide a useful and relevant working environment for the distribution, integration, visualization, and analysis of these data. However, compared to other scientific areas, the application of geographical information system (GIS) tools was for a long time a clumsy process within meteorology and climatology, and especially within most national meteorological services (NMS); because of the shortcomings of GIS related to the underlying data model and missing interfaces to standard meteorological tools (e.g. weather forecast model). While the GIS data models are highly static based, meteorological data models have a need for a strong dynamical component with causal dependencies in the space/time domain (see for example [CHR 02]). Nativi et al. [NAT 04] describe the differences between both underlying data models and advocate models that are supported by so-called interoperability services. In addition to these differences in the data models, there are significant differences in the spatial modeling approaches. In general, GIS environments have implemented the geo-statistical modeling tools that are based on one temporal realization only, whereas meteorological data offer the temporal sample in addition to the spatial sample, which results in different spatial modeling approaches [SZE 04]. However, within the last few years efforts for integration of meteorological data models in GI environments were quite successful, and well-established GIS web-mapping standards and spatial infrastructures have gained increasing importance in meteorology and climatology. Thus parallel efforts and development currently appear to be resolved [SHI 05].

Information to be derived from climate variability analyses is strongly dependent not only on the spatiotemporal density, but also on the quality of the available data. Today it is a well-established fact in climatology that the climate signal from measurements, beside the statistical noise, is by inhomogenities. Therefore, a primary step of climate studies is to analyze the input data used with respect to

Chapter written by Wolfgang SCHÖNER.
quality and homogeneity, which makes the results uncertain regarding input data, in addition to the uncertainty of the approach. Quantifying quality and homogeneity of the data require data about the data itself (the metadata). Nowadays, metadata are highly standardized for GI data (e.g. see the Open GIS Consortium activities), but the information obtained from metadata regarding climate data is still heterogenous. However, the NMSs and the WMO (World Meteorological Organization) are aware of the importance of climate metadata, which resulted in several efforts for standardization of metadata (see e.g. [AGU 03]). To summarize these efforts, it can be stated that in climatology metadata information is related to the documentation of the “where”, “when”, and “how” of measurements, whereas metadata in GI science also add emphasis to the usability of the data.

In this chapter, basic concepts of climate networks and climate data are presented. This includes an overview of standards of climate measurements, description of climate data types, spatial reference of data, as well general comments on accessing the data. The areas of climate data quality and homogeneity are reviewed in depth, covering the important aspects of metadata description. The chapter does not tackle climate model data and only introduces climate reanalysis data.

1.1. Data measurements and observations in climatology

1.1.1. Networks and concepts for meteorological/climate data

Meteorological measurements are motivated by the primary aim of predicting the Earth’s weather with the highest possible precision. This aim results in measurements covering the entire Earth (for both the land and the sea), but also encompassing the third dimension (vertical sounding of the atmosphere by radio sounds, satellite sensors, radar, etc). Beside weather forecasting, meteorological services are responsible for monitoring the state and spatiotemporal change of the climate. As these two basic aims do not coincide with respect to network performance, two different networks have been established in public weather services, the synoptic and the climate network. Whereas, the stations and instruments are identical, the networks differ in their interval, quantity, availability and time of observations. Moreover, the synoptic network is characterized by the need for a much larger spatial extent and more detailed information on past and current weather situations. In contrast, climate networks are characterized by higher demands on data quality. All national meteorological/climatological networks are coordinated on an international level by the WMO.

The need for meteorological/climatological networks is met by in situ measurements and by remote sensing techniques. Consequently, the WMO Global Observing System is composed of the surface-based subsystem and the space-based subsystem. The surface-based subsystem includes different types of station networks (e.g. surface synoptic stations, climatological stations), whereas the space-based subsystem comprises, for example, on-board sounding from spacecraft. The
observational requirements of a climatological station or synoptical station are
detailed in [WMO 03] and include: present weather, past weather, wind direction
and speed, cloud amount, cloud type, cloud-base height, visibility, air temperature,
relative humidity, air pressure, precipitation, snow cover, sunshine duration or solar
radiation, soil temperature, evaporation.

Figure 1.1. Layout of an observing station in the northern hemisphere showing minimum distances between installations (from [WMO 08])
An important concept behind climate observations is representativeness, which is the degree to which the observation accurately describes the value of the variable needed for a specific purpose. Therefore, it is not a fixed quality of any observation, but results from joint appraisal of instrumentation, measurement interval and exposure against the requirements of some particular application [WMO 08]. An estimate of spatiotemporal representativeness of air temperature and precipitation is shown in Figure 1.2, with much higher spatial correlation for air temperature compared with precipitation. It can be concluded from this results that station density has to be much higher for precipitation compared with air temperature and that station density has to be increased for investigations with increasing temporal resolution.

![Figure 1.2. Average decorrelation distances ($r^2$ decreasing below 0.5) for air temperature and precipitation in four time resolutions. Samples: daily values for all of Europe; monthly, seasonal and annual for the Greater Alpine Region (from [AUE 05])](image)

Various meteorological applications have their own preferred timescale and space scale for averaging, station density, and resolution of phenomena. From there, for example, weather forecast requires more frequent observations compared to climate monitoring. The spatio-temporal dependency of meteorological phenomena results in simple scaling convention (see Table 1.1).

<table>
<thead>
<tr>
<th>Type of motion</th>
<th>Spatial scale (m)</th>
<th>Temporal scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eddy</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Micro turbulence</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Tornado</td>
<td>100</td>
<td>60</td>
</tr>
<tr>
<td>Cumulus convection</td>
<td>1,000</td>
<td>20 min</td>
</tr>
<tr>
<td>Cumulunimbus</td>
<td>100,000</td>
<td>1 h</td>
</tr>
<tr>
<td>Front</td>
<td>100,000</td>
<td>3 h</td>
</tr>
<tr>
<td>Hurricane</td>
<td>100,000</td>
<td>3 h</td>
</tr>
<tr>
<td>Cyclone</td>
<td>1,000,000</td>
<td>1 d</td>
</tr>
<tr>
<td>Planetary waves</td>
<td>10,000,000</td>
<td>10 d</td>
</tr>
</tbody>
</table>

Table 1.1. Spatial and temporal scales of meteorological phenomena
The design of a meteorological station has to be according to the network requirements. In particular, the station site, instrument exposure and location of sensors has to be treated according to regulations. As an example, Figure 1.1 shows the layout for a typical synoptic/climatological station according to WMO regulations.

1.1.2. Standards for climate data measurements

The term “standard” is related to the various instruments, methods, and scales used to estimate the uncertainty of measurements. Amongst others, nomenclature for standards of measurements is given in the International Vocabulary of Basic and General Terms in Metrology issued by the International Organization for Standardization (ISO) [ISO 93]. The following standards are included: measurement standard, international standard, national standard, working standard, transfer standard, traceability, etc.

Meteorological observations and measurements are highly standardized from WMO or from NMSs. Such standardization is obvious if we take into account the influence of station surroundings (surface properties, influences from nearby buildings, trees, etc) or of measurement observation procedures. From there meteorological (climatological) measurements are standardized especially with respect to:

– surface conditions in the nearby of the sensor;
– station surrounding;
– sensor-height above ground;
– procedure of reading;
– observation time.

However, practices are different and measurements are occasionally performed under conditions that are different from the required standard, which have to be archived in metadata information. This is especially true for the surface conditions in the areas around the sensor and the station, whereas sensor height and observation are generally in accordance with the standards. Standards are more accurately considered in climate networks of weather services compared with networks from other operators. When incorporating data from various other sources, the standardization regulations of the data providers should be carefully considered.

1.1.3. Climate data types

Classification of data types can be undertaken from different perspectives. Using classical classification schemes used in GI science the following types of data are used in meteorology and climatology.
– **Spatial irregularly distributed point data**: e.g. the station measurements and observations, vertical radio sounding data if some generalization is taken into account;
– **Raster data**: e.g. the different field from weather forecast models or from climate models;
– **Image data**: e.g. satellite data, weather radar data.

Meteorological data can also be classified into scalar data (air temperature) and vector data (wind with wind speed and wind direction). According to the classical basic of statistics meteorology/climatology include all types of scales of measurements:

– **Nominal scale**: e.g. cloud type, present weather, weather type;
– **Ordinal scale**: e.g. cloud density;
– **Interval scale**: e.g. air temperature;
– **Ratio scale**: e.g. precipitation, air pressure.

In addition to these statistical or GIS-related classification schemes, there are also such from meteorology/climatology schemes based on the idea of a Global Observing System [WMO 08]:

– Surface-based subsystem: comprises a wide variety of types of stations according to the particular application (e.g. surface synoptic stations, upper-air stations, climate stations);
– Spaced-based subsystem: comprises a number of spacecraft with on-board sounding missions and the associated ground segment for command, control and data reception.

1.1.4. **Access to climate data: spatial data infrastructure in meteorology and climatology**

Since the foundation of the NMS, the weather forecast has been highly dependent on efficient spatial data infrastructure, which today is called the Global Telecommunication System (GTS) and covers the entire Earth. Station observations and other data are shared with GTS worldwide within the hour according to standardized regulations, in order to get a “snapshot” of the current state of atmosphere and weather conditions and as an input for weather forecast models. The GTS data infrastructure is highly standardized and secures the data transfer between the NMSs, but does not fully meet the needs of the increasing number of users outside the NMSs’ networks. As a result, international NMSs networks, such as EUMETNET (The Network of European Meteorological Services), established projects to address this, e.g. UNIDART (uniform data request interface, http://www.dwd.de/UNIDART). During the last few years, and based on GTS, WMO initiated the WIS (WMO Information System), which distributes information globally for real-time weather forecasting and climate monitoring using a service-oriented architecture.
The operability of the GTS is dependent on data exchange and a related policy of data holders. Today, each National Meteorological and Hydrological Service (NHMS) has its own data access policy ranging from free access to highly commercially oriented data selling. Even within the European Union (EU), meteorological data policy is quite heterogenous and the exchange of data between NMHS, apart from for weather forecast purposes, such as for climate monitoring, is sometimes limited. Generally, important information on meteorological data is provided in table or map form by basic metadata of the station network, which provides information about location, geographical coordinates, altitude, sensor equipment and data availability, etc., without any charge from NMHS. Easy access to such information is still not guaranteed, but there is a move towards providing a greater amount of information without restriction.

In Europe, the idea of spatial data infrastructure (SDI) was substantially supported by the INSPIRE (http://inspire.jrc.it/home.html) initiative. INSPIRE is an EU directive that forces EU member states to provide spatial data to different users according to OGC’s SDI standards. As a result of INSPIRE and as a general need of climate research, European NMSs started with efforts to meet INSPIRE needs. Within the frame of EUMETNET, the EUROGRID was formulated with its first step as a showcase (S-EUROGRID, see www.eurogrid.eu, [KLE 08]). EUROGRID aims to provide a SDI for climate data according the OGC standards. In addition to this multinational initiative, climatological/meteorological SDIs were established on national levels. SeNorge, a common meteorological and hydrological effort in Norway, is a good example (www.senorge.no). Due to the user-friendly data policy in Norway, SeNorge not only displays climate data fields on a monitor screen according to the OGS WMS (Web Map Service) standard, but users can also obtain and integrate data of interest according to the WFS (Web Feature Service) and WCS (Web Coverage Service) standard. These OGC standards for web-mapping have received substantial interest in the field of meteorology over the last few years.

Another well-established OGC standard used in addition to meteorological and climatological applications is the Google Earth KML format for many web services. Integration of OGC-compliant spatial infrastructure for distribution of climate data received much earlier support in the USA compared with Europe. The NOAA (National Oceanic and Atmospheric Administration), and in particular NCAR (National Centre for Atmospheric Research), supported the OGC ideas of interoperability for meteorological data. Special attention was given to the ArcGIS Atmospheric Data Model, a collaborative initiative among ESRI, UCAR, NCAR, Raytheon, Unidata, and NOAA. The ArcGIS Atmospheric Data Model aims to represent each of these data objects in a uniform manner, enabling their superposition and combined analysis in the ArcGIS desktop environment. For the first time, the ArcGIS 9.2 [ESR 09] release supported both the NetCDF and HDF-5 data format through a new tool from the ArcGIS toolbox list. Both the NetCDF and HDF data models are commonly used in atmospheric sciences, e.g. data fields from climate model runs are available in NetCDF. Through this data model, a fundamental linkage between the GI community and atmospheric sciences community was established.
Figure 1.3. Simplified scheme of OGC compliant web services (GDAL stands for Geospatial Data Abstraction Library, figure adapted from [VAN 08])

In particular, the GALEON IE (Geo-interface for Atmosphere, Land, Earth and Ocean netCDF) Interoperability Experiment supports open access to atmospheric and oceanographic modeling and simulation outputs. The geo-interface to netCDF datasets is established by the Web Coverage Server (WCS 1.0) protocol specifications. Additionally, UNIDATA unified the OpenDAP, netCDF and HDF5 data models to the new CDM (Common Data Model) and introduced a new API (application programming interface), NcML, an XML (extensible mark-up language) representation of netCDF using XML syntax. On a long-term perspective, GALEON will analyze FES (Fluid Earth Sciences) requirements for simple and effective interface specifications to access datasets and will define a more general data model for CF-netCDF. This new data model should include non-regular data grids and should establish metadata encodings (e.g. Climate Service Modeling Language CSML, ncML-G). CSML is a standard-based data model described in Unified Modeling Language (UML), and an XML mark-up language that
implements this data model [WOO 06]. The model describes climate science data (e.g. observational data, model runs) at the level of the actual data values; CSML is not a high-level discovery metadata model [LOW 09]. An example of a simplified structure of an OGC compliant web service for integration of meteorological/climatological data in geospatial services or applications is shown in Figure 1.3.

Another major OGC initiative with increasing interest from meteorology is Sensor Web Enablement (SWE). The ultimate goal of SWE is to make all kinds of sensors discoverable, accessible, and controllable via the web, which should result in “plug-and-play” web-based sensor networks. Beside others, SWE include Sensor Observation Service (SOS), Sensor Planning Service (SPS), and Sensor Alert Service (SAS). SOS aims to provide access to observations from sensors in a standardized way that is consistent for all sensor systems, including remote, in situ, fixed and mobile sensors.

As mentioned previously, the time dimension is an important domain in meteorology and climatology not adequately covered by GIS (see e.g. [WOO 05]). Moreover, climatology and weather forecast are highly interested in slices of time, showing climate fields on axes of latitude and time or longitude and time. Such diagnostic slices are required in future GIS standards. In addition to the time dimension, the representation of gridded meteorological fields could result in problems. For instance, meteorological grids can be non-regularly spaced or, in the case of models formulated in spectral coordinates, could have fewer longitudinal grid-points towards the poles. These shortcomings need to be addressed in the future by additional cooperating standardization work between GIS and meteorology.

1.1.5. Spatial reference for climate data

The position of climatological/synoptic station has to be measured in the World Geodetic System 1984 (WGS-84) or Earth Geodetic Model 1996 (EGM96). The coordinates of a station includes [WMO 08]:

a) the latitude in degrees with a resolution of 1 in 1,000;

b) the longitude in degrees with resolution of 1 in 1,000;

c) the altitude of the station above mean sea level to the nearest meter.

The elevation of the station is defined as the altitude above mean sea level of the ground on which the rain gauge stands or, if there is no rain gage, the ground beneath the thermometer screen. If there is neither a rain gauge nor screen, it is the average level of terrain in the vicinity of the station. If the station reports air pressure, the elevation of air pressure sensor must be specified separately.

Within the last few years, the increasing number of spatial modeling tools with increasing spatial resolution used in meteorology and climatology also enforced the pressure on the accuracy of station coordinates. Previously station coordinates were digitized from topographical maps; currently, station coordinates are surveyed by
GPS measurements. In addition to the spatial reference in geographical coordinates according to WMO, there are still a great number of national reference systems in use. However, the increasing use of the UTM system will overcome this variety of spatial reference systems in the future. Problems from different national reference systems could appear in the case of merging datasets (especially gridded fields) from different data holders, which could result in certain differences in overlapping areas.

In addition to altitude, several other vertical coordinate systems are used in meteorology including pressure, isentropic, or terrain-following coordinates, which are used for upper-air observations or for weather and climate models. Besides upper-air observations, such coordinated systems are used for weather and climate models. Such systems are not established in the traditionally predominately two-dimensional GIS world [WOO 05]. Providing the full richness of vertical coordinate systems will be an important requirement for full integration of GIS in meteorology and climatology, and thus, an important area of OGC activity.

1.1.6. Climate reanalysis data

Climate reanalysis aims to produce meteorologically consistent datasets of the atmosphere covering the entire Earth with state-of-the-art methods. In particular, they combine the full set of meteorological observations, including, e.g. surface stations, radio sounds, and satellite data with weather forecast models using data assimilation methods. Climate reanalysis datasets are among the most important climate datasets in climate research, including climate impact studies. Standard data formats for climate reanalysis are GRIB or NetCDF. Due to the NetCDF data format, these datasets are already standardized for direct use in GIS applications (see section 1.1.4). Climate reanalysis data are provided in Europe by the ECMWF (European Center for Medium-range Weather Forecast, UK) from the following projects: ERA15 covering the period 1979-1993 and ERA40 covering the period 1957-2002. In the USA, reanalysis projects have been run by the NOAA and NASA within: NOAA-NCEP covering the period 1948 onwards and NASA/DAO covering the period 1980-1995, and from Japan Meteorological Agency: JRA-25 covering the period 1979 onwards.

New reanalysis projects are currently under way (ERA interim) or planned (NCEP, JRA). In addition to meteorological consistency, the most important product of reanalysis data is their full spectrum of data covering the entire atmosphere in similar way as weather forecast models in high temporal resolution (e.g. 6 hourly fields for ERA40), but also with similar spatial resolution of the gridded fields.

Climate reanalysis is derived by data assimilation methods, which is today a four-dimensional (4D) variational analysis in the case of ERA [AND 08]. 4D-Var performs a statistical interpolation in space and time between a distribution of meteorological observations and an a priori estimate of the model state (the background). This is done in such a way that the dynamics and physics of the
forecast model is taken into account to ensure the observations are used in a meteorologically consistent way. The idea behind 4D-Var data assimilation is shown in Figure 1.4. For a single parameter $x$ the observations are compared with the short-range forecast from a previous analysis over a 12-hour period. The model state at the initial time is then modified to achieve a statistically good compromise, $x_a$, between the fit $J_b$, to the previous forecast, $x_b$, and the fit $J_0$ to all observations within the assimilation window. $J_b$ and $J_0$ are referred to as cost functions [AND 08]. This 4D-Var approach replaced earlier approaches that were based on the optimum interpolation method.

![Figure 1.4. The idea of 4D-Var data assimilation technique, see the text for a detailed explanation (from [AND 08])](image)

Climate re-analyses are also subject to a detailed validation against independent observations. A detailed description of climate reanalysis goes far beyond the scope of this chapter; for more details the interested reader should refer to the literature (e.g. [UPP 04]).

### 1.1.7. Climate data providers outside NMHs

Climate data are not only provided by meteorological and hydrological services but also by other data providers, in particular, universities. The majority of these data centers provide surface climate data that are either station data or gridded data. Usually, these data providers use data from NMHs and improve their data quality or spatial coverage. From the GIS perspective, standardization of these datasets is still weak and both data formats and metadata are quite heterogenous. Consequently, the import of the data to GIS needs some data preparation. However, OGC standardization of the NetCDF format is expected to overcome this shortcoming in the near future.

A major climate data provider is the Climate Research Unit (CRU) from the University of East Anglia (UK). In particular, CRU provides long-term climate data with global coverage, which is an important base for global climate monitoring.
1.2. Data quality control and data homogenization in climatology

1.2.1. The importance of data quality control and homogenization

Data quality control (DQC) is applied to detect errors in the process of recording, manipulating, formatting, transmitting and archiving data. DQC is not identical to homogenization as homogenization goes far beyond the aims of DQC. For homogenization, long-term series of climate data are needed, which enable non-climatic breaks to be detected in the series resulting from changes of station location, observer, observation time, sensor type, station surrounding, etc. In fact, homogenization is a two-step procedure including detection of breaks with statistical tests and adjustment of breaks.

<table>
<thead>
<tr>
<th>Country</th>
<th>Data provider</th>
<th>Means calculus</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Central Inst. for Meteorology and Geodynamics</td>
<td>ZAMG $(t_1 + t_9 + 2*t_{14})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td></td>
<td>Hydrographical Service (yearbooks)</td>
<td>HZB $(t_1 + t_9)/2$</td>
<td>LMT</td>
</tr>
<tr>
<td>Bosnia and Hertegovina</td>
<td>Federal Meteorological Inst.</td>
<td>Meteo BiH $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td></td>
<td>Federal Meteorological Inst. (historic Yugoslavian yearbooks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>Meteorological and Hydrological Service of Croatia</td>
<td>DHMZ $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Czech Hydrometeorological Inst.</td>
<td>CHMI $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td>France</td>
<td>Météor-France</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>German Meteorological Service</td>
<td>DWD $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>1961-86: LMT 1987-90: CET +30'</td>
</tr>
<tr>
<td>Hungary</td>
<td>Hungarian Meteorological Service</td>
<td>OMSZ $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td>Italy</td>
<td>Italian National Research Council, Inst. of Atmospheric Sciences and Climate</td>
<td>ISAC-CNR $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>University of Milan, Dept. of Physics</td>
<td>UNIMI $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>University of Padua, Treeline Ecology Research Unit</td>
<td>UNIPD $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>University of Pavia, Dept. of Territorial Ecology and Terrestrial Environments</td>
<td>UNIPV $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>University of Turin, Dept. of Agronomy, Forest and Land Management</td>
<td>UNITO $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Giancarlo Rossi, private data collection</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Italian Meteorological Society, Aosta Valley, Piedmont</td>
<td>SMI $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>University of Turin, Department of Earth Science, Piedmont</td>
<td>UNITO $(t_1 + t_9)/2$</td>
<td>-</td>
</tr>
<tr>
<td>Slovenia</td>
<td>Environmental Agency of the Republic of Slovenia, Climatological Dept.</td>
<td>ARSO $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td>Slovakia</td>
<td>Slovak Hydrometeorological Inst.</td>
<td>SHMU $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>LMT</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Federal Office of Meteorology and Climatology</td>
<td>Meteo Swiss $(t_1 + t_{14} + 2*t_{12})/4$</td>
<td>CET +30'</td>
</tr>
</tbody>
</table>

Table 1.2. Example of the heterogeneity of air temperature station networks for the Greater Alpine Region GAR used for spatial modeling of climate normal fields 1961-90 (tn=mean daily minimum temperature; tx=mean daily maximum temperature; TRM=true mean; CET=Central European Time; LMT=local mean time, adapted from [HIE 09])
As a minimum requirement, a yes/no answer is recommended to indicate whether DQC has been applied or not. If the answer is positive, it would be good practice to describe the degree of DQC applied to the data (e.g. subjected to logical filters only; compared for internal coherency in sequence of observations, for spatial consistency among suitable neighboring stations, for coherency with its climatological values and limits) and to provide details on the employed techniques and their application [AGU 03].

A simple example of inhomogenity in climate data series results from different approaches for the computation of daily or monthly means of e.g. air temperature from either observations at fixed times or from daily extremes of temperature (Table 1.1). The example shown in Table 1.2 is taken from the work of a new air temperature map for the Greater Alpine Region (GAR) [HIE 09] using powerful spatial modeling approaches including GIS techniques. However, before spatial modeling could be started, station measurements had to be transformed to common mean formula. Beside the formula for mean computation, the time reference system used is also heterogenous within the GAR study region. It is obvious from this simple example (which only tackles one out of several inhomogenities in climate datasets) that DQC and data homogenization, in particular, are a laborious part of climate modeling studies. Exclusion of this part of the modeling study could result in systematic biases.

![Figure 1.5](image)

**Figure 1.5.** Evolution through the year of the difference between various 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 (from [AGU 03])

Long-term series from measurements of automatic weather stations with hourly values make it easy to compute the differences between various computation formulas of daily means of air temperature used by NHMs. Selected examples of differences between commonly used mean formulas and a 24-hourly mean are shown in Figure 1.5. In fact, the widely used formula of \((\text{max}+\text{min})/2\) show differences of up to 1°C to the 24-hourly mean, which turns out to be larger than the
final standard error of spatial modeling of air temperature in the case of the GAR study example. Similarly, it was shown by many studies that inhomogenities can even exceed the climate change signal in climate time series (see e.g. [AUE 07]). It is quite easy to understand from these findings that treatment of data homogenity is essential in the analysis of spatial or temporal variability in climate data.

Although adjustment of errors originating from different means calculations can be performed quite easily in the case of longer time series from automatic weather stations, adjustment of inhomogeneity originating from e.g. urbanization effects of villages is not that simple. Urbanization does not cause a sudden break in series but instead a gradual inhomogenity trend (Figure 1.6). In the case of homogenization of the urbanization effect, it is very useful to collect information on changing building density and changing land-use.

**Figure 1.6.** 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 [BOH 98]

Another inhomogeneity in climate networks results from the change of sensors of the same type or different types. The increasing number of automatic weather stations causes such a systematic shift of sensors. Parallel measurements with both the old and the new sensor correctly merge the datasets of different sensors. However, such parallel measurements are not performed on a regular basis, and even if parallel measurements are undertaken, they are quite often undertaken over a very short period. An example of inhomogeneity from different sensors is shown in Figure 1.7 for measurement of sunshine duration in Austria, replacing the Campbell-Stokes sunshine autograph with the Haenni-Solar sensor.
Figure 1.7. Top: two types of instruments to record sunshine duration used in the Austrian meteorological network: Campbell-Stokes sunshine autograph and Haenni Solar system of automatic weather stations. Down: Consequences: Mean annual course of the breaks in Austrian sunshine series due to a change from the traditional Campbell-Stokes recorders to the Haenni-Solar sensors of the automatic network (new minus old in %, sample 1986-1999, dark: mean of four low-level sites, light: mean of three high-level sites) [AUE 01]

<table>
<thead>
<tr>
<th>Table 1.3. Result of a homogeneity study from monthly multiple climate series from the Greater Alpine Region (from [AUE 07])</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of series</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>air pressure</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>10215</td>
</tr>
<tr>
<td>141.9</td>
</tr>
<tr>
<td>258</td>
</tr>
<tr>
<td>31.1</td>
</tr>
<tr>
<td>638</td>
</tr>
<tr>
<td>4217</td>
</tr>
<tr>
<td>3.4</td>
</tr>
</tbody>
</table>
A very detailed study on climate data homogenity is available for the Greater Alpine region from several research projects (e.g. [AUE 07]). Some important results of these studies are summarized in Table 1.3 showing that data inhomogenity is immanent to climate data studies even on a very short time scale. The number of detected breaks is quite high and the mean length of homogenous sub-interval is in the range of 10-30 years for all climate variables shown. Results in this table are derived from selected monthly data series covering monthly means and monthly sums. If, however, climate extremes or daily data series are studied, the problem of data homogenity is even more pronounced.

1.2.2. Methods for climate DQC

DQC is part of the core of the whole data-flow process. In fact, it has to ensure that data are checked and is as error-free as possible. All erroneous data have to be eliminated and, if possible, should be replaced by corrected values (while retaining the original values in the database).

Useful tools of DQC for climate data are (Aguilar et al., 2003):

a) *Gross error checking*: report what kind of logical filters have been utilized to detect and flag obviously erroneous values (e.g. anomalous values, shift in commas, negative precipitation, etc).

b) *Tolerance test*: documents to which tests have been applied, to flag those values considered as outliers with respect to their own climate-defined upper/lower limits. The tests provide the percentage of values flagged and the information on the approximate climate limits established for each inspected element.

c) *Internal consistency check*: indicate whether data have undergone inspection for coherency between associated elements within each record (e.g. maximum temperature < minimum temperature; or psychrometric measurements, dry-bulb temperature ≤ wet-bulb temperature).

d) *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.

e) *Spatial coherency*: notify if any test is used to determine whether every observation is consistent with those taken at the same time in neighboring stations affected by similar climatic influences.

Figure 1.8 shows the results from a detailed homogenization study of climate time series for the GAR, which also included estimation of outliers and gap filling. Whereas the time series of outlier rates (figures on the left) indicate more about internal system stability of meteorological networks the gap rates (figures on the right) seem to react more to external influences. It is interesting to see from Figure 1.8 that both outliers and gaps increased since the 1980s, which was the beginning of automation of climate networks in the study region.
Figure 1.8. Time series of outlier rates (left panels) and gap rates (right panels) in the HISTALP series of (a) air pressure, (b) air temperature, and (c) precipitation. Outlier and gap rates in percentage (in relation to the amount of available data) (from [AUE 07]).

Generally, methods listed under a) to e) are implemented in the standard workflow of the climate data section of NMHs. However, these quality control procedures are always a compromise between strength of regulations and acceptance of outliers. Moreover, everyday practice in data handling and database management still produces erroneous data after passing the quality control procedures mentioned above. Therefore, additional efforts are necessary to improve the data quality, especially for longer-term climate data series. Such efforts should not only include a quality check according to a) to e) but also a validation against independent data as, for example:

– measured discharge against discharge at catchment level from hydrological model forced by meteorological observations;

– measured snow water equivalent or snow height against modeled data from a snow cover model forced by precipitation and air temperature;

– measured glacier mass balance against modeled data from glacier mass balance model forced by meteorological data.
Such independent validation is a powerful tool to show, for example, consistency between air temperature, precipitation, and snow height or snow water equivalent. Discrepancies can be, for example, well associated either to problems in air temperature or precipitation measurements. Similar validation approaches were also applied to climate reanalysis data.

1.2.3. Methods for climate data homogenization

Figure 1.9. Schematic representation of homogenization procedures for monthly to annual climate records (from [AGU 03])

Whereas DQC is a well-established part of the standard workflow for the database of NHMs, data homogenization still has a strong link to climate research and not to standard procedures. This originates, in part, from the fact that powerful homogenization methods were only developed during the last approximately 15 years. Today there exists various types of homogenization tools for climate data involving different homogenization philosophies and with different strengths and weaknesses regarding climate element, geographical region, or temporal resolution of the dataset ([AGU 03]). It can be concluded from literature (e.g. [AUE 07],
that various homogenization tools work well for monthly time series, but homogenization of daily series or series with even higher temporal resolution is still not solved properly.

Figure 1.10. Inhomogenities in monthly data series of air pressure, cloudiness, air temperature, precipitation and sunshine for the GAR (from [AUE 07]). Black bold lines show the mean of inhomogenity series, which implies that even averaging over many series cannot level out inhomogenities.

As previously outlined climate time series homogenization includes two important steps: first, the detection of breaks in the series (which need the creation of a reference time series), and second, data adjustment. Homogenization should be based on available metadata, which means that breakpoints should, as much as possible, be reflected by metadata information. However, many inhomogenities of climate time series are not captured by metadata (only perfect metadata would include all breaks). If available, parallel measurements from sensors, station locations, etc, should be used for adjustments. Today all homogenization tools for breakpoint detection use relative homogenity tests, which mean that they detect...
inhomogeneities from statistical evaluation of the differences or ratios to a reference series computed from neighboring stations in similar climate conditions considering statistical significance of breakpoint. Absolute homogeneity tests fail to differentiate between climatic and non-climatic inhomogeneities.

Figure 1.9 summarizes the procedure of climate time series homogenization used by most of the homogenization tools used in climatology. Also the procedure varies with the different homogenization tools; however, it is laborious work, independent of the method used. Reference series are not only used for detection of breakpoints but also for adjustment of breaks. A useful summary of available homogenization methods (for both the breakpoint detection and the adjustment procedure) is given by [AGU 03].

It is important for homogenization work to consider the underlying statistical distribution of data to be analyzed. Non-normally distributed datasets need other approaches compared with datasets with normally distributed data.

Figure 1.10 shows results from homogenization of a multiple climate dataset for the GAR. It can be concluded from this figure that climate networks contains systematic changes of the network inherent to all stations and that averaging over a larger sample of climate series can not remove inhomogeneities.

1.3. Metadata: documenting quality and usability

1.3.1. Short characteristic of metadata from a climate data perspective

Information on data, known as metadata, enables the operator of an observing system to take the most appropriate preventive, corrective, and adaptive actions to maintain or enhance data quality. Metadata, therefore, include detailed information on the observing system itself and, in particular, on all changes that occur during the time of its operation. Within the WMO, the role of ISO 19115 is stressed as the metadata standard for describing meteorological and climatological data. This standard has been developed for the geographical community, but it is also useful for meteorology as both disciplines deal with spatial data. Whereas the documentation of metadata related to the quality of the data is well established in meteorology and climatology, the description of the context of the data (e.g. access rules) is much less developed. Detailed and standardized metadata are increasingly important in the case of data access via the internet.

Elements of a metadata database (according to [WMO 08]):

A metadata database contains initial set-up information together with updates whenever changes occur. Major elements include the following:

(a) Network information:
   (i) the operating authority, and the type and purpose of the network;
(b) Station information:
(i) administrative information;
(ii) location: geographical coordinates, elevation(s);
(iii) descriptions of remote and immediate surroundings and obstacles. (It is necessary to include maps and plans on appropriate scales);
(iv) instrument layout;
(v) facilities: data transmission, power supply, cabling;
(vi) climatological description.

(c) Individual instrument information:
(i) type: manufacturer, model, serial number, operating principles;
(ii) performance characteristics;
(iii) calibration data and time;
(iv) sitting and exposure: location, shielding, height above ground;
(v) measuring or observing program;
(vi) times of observations;
(vii) observer;
(viii) data acquisition: sampling, averaging;
(ix) data-processing methods and algorithms;
(x) preventive and corrective maintenance;
(xi) data quality (in the form of a flag or uncertainty).

A simple template for the description of station exposure is shown in Figure 1.11. As ideal exposure of sensors, according to the WMO standards, is seldom available, the documentation of exposure is of high importance for the usage of data retrieved from a particular station, otherwise the reliability of the observations cannot be determined.

Metadata on meteorological data are very important because they are highly dependent on observational practices, such as type of sensor, its exposure, observational procedures, observation times, etc. The full potential of data can only be explored when sufficient metadata are available; As stated by the WMO [AGU 02]: “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”.

An important part of metadata is information on the local environment. Geographical coordinates and elevation of a station does not provide enough information in this context. Even if coordinates are documented with sufficient accuracy, the local environment can not be reconstructed. In particular, influences on station measurements act at different spatial scales. Therefore, documentation of the local environment should cover [AGU 03]:

– updated mapping in some form of the mesoscale region at ca. 1:100,000;
– toposcale map (ca. 1:5,000), updated each year, as specified by the WMO Technical Commission for Instruments and Methods of Observations [WMO 96];
– radiation horizon mapping, updated each year;
– photos taken from all points of the compass and sufficient area 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 when other significant changes occur.

![Figure 1.11. General template for station exposure metadata (from [WMO 96])](image)

A detailed picture of metadata information can be derived from Figure 1.12 for the example of the climate station at Sonnblick in the Austrian Alps. It is clear from this figure that many changes occur within short intervals, although this station is outstanding with respect to station relocations showing no changes over the entire observation period.
**Figure 1.12.** Meta quick-look for the climate station Sonnblick (Austrian Alps) (from [AUE 01]), (see color section)
Figures 1.12 shows the documented station environment within the Austrian climate network since the beginning of systematic weather observations. It shows the decreasing trend in the number of urbanized stations in the network compared with rural stations. The systematic change from manned to automatic stations can be seen in Figure 1.13. Today volunteer and professional observers are available only at a very limited number of climate stations.
1.3.2. Catalog services

Metadata are an important basis of catalog services. The OGC Catalog Service defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources. Providers of resources use catalogs to register metadata that conform to the provider’s choice of an information model (e.g. descriptions of spatial references, thematic information). Catalog services are either web-based or client applications that can be searched for geospatial data and services in very efficient ways. Catalog services need standardized metadata according to ISO. The most important catalog service in Europe is INSPIRE the European geospatial data infrastructure. Catalog services will increase in importance in meteorology and climatology, but will need considerable effort in metadata servicing and handling.

1.4. Future perspectives

The integration of meteorology and climatology into GIS is an ongoing, extremely dynamic process, for which future perspectives are hard to predict. However, it can be assumed that meteorology and climatology will strongly benefit from integration into the GIS community because of the technical progress associated with GIS. In particular, the exchange and transfer of data with users and across disciplines will be significantly improved, and thus, it will be much easier for weather and climate models to integrate relevant (and updated) datasets from other disciplines (e.g. land use, soil, vegetation). The strong innovation and technological progress in GIS and emerging tools and applications will even stimulate the already well-established tools for weather forecasting in the future. The pressure towards web-based services will also force meteorology and climatology to more standardized metadata according to ISO standards. Finally, there is currently also great interest from GIS industry to incorporate meteorology and climatology and especially considering the special demands (in the OGC standards) from these disciplines.

Spatial data infrastructures like INSPIRE will additionally push forward the integration of meteorology and climatology into GIS. Geospatial data interoperability will soon include climate and weather data. OGC standards like WMS, WCS, WFS and new standards to be developed will soon fully be adopted by meteorology and climatology, and thus, enable the incorporation of meteorological and climatological data into different services. However, an important drawback still exists because of the different data policy of the various data providers (NHMs in particular), which could be a significant barrier to further development. Aside from this impedance, Shipley [SHI 05] stated that these parallel universes are already joined, or will soon unite in the future.
1.5. Bibliography


