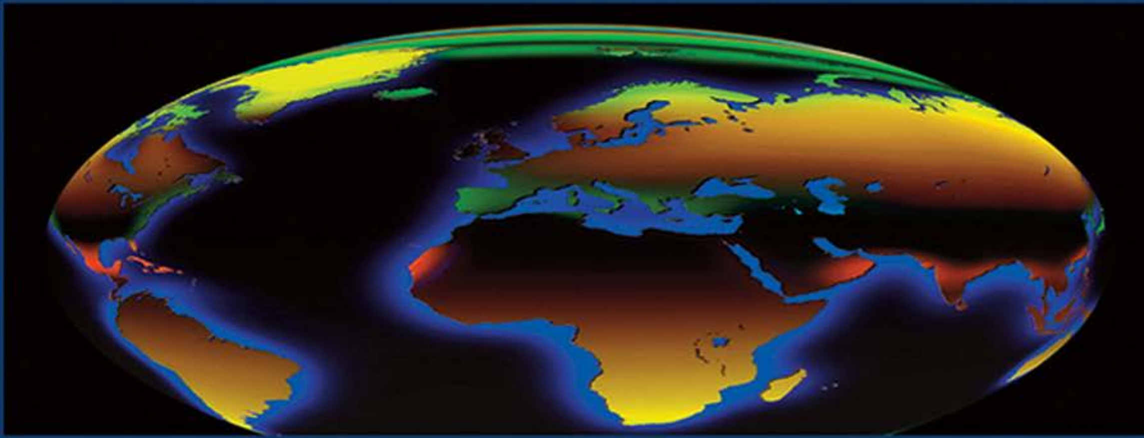
GIS

GEOGRAPHICAL INFORMATION SYSTEMS SERIES



Geographical Information and Climatology

Edited by Pierre Carrega

ISTE

WILEY

Geographical Information and Climatology

Geographical Information and Climatology

Edited by
Pierre Carrega

iSTE

WILEY

First published 2007 in France by Hermes Science/Lavoisier entitled: *Information géographique et climatologie* © LAVOISIER 2007
First published 2010 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

Information géographique et climatologie. English.
Geographical information and climatology / edited by Pierre Carrega.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-185-8
1. Climatology. 2. Geographic information systems. I. Carrega, Pierre. II. Title.
QC871.I4313 2010
551.6--dc22
2009045210

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-84821-185-8

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne

FSC
Mixed Sources
Product group from well-managed
forests and other controlled sources

Cert no. SGS-COC-2953
www.fsc.org
© 1996 Forest Stewardship Council

Table of Contents

Preface

Geographic information is used in many different themes and is also used as a source of information for a large number of different domains. The aim of this book is to highlight the relationship that exists between geographic information and the world of climatology. It is always a good idea to provide a definition of the subject that is being written about, so that readers do not have any misunderstandings or misinterpretations of the subject in question.

The word geography comes from the ancient Greek geo (Earth) and graphein (write). In the beginning the role of geography was to describe the Earth by creating maps. Maps are models, a way of representing what exists in reality. They are also seen as a model that can be used to transmit geographic information. Nowadays, however, the world of geography no longer only locates, observes and describes what is happening in an area. The term geography can also be applied to the study of human behavior and the environment, and whenever bio-physical areas are being studied the world of geography reminds us that these areas are very closely linked to man. Other disciplines study similar areas but what makes each of these individual disciplines different from one another is their "project", more than their actual subject content. The world of geography focuses less on the relationship that exists between man and nature, than on its spatial vision of certain phenomena. Space is to geography as time is to history, and for this reason many different studies have been carried out in areas at all levels, including studies carried out on a country, regional or territorial level, etc.

What exactly is geographical information? In order to answer this question several different responses are required so that the different chapters of this book can be understood. The term geographic can be understood here as being everything that relates to the Earth, to the interface that exists between the lithosphere, hydrosphere, and atmosphere, to the the Earth surface occupation and not only the land-use types. Every definition has its limits: do sub-soils, as well as the deepest water of the oceans form part of our study? The same logic can be applied to the air (i.e. what is not part of climatology?); does everything that exists in the air form part of our study? To avoid endless numbers of debates, which could occur on this subject, perhaps we need to adopt a certain level of pragmatism and link all of these different areas to one geographical space whenever these different areas are indirectly associated with one of the studies. In other words, if we are to understand these different areas, it is necessary to study them as closely as possible.

If geographic information takes into consideration the state of the surface of the Earth and its surrounding environment from a spatial point of view, then there is another important point that arises and needs to be dealt with: the type of geographic information that is to be produced. Nowadays, whenever the term geographic information is used it immediately involves the use of a tool known as a Geographic Information System (GIS). GISs are everywhere and are not only found in research laboratories (where they have been in use for a long time), but they can also be found in planning departments and in many administrative and local authorities. The important idea of linking one point or one pixel to a series of information with the aim of describing the point or pixel in the best possible way has been carried out by using powerful software. And the use of raster or vector GIS allows us to adapt to the different characteristics of the areas that are being studied.

It is difficult to state where the limit between the quantitative and qualitative worlds can be found in geographic terms. In this book, geographic information is more often than not the subject of the quantitative world, although this is not always the case. In the beginning, geographic information was considered in the widest terms possible and it included also the qualitative, as are often the metadata in climatology, for example. But it is true that the numbers (quantitative data) are much easier to process and deal with, as is shown in some of the different chapters of this book. Geographic information is a basis, a starting point for a series of sometimes complex operations that require multiple super positions or combinations so that a fixed goal can be reached. If quantification is compulsory each time a digital response is required, then the quantification process is also an impoverishment that is compulsory, and this is dealt with by some of the authors of the different chapters of this book: converting a measurement site into figures loses information, but what other method can be used? Several authors of this book and in particular, D. Joly, J.-P. Laborde, and P. Carrega, have been and are still faced with the following problem: as we have to digitalize information what method can we use in order to improve the process? The old issue of carrying out research on the field is raised once again. Some scientists think it is a necessary process, whilst others think that it is a time-wasting process. Although the different opinions of the different people concerned are based on strong arguments, they also depend on the individual person, for example, how they think, their memory and their mental understanding of their environment. The more operational scientists (in other words those who are committed to using concrete results) are normally those who carry out their research on site, in the field, at least in the short term.

Climatology is seen as being a domain that is capable of challenging geographic information. The field of climatology is an extremely large domain in which the number of climatologists has increased by a scale of 30 in a period of only 20 years. This book does not discuss the differences that exist between the worlds of climatology and meteorology. One major difference that does exist between these two worlds, and that should be mentioned, is the time scale that each of these domains focuses on. Meteorology focuses on forecasting what is going to happen over the short term (over a period of a few hours to a few days), whereas climatology focuses on defining, ranking, and describing events that have occurred over a longer period of time (regardless of whether this time refers to the past or to the future). The expression, “the climate was really nice today”, is not used and this is due to the fact that the term climate is used to describe a relatively long period of time (at least 30 years). This means that when climate is being studied, it is possible to observe key values (average, median, etc), as well as observing the distribution of these values, and thus, the extreme values. Therefore, the difference between climatology and meteorology is more functional and temporal than it is spatial. However, the limits as to where one ends and the other begins are quite unclear. As far as the future is concerned, how do we know when the notion of climatology takes over from meteorology? This is where the notion of functionality comes into force. As far as weather forecasting is concerned, there are not very many methods that are used that can provide an accurate forecast for a period of more than 15 days. The American meteorological model known as GFS publishes a weather forecast online for up to 384 hours after the current date (in other words up to 16 days after). The European model, however, does not take as many risks and publishes a weather forecast for up to 240 hours after the current date (in other words up to 10 days after). The temporal limits of physics laws, and deterministic processes, appears when we try to predict what the weather will be like for any particular day in the future because of the non-linearity of the equations that are used in forecasting, and also because of the fact that the initial state of the atmosphere is never fully known whenever the forecasts are being calculated.

There has been an undeniable amount of progress made in the world of meteorology over the last 20 years thanks to the use of such meteorological models, and the use of other complex solutions, which P. Bessemoulin describes in Chapter 4. However, these models and solutions have spatial and temporal limits. Nevertheless, this logic (and its future updated versions), is used to forecast the average state of the atmosphere in 20, 30 or even in 100 years time. This logic is also the subject of many current debates that are taking place and which deal with the following themes: what will the climate be like in the future? What do we all need to do so that climate change can be limited? What do we all need to do in order to adapt to the changes that will inevitably take place?

Climatology is also a field that is empirical and dominated by statistics, when models, which are traditionally used in the world of physics, are unable to respond to or have difficulties in responding to the demands that exist in climatology. If a new embankment is going to be built, working out its height involves considering the water levels that were measured in the area in the past. If these measurements are adjusted by Gümbel’s distribution (for example), it then becomes possible to work out the probability that a certain level of water will be exceeded, and thus its “return period”. All of this information should form part of what is known as a stationarity hypothesis, which nowadays is not normally validated.

Bringing together models from both the worlds of physics and statistics is a usefull exercise from an an intellectual point of view, although formal, because interactions between models from these two domains occur on a daily basis. Each physical model relies ever so slightly on the use of calibration coefficients that are determined by statistics, and inversely, each effective and operational statistical model that is used in climatology relies on the use of different fundamentals that stem from the world of physics.

The most common methods that are used today include: multiple regressions, and geostatistics based on spatial autocorrelation (kriging in particular), which are sometimes combined. The use of neural networks is not as widespread, and this method does not seem to solve many of the issues that people thought it would be able to a few years ago.

Remote sensing is a term that is used to group together all of the different tools that are able to record information from a distance, which is usually done using airplanes or satellites. The multiple sensors, which can be found on board these vessels, can contribute to collecting geographic information that can be used to recreate the relief of an area or to evaluate how well a particular crop is growing from a phenological point of view, etc. Sensors can also be used to measure different climatological variables, such as the temperature of the Earth's surface or the temperature of the clouds as is explained in Chapter 3 by Dubreuil. What makes remote sensing different from other methods is that it can provide data on two different pieces of information that are being researched at the same time, or at least in part.

There is one fundamental issue that affects geographic information and the relationship that it has with the world of climatology: how is it possible to make these two different domains evolve together in the future? Roussel, the author of Chapter - 6, reminds us that the geographic information produced depends on the metrological and political context in which it is used. With this in mind, different rules and regulations, as well as different socio-economic contexts and the mentality of the general public, will influence how the geographic information is used. Advances in technology in the future will probably change the way in which geographic information is measured, and as a result what is actually being measured. Will financial fluxes be a more important part of geographical information in the future?

This book is made up of eight chapters, and can be divided into two main parts.

The first part of the book is devoted to the technical aspect and the tools used to gather geographic information. In Chapter 1, Wolfgang Schoner analyses the bases of climatological observations for GIS applications, while in Chapter 2 Daniel Joly focuses on spatial analysis and cartography, and throughout the chapter he elaborates on the use of the statistical approach. In Chapter 3, Vincent Dubreuil shows how remote sensing can be used to provide us with both geographic information and information relating to the climate. In Chapter 4, Pierre

Bessemoulin provides us with an explanation of a number of key elements that are used to give us a better understanding of the way in which meteorological and climate models exploit geographic information so that they can be used effectively.

The second part of the book is devoted to how the geographic information is applied to different domains. The themes that we have chosen to focus on are associated with risks or certain constraints. The characteristics of the climate as it is today are associated with the actions of man, his needs and his limits. The research that we have carried out focuses on these limits. In Chapter 5, Maria Joao Alcoforado shows the necessity of geographical information to understand the specificity of urban climates; and, in Chapter 6 Isabelle Roussel focuses on the complexity of the relationship that exists between climatology, atmospheric pollution, and geographic information. Throughout the chapter she shares her views on what the term geographic information means and in some cases questions the term itself.

In Chapter 7, Jean-Pierre Laborde, who is a passionate hydrologist and renowned technician, proves that it is necessary to take a step back to understand exactly what a simple water flow or fload means. By taking spatialized geographic information into consideration he places a lot of importance on climatology. Finally, in Chapter 8, Pierre Carrega defines meteorological risk levels associated with forest fires. He bases his research on two different methods that can be used to generate the meteorological risk level index and compares them throughout the chapter. The two methods are both part of geographic information and the world of climatology.

Pierre CARREGA

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 SCHOENER.

quality and homogenity, which makes the results uncertain regarding input data, in addition to the uncertainty of the approach. Quantifying quality and homogenity 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 homogenity 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.

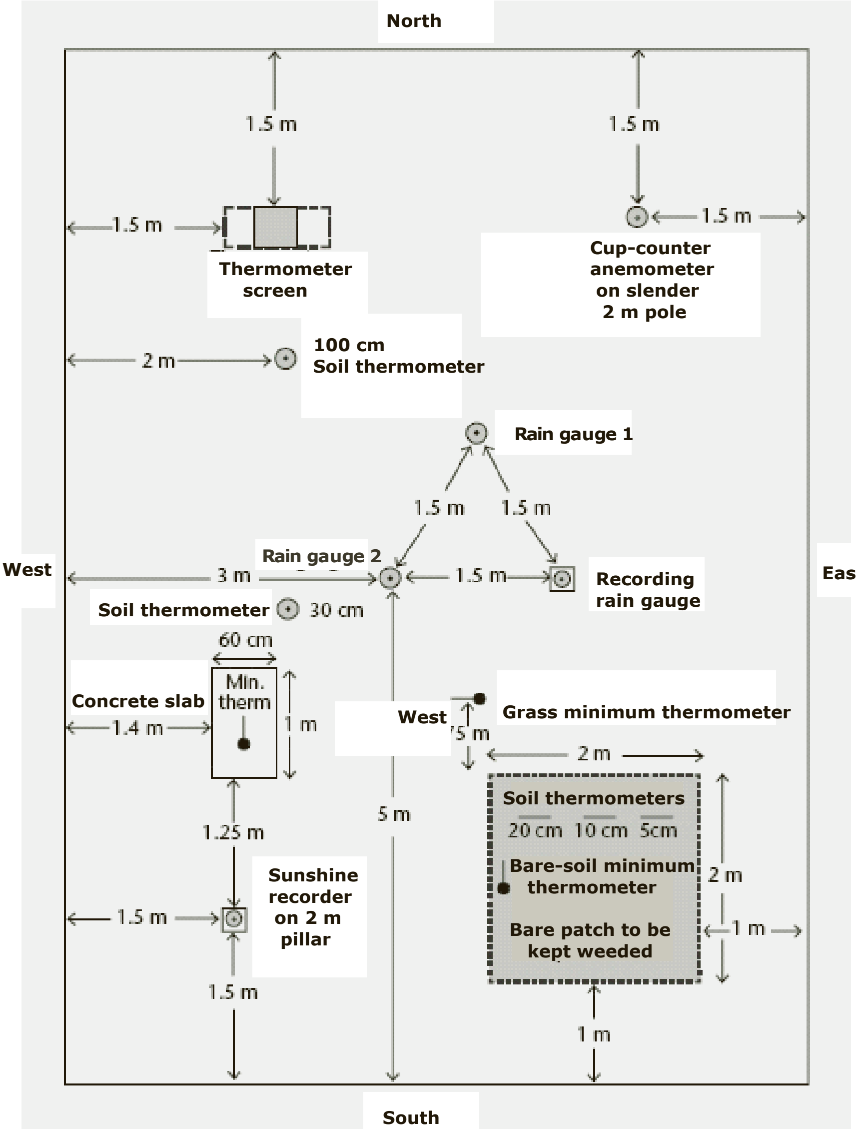


Figure 1.1. *Layout of an observing station in the northern hemisphere showing minimum distances between installations (from [WMO 08])*

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.

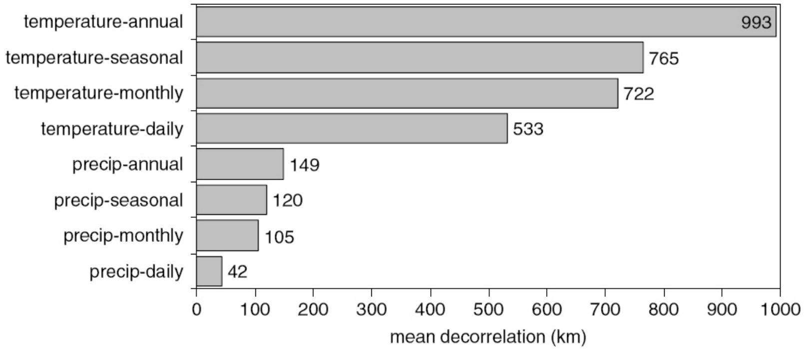


Figure 1.2. *Average decorrelation distances (r^2 decreasing below 0.5) for air temperature and precipitation in four time resolutions. Samples: daily values for all of Europe; monthly, seasonal and annual for the Greater Alpine Region (from [AUE 05])*

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).

Type of motion	Spatial scale (m)	Temporal scale
Eddy	0.001	0.001
Micro turbulence	10	10
Tornado	100	60
Cumulus convection	1,000	20 min
Cumulunimbus	100,000	1 h
Front	100,000	3 h
Hurrican	100,000	3 h
Cyclone	1,000,000	1 d
Planetaric waves	10,000,000	10 d

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 takes 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 date*: 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 Meterological 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 OGCs 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.

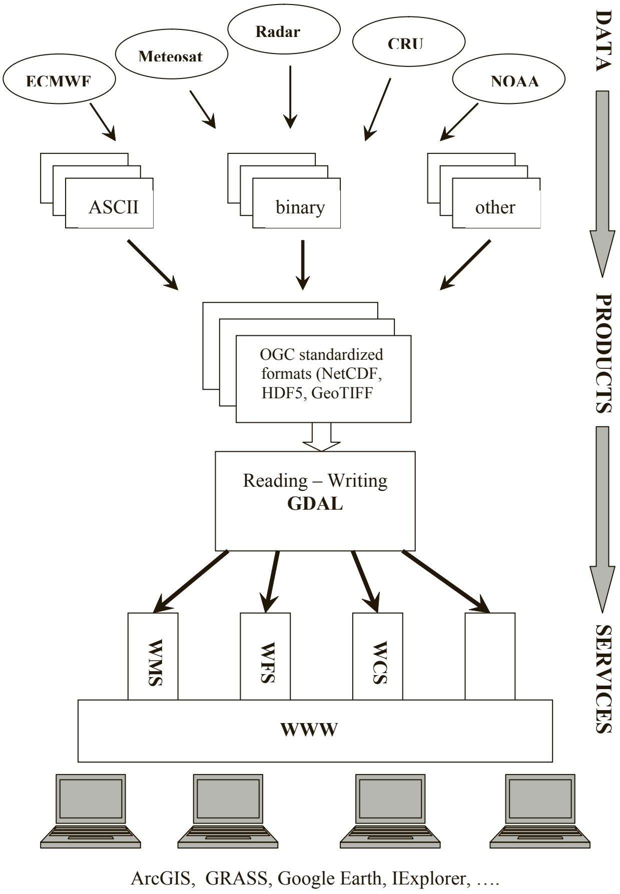


Figure 1.3. *Simplified scheme of OGC compliant web services (GDAL stands for Geospatial Data Abstraction Library, figure adapted from [VAN 08])*

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. Beside 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.

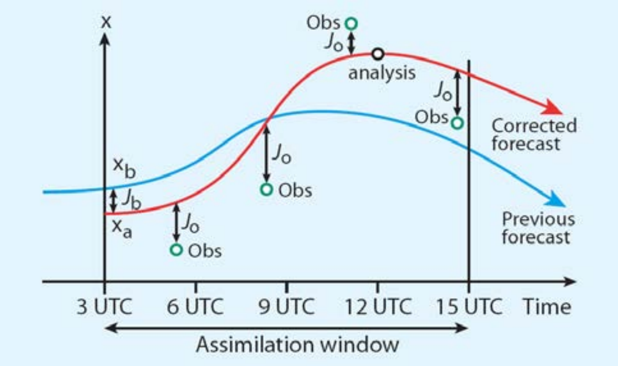


Figure 1.4. *The idea of 4D-Var data assimilation technique, see the text for a detailed explanation (from [AND 08])*

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.

Country	**Data provider**		**Means calculus**	**Time**
Austria	Central Inst. for Meteorology and Geodynamics	ZAMG	$(t_1 + t_\mu + 2*t_p)/4$	LMT
	Hydrographical Service (yearbooks)	HZB	$(t_I + t_n)/2$	LMT
Bosnia and Hertzegovina	Federal Meteorological Inst.	Meteo BiH	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
	Federal Meteorological Inst. (historic Yugoslavian yearbooks)		$(t_1 + t_{14} + 2*t_{21})/4$	LMT
Croatia	Meteorological and Hydrological Service of Croatia	DHMZ	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
Czech Republic	Czech Hydrometeorological Inst.	CHMI	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
France	Météro-France		$(t_I + t_n)/2$	-
Germany	German Meteorological Service	DWD	$(t_1 + t_{14} + 2*t_{21})/4$	1961-86: LMT 1987-90: CET +30'
Hungary	Hungarian Meteorological Service	OMSZ	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
Italy	Italian National Research Council, Inst. of Atmospheric Sciences and Climate	ISAC-CNR	$(t_I + t_n)/2$	-
	University of Milan, Dept. of Physics	UNIMI	$(t_I + t_n)/2$	-
	University of Padua, Treeline Ecology Research Unit	UNIPD	$(t_I + t_n)/2$	-
	University of Pavia, Dept. of Territorial Ecology and Terrestrial Environments	UNIPV	$(t_I + t_n)/2$	-
	University of Turin, Dept. of Agronomy, Forest and Land Management	UNITO	$(t_I + t_n)/2$	-
	Giancarlo Rossi, private data collection		$(t_I + t_n)/2$	-
	Italian Meteorological Society, Aosta Valley, Piedmont	SMI	$(t_I + t_n)/2$	-
	University of Turin, Department of Earth Science, Piedmont	UNITO	$(t_I + t_n)/2$	-
Slovenia	Environmental Agency of the Republic of Slovenia, Climatological Dept.	ARSO	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
Slovakia	Slovak Hydrometeorological Inst.	SHMU	$(t_1 + t_{14} + 2*t_{21})/4$	LMT
Switzerland	Federal Office of Meteorology and Climatology	Meteo Swiss	$(t_1 + t_{14} + 2*t_p)/4$	CET +30'

Table 1.2. *Example of the heterogenity 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.

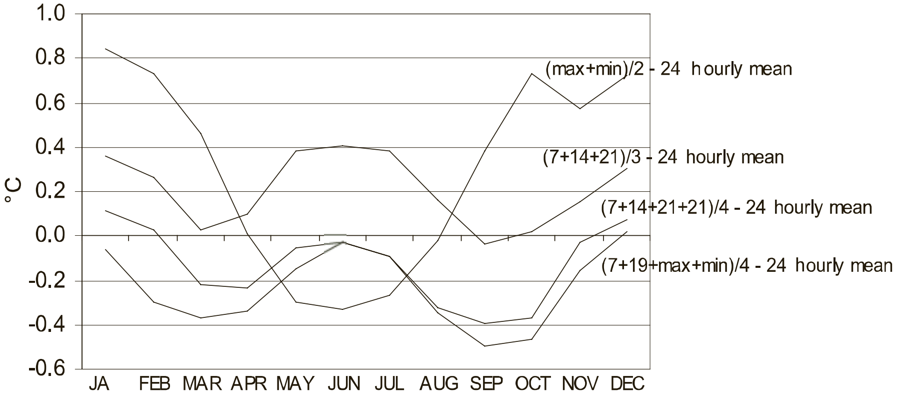


Figure 1.5. *Evolution through the year of the difference between various ways of calculating daily mean temperature and 24-hourly observations average for the inner-alpine station Puchberg in Austria, 1987-1996. Data source: Central Institute for Meteorology and Geodynamics, Vienna, Austria (from [AGU 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 (max+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 inhomogenity 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.

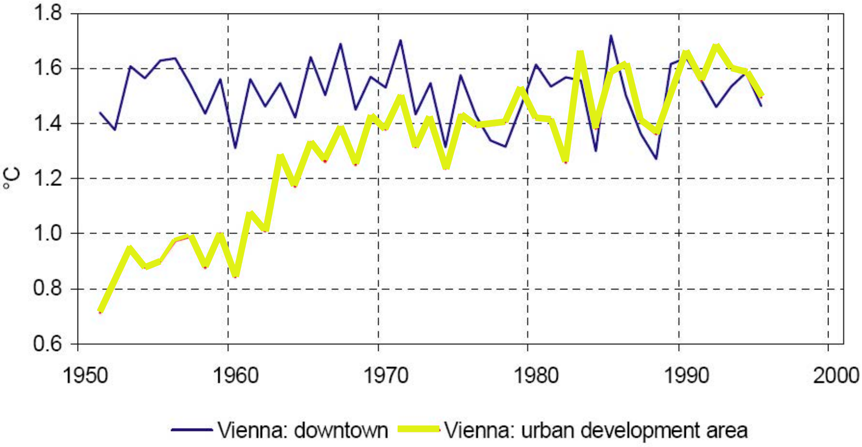


Figure 1.6. *Time series of annual mean urban temperature excess (relative to rural mean 1951 to 1995) based on height-reduced temperature records. The station in the densely built-up area shows a stable temperature excess against the rural surroundings, whereas the trend of temperature excess at the station in the urban development area is 0.18°C per decade. Data source [BOH 98]*

Another inhomogenity 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 inhomogenity 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.

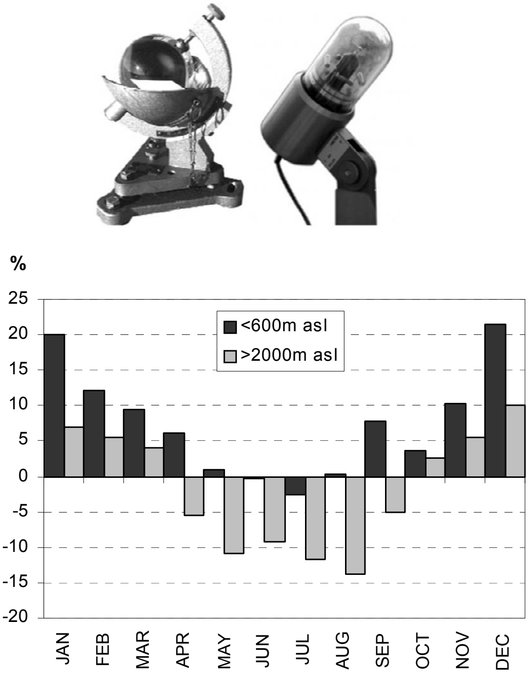


Figure 1.7. *Top: two types of instruments to record sunshine duration used in the Austrian meteorological network: Campbell-Stokes sunshine autograph and Haenni Solar system of automatic weather stations. Down: Consequences: Mean annual course of the breaks in Austrian sunshine series due a change from 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]*

	air pressure	temperature	precipitation	sunshine	cloudiness	all	
no. of series	72	131	192	55	66	516	series
available data	10215	19312	26063	7886	7669	71145	years
mean length of series	141.9	147.4	135.7	88.8	119.5	137.9	years
detected breaks	256	711	966	366	234	2533	breaks
mean homogeneous sub-interval	31.1	22.9	22.7	11.6	26.3	23.4	years
detected real outliers	638	4175	529				outliers
filled gaps	4217	12392	14927	2011	3513	37060	months
mean gap rate	3.4	5.3	4.8	2.1	3.8	4.3	%

Table 1.3. *Result of a homogenity study from monthly multiple climate series from the Greater Alpine Region (from [AUE 07])*

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.