



Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

D5.8 Assessment of the available station density for the gridding and daily data quality/homogeneity

Due date of deliverable: February 2006

Actual submission date: 3 March, 2006

Start date of project: 1 September 2004

Duration: 60 Months

KNMI, The Netherlands

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	YES
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

**D5.8 Assessment of the available station density for the
gridding and daily data quality/homogeneity**

**WP5.1 Development of daily high-resolution gridded
observational data sets for Europe**

March 2006

Royal Netherlands Meteorological Institute (KNMI)

Lisette Klok

Albert Klein Tank

Federal Office of Meteorology and Climatology (MeteoSwiss)

Michael Begert

Evelyn Zenklusen

Overview

1	Introduction.....	5
2	Data collection	7
2.1	Introduction	7
2.2	Data providers	7
2.3	Spatial coverage.....	11
2.4	Temporal coverage.....	13
2.5	Database, blending and updating	14
3	Quality control	15
3.1	Introduction	15
3.2	Quality control tests	15
3.2.1	Precipitation.....	15
3.2.2	Temperature	16
3.2.3	Air pressure	16
3.2.4	Snow depth	17
3.3	Quality control results	17
4	Homogeneity	19
4.1	Introduction	19
4.2	Absolute testing	20
4.2.1	The AT method.....	20
4.2.2	Results	21
4.3	Relative testing.....	23
4.3.1	The RT method.....	23
4.3.2	Results	26
4.4	Assessment of homogeneity	30
5	Summary and Conclusions	33
	References	35

1 Introduction

Observational data sets of climate variables are very important for climate research and can be used for climate model validation, scenario construction and impact assessments, and for the analysis of the past climate. Within RT5, “Independent comprehensive evaluation of the ENSEMBLES simulation-prediction system against observations/analyses”, observational data sets are developed for the evaluation of regional climate models (RCMs) that will focus on modeling changes in extreme events in Europe. Such observational data sets should preferably have a daily time resolution, be based on a dense network of quality-controlled station data and be available in a gridded format. The aim of WP5.1 is therefore to build high-resolution gridded data sets covering the standard RCM domain as defined in RT3 (see Figure 1), which is Europe and the countries bordering the Mediterranean Ocean. The data sets will include five climate variables: minimum and maximum temperature, precipitation, snow depth and air pressure.



Figure 1: Domain for gridded data sets; similar to the RCM domain as defined in RT3.

The gridded data sets will have a daily time resolution and will date back as far as data availability allows. However, the focus is the time span 1960-2004. For this period the regional climate models will simulate the present climate conditions. The period is overlapping the ECMWF and NCEP reanalysis. As such, the gridded data sets can serve as an alternative for using reanalyses for model evaluation. The spatial resolution of the gridded data sets will depend on the station density and data quality, but the aim is 25 km for the greater part of Europe and higher for some specific areas. In 2007, the data sets will become available for other WPs and RTs in ENSEMBLES and they will be accompanied by estimates of data uncertainty. The data will also be incorporated in KNMI's Climate Explorer (climexp.knmi.nl) for easy climate model comparisons. More and up to date

information about the gridded data sets can be found on WP 5.1's website via <http://www.ensembles-eu.org/>.

Although Europe has a long history of routine meteorological observations, high-resolution gridded daily data sets have not yet been developed plenty. The reason is that most station series are difficult to access. Especially high-resolution daily time series are often not available because of user restrictions at the National Meteorological Services or due to the amount of time and effort needed for the climate services to extract a full set of variables from their databases. Another reason is that some data have not yet been digitized. In fact, high-resolution gridded daily data sets covering Europe for a 45-year period have not yet been developed. The currently existing gridded data sets have either a coarser grid resolution, a shorter time span, or do not cover entire Europe. For instance, HadGHCND (Caesar *et al.*, 2005) is a global gridded daily data set based upon near-surface maximum and minimum temperature observations. It spans the years 1946 to 2000 but is available on a 2.5° latitude by 3.75° longitude grid only. JRC in Ispra, Italy houses the MARS-STAT Data Base containing interpolated meteorological data for Europe onto a 50 km grid, but only from 1975 up to present. The Alpine precipitation gridded analyses (Frei and Schär, 1998) are based on 6700 daily precipitation series and cover the period 1966-1995. The spatial resolution of this data set is 25 km and the region encompasses all Alpine countries.

This report describes the collected station series for the gridding project of ENSEMBLES WP 5.1 as well as their quality and homogeneity. A companion ENSEMBLES WP 5.1 deliverable (D5.9) will report on the analysis of possible gridding methods for the project. Chapter 2 gives an overview of the collected digitized daily data series, and shows the spatial and temporal coverage of the data network. The collection of data was primarily carried out by KNMI, who also hosts the European Climate Assessment and Data set (ECA&D). ECA&D served as the starting point for the ENSEMBLES data set and its database infra structure was used for ENSEMBLES as well (Section 2.5). All meteorological data were quality controlled and analyzed to avoid that non-climatic changes affect later analyses. The quality control procedures and their results are presented in Chapter 3. MeteoSwiss and KNMI performed separate homogeneity tests on the series. These tests and their results are described in Chapter 4. Finally, summary and conclusions on the available station density and its quality are given in Chapter 5.

Note that at the time that this report was written, not all data delivered for the project were implemented in the database yet. Especially updates for the ECA&D stations and data from 78 UK stations were still waiting for implementation. Also, we were still expecting new data, and the original series of the MARS-STAT Data Base at JRC in Ispra (243 maximum and minimum temperature series and 215 precipitation series), snow depth data from the INTAS project and additional station series from some ECA&D partners were in the pipeline. Therefore, this report discusses the status of the database at January 2006. It is expected that in summer 2006, the most substantial data collection will be finalized. This is well in time for the actual gridding process, which is planned for the next phase of WP 5.1.

2 Data collection

2.1 Introduction

At the start of this project (February 2005), only data from the European Climate Assessment and Data set (ECA&D) were at our disposal. At that time, this data set included about 250 stations, which is insufficient for the purpose of high-resolution gridding. The preferred station density for high-resolution (25 km) gridding was estimated at 1 station per 2,500 km². Since Europe's surface area is approximately 10,000,000 km², the intended number of stations to collect was 4000. We therefore gained additional station series from earlier EU projects and other research projects and requested various National Meteorological Services directly to send us more station data. Contacting the services directly was carried out in collaboration with ECA&D. These efforts resulted in an increase in the number of stations from the original 250 to 1670 stations at present. Section 2.2 lists the providers of these station series and Section 2.3 presents the station network. Section 2.4 describes the temporal coverage of the time series and the last section (Section 2.5) briefly explains the database structure in which the data are stored.

The collected climate variables are: daily minimum (tn) and maximum (tx) temperature, daily precipitation amount (rr), daily mean snow depth (sd) and daily mean air pressure corrected to sea level (pp). Daily mean temperature (tm) was often provided along with the maximum and minimum series or calculated by averaging minimum and maximum temperature. Although tm was not selected as a variable for gridding, we incorporate it in this quality assessment. Air pressure was not always corrected to sea level before delivery (for instance series from Germany, Croatia and Switzerland). In these cases, we performed this correction following Slonosky *et al.* (2001). Air pressure series measured at stations above 1000 m a.s.l. were not corrected but flagged as suspicious (Section 3.2.3)

2.2 Data providers

The daily time series have been gained either directly from the National Meteorological Services, or indirectly from various projects, such as STARDEX and EMULATE (two European FP5 projects), or from existing data sets, such as the GCOS Surface Network, GHCND and MAP. Details on these projects and data sets are given below. Table 1 lists the data providers that have contributed to these projects or from whom we received data directly, often for the joint purpose of ECA&D and ENSEMBLES. Some series came with restrictions for use. All data providers should be acknowledged in the papers in which the data series are used.

Examples of data sets that were unfortunately not or not yet available for our project due to legal aspects are the daily precipitation data collection from the Global Precipitation Climatology Centre (GPCC) and the original daily series in the Alpine precipitation gridded analyses (Frei and Schär, 1998).

ECA&D (<http://eca.knmi.nl/>)

ECA&D is a project of EUMETNET/ECSN. The project's objective is to gather daily series, quality control these series, analyse extremes and disseminate both daily data and the analysis results. New versions of the daily data set are issued at regular intervals. The network of participants covers 41 countries in Europe and some countries in the Middle East and North Africa. In March 2005, all ECA&D participants were asked to send updates and new series especially for the purpose of the ENSEMBLES gridding project. We are very pleased that many countries responded positively to this request, leading to a growth of the number of ECA&D stations from 250 (January 2005) to 550 (January 2006). Since not all ECA&D partners have responded yet, more station series are expected.

EMULATE (<http://www.cru.uea.ac.uk/cru/projects/emulate/>)

EMULATE (European and North Atlantic daily to MULTidecadal climATE variability) is a European 5th framework programme. This project aimed at extending the availability of daily historic records of air pressure measurements over the Atlantic and Europe for the period 1850 to present. Atmospheric circulation patterns have been derived from these records and relationships between the circulation, sea-surface temperatures and surface temperature and precipitation patterns across Europe have been evaluated. Air pressure series were downloaded from the EMULATE website and mean, minimum, and maximum air temperature, and precipitation were sent by the project co-ordinator after each EMULATE data provider had given us the permission to use the EMULATE series for ENSEMBLES. Due to overlap with the ECA&D data set a total of about 200 EMULATE stations were added to the database. Unfortunately, not all EMULATE stations contain data for all five gridding variables.

STARDEX (<http://www.cru.uea.ac.uk/cru/projects/stardex/>)

STARDEX (Statistical and Regional dynamical Downscaling of Extremes for European regions) is also a European 5th framework programme. STARDEX compared and evaluated statistical, dynamical and statistical-dynamical downscaling methods for the construction of scenarios of extremes. Daily mean, minimum, maximum temperature and precipitation series of 250 non-overlapping stations were added to the database after each STARDEX data provider had given us permission to use their series for ENSEMBLES purposes. All STARDEX series cover the period 1958-2000.

GCOS Surface Network (<http://www.wmo.ch/web/gcos/gcoshome.html>)

The Global Climate Observing System (GCOS) was established in 1992 to ensure that the observations and information needed to address climate-related issues are obtained and made available. It is sponsored by the World Meteorological Organization, the Intergovernmental Oceanographic Commission of UNESCO, the United Nations Environment Programme and the International Council for Science. To monitor the world's climate, GCOS selected the best stations in each region of the world to build a global surface reference climatological station network. Presently, this global network contains daily series from 953 stations in 126 countries. We downloaded series of 49 European stations from the GCOS website, which were not overlapping with ECA&D, and implemented them in the database. These series contain mean, maximum, minimum temperature, precipitation, and sea level pressure.

GHCND (<http://www.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html>)

The Global Historical Climatology Network – Daily (GHCND) is a compilation of global daily data developed by the National Climatic Data Center (NCDC). This data set is developed for researchers, weather related issues, agriculture and policy makers. 643 stations from this data set (DSI-9300) were taken for incorporation in the database. These stations mainly reside in the former Soviet-Union countries and include precipitation and temperature series.

MAP (<http://www.map.meteoswiss.ch>)

The Mesoscale Alpine Programme (MAP) is an international project that aims at studying the atmospheric and hydrological processes over the Alps to improve the current forecasting capabilities. This project supplied us with mainly precipitation series of 59 Italian stations.

SYNOP and CLIMAT (<http://www.ecmwf.int/services/archive/>)

Additionally, data were retrieved from the synoptical messages. These SYNOP and CLIMAT data were only used for updating, extending and filling gaps in existing station series (see Section 2.5). The series were retrieved from the ECMWF MARS-archive.

Table 1: List of data providers.

ECA&D	
Hydrometeorological Institute	ALBANIA
Office National de la Météorologie	ALGERIA
Hydrometeorology and Environmental Agency	ARMENIA
Central Institute for Meteorology and Geodynamics	AUSTRIA
Koninklijk Meteorologisch Instituut	BELGIUM
Federal Meteorological Institute of Bosnia and Herzegovina	BOSNIA AND HERZEGOVINA
National Institute of Meteorology and Hydrology	BULGARIA
Meteorological and Hydrological Service of Croatia	CROATIA
Meteorological Service of Cyprus	CYPRUS
Czech Hydrometeorological Institute	CZECH REPUBLIC
Danish Meteorological Institute	DENMARK
Estonian Meteorological and Hydrological Institute	ESTONIA
Finnish Meteorological Institute	FINLAND
Météo-France	FRANCE
Deutscher Wetterdienst	GERMANY
Hellenic National Meteorological Service	GREECE
Hungarian Meteorological Service	HUNGARY
Icelandic Meteorological Office	ICELAND
Met Éireann	IRELAND
National University of Ireland, Galway	IRELAND
Israel Meteorological Service	ISRAEL
Università degli Studi di Milano	ITALY
Servizio Meteorologico dell' Aeronautica	ITALY
Latvian Hydrometeorological Agency	LATVIA
Institute of Geology and Geography	LITHUANIA
Service Météorologique du Luxembourg	LUXEMBOURG
Centre de Recherche Public Gabriel Lippmann	LUXEMBOURG
Royal Netherlands Meteorological Institute	NETHERLANDS
Det Norsk Meteorologiske Institutt	NORWAY
Institute of Meteorology and Water Management	POLAND

Instituto de Meteorologia	PORTUGAL
Universidade do Algarve	PORTUGAL
Republic Hydrometeorological Institute	REPUBLIC OF MACEDONIA
National Institute of Meteorology and Hydrology	ROMANIA
Russian Federal Service for Hydrometeorology and Environmental Monitoring	RUSSIAN FEDERATION
Republic Hydrometeorological Institute of Serbia	SERBIA AND MONTENEGRO
Slovak Hydrometeorological Institute	SLOVAKIA
Environmental Agency of the Republic of Slovenia	SLOVENIA
Instituto Nacional de Meteorologia	SPAIN
Swedish Meteorological and Hydrological Institute	SWEDEN
Stockholm University	SWEDEN
MeteoSwiss	SWITZERLAND
Devlet Meteoroloji Genel Mudurluou	TURKEY
Central Geophysical Observatory	UKRAINE
Met Office	UNITED KINGDOM
EMULATE	
Central Institute for Meteorology and Geodynamics	AUSTRIA
Meteorological and Hydrological Service of Croatia	CROATIA
Czech Hydrometeorological Institute	CZECH REPUBLIC
Finnish Meteorological Institute	FINLAND
Potsdam Institute for Climate Impact Research	GERMANY
Universität Augsburg, Institut für Geographie	GERMANY
National Observatory of Athens	GREECE
Icelandic Meteorological Office	ICELAND
Università di Milano, Istituto di Fisica Generale Applicata	ITALY
Instituto Nazionale di Astrofisica, Osservatorio Astronomico di Palermo	ITALY
Royal Netherlands Meteorological Institute	NETHERLANDS
University of Barcelona, Department of Astronomy and Meteorology	SPAIN
Universitat Rovira I Virgili, Physical Geography	SPAIN
Swedish Meteorological and Hydrological Institute	SWEDEN
Uppsala University, Department of Earth Sciences, Air and water Science-Meteorology	SWEDEN
MeteoSwiss	SWITZERLAND
Met Office	UNITED KINGDOM
Armagh Observatory	UNITED KINGDOM
GCOS Surface Network	
National Climate Data Center	USA
GHCND	
National Climate Data Center	USA
MAP	
Ufficio Idrografico Provincia Autonoma di Trento	ITALY
Servizio Meteorologico dell' Aeronautica	ITALY
STARDEX	
Central Institute for Meteorology and Geodynamics	AUSTRIA
Czech Hydrometeorological Institute	CZECH REPUBLIC
Danish Meteorological Institute	DENMARK
Finnish Meteorological Institute	FINLAND
Meteo France	FRANCE
Deutscher Wetterdienst	GERMANY
Department of Meteorology and Climatology, Aristotle University of Thessaloniki	GREECE
Hungarian Meteorological Service	HUNGARY
Servizio Meteorologico Regional, ARPA-Emilia Romagna	ITALY

Royal Netherlands Meteorological Institute	NETHERLANDS
Det Norsk Meteorologiske Institutt	NORWAY
Instituto de Meteorologia	PORTUGAL
Russian Federal Service for Hydrometeorology and Environmental Monitoring	RUSSIAN FEDERATION
Instituto Nacional de Meteorología	SPAIN
Swedish Meteorological and Hydrological Institute	SWEDEN
MeteoSwiss	SWITZERLAND
University of East Anglia, Climate Research Unit	UNITED KINGDOM
SYNOP	
ECMWF	United Kingdom

2.3 Spatial coverage

Figure 2 depicts the present station density for the ENSEMBLES data set. It is still sparse in some areas: southern Italy, Poland, Slovakia, Hungary, the Balkan countries, Turkey and North Africa. The station network in the former Soviet Union is relatively dense and is mainly based on stations from the GHCND data set. The densest station network is found in the Netherlands where each station represents an average area of 300 km². Here, the required density for gridding is reached (see Section 2.1).

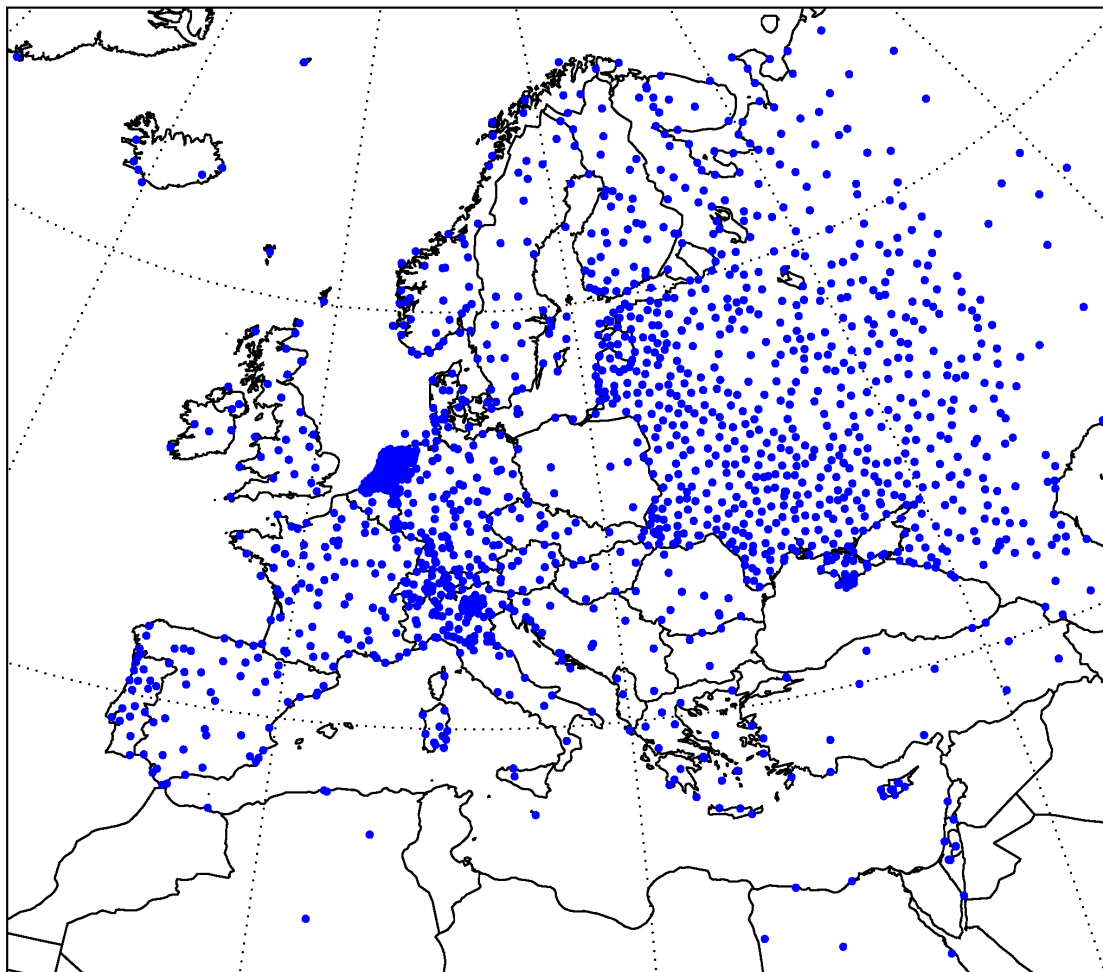


Figure 2: Station network for ENSEMBLES (status: January 2006).

Although, the number of collected stations in the database is 1670 at present, not all stations contain series for all variables. Most of the collected stations contain precipitation series, but from only a few stations, we received snow depth data. Currently, the database stores 1456 precipitation series, 966 maximum temperature series, 970 minimum temperature series, 639 mean temperature series, 239 air pressure series and 134 snow depth series. Figure 3 shows the spatial distribution of the series for each of these variables. It is clear that only a few countries have contributed to the snow depth data set so far.

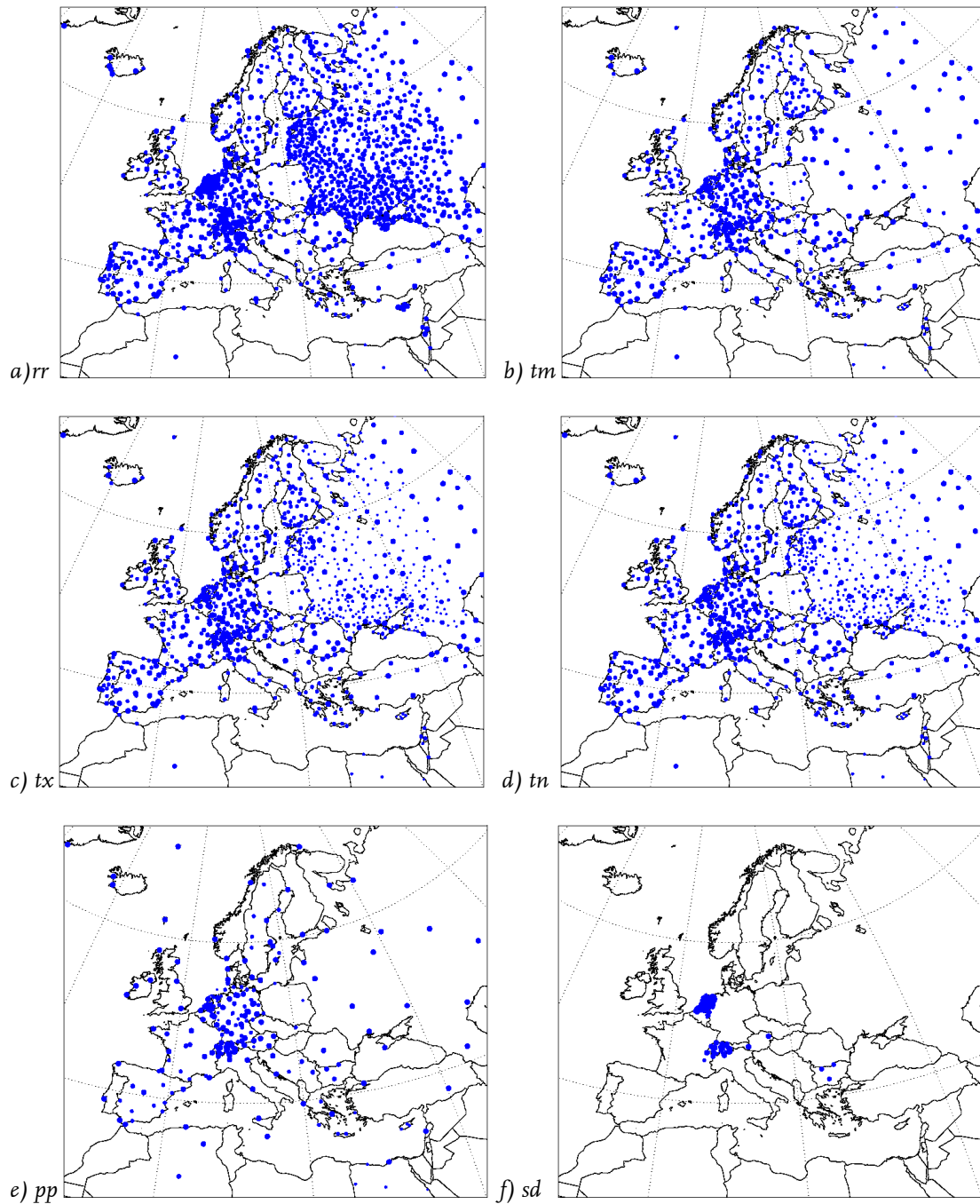


Figure 3: Stations with daily time series for precipitation (a), mean temperature (b), maximum temperature (c), minimum temperature (d), air pressure (e) and snow depth (f). Large dots represent series that start before 1960 and end after 2004. Small dots indicate shorter series.

2.4 Temporal coverage

Unfortunately, not all of the depicted stations in Figure 3 contain series that cover the entire period 1960-2004. The series that cover this period are indicated with larger symbols: 92% for precipitation, 90% for mean temperature, 62% for maximum temperature, 62% for minimum temperature, 80% for air pressure and 90% for snow depth.

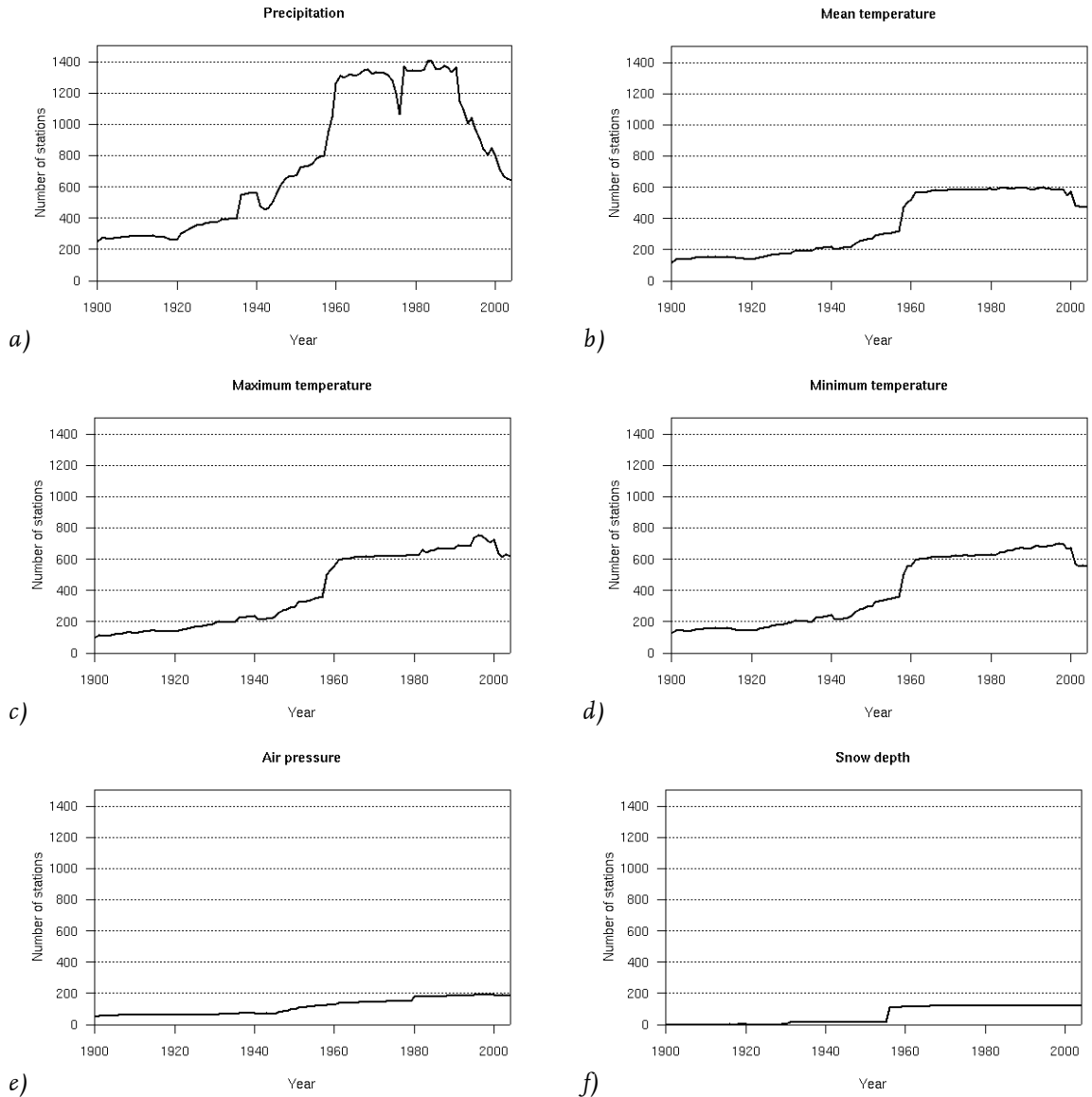


Figure 4: Number of stations with for each year the number of stations for precipitation (a), mean temperature (b), maximum temperature (c), minimum temperature (d), air pressure (e), and snow depth (f). A station was added year when at least 292 days of that year contain valid data.

Figure 4 shows the number of series for each variable as a function of time. Here, a series was assigned to a specific year when at least 80% of the year (=292 days) contains valid data (*i.e.* data that were flagged as useful; see Section 3.1). The plots show that the number of precipitation series in the database is by far the greatest.

For all variables, the best data coverage is achieved between 1960 and 2000. Data from the most recent years are often missing, but we do use SYNOP and CLIMAT data to update the series (see Section 2.5). The strong decline in the precipitation series over the last 15 years is mainly caused by the GHCND precipitation series from the former Soviet Union, which stop in the early nineties.

2.5 Database, blending and updating

All data are stored in the operational system of ECA&D. This is a relational data model, which makes use of a MySQL database. It consists of two types of tables: core tables that contain raw data and derived tables that contain data that are derived from the raw data and regularly renewed. For instance, metadata and raw station series received from the data providers are stored in core tables, while results of the homogeneity tests and blended data (see below) are stored in derived tables. In addition to data gathering and archiving, the system also performs a quality control (see Section 3) and appoints permission rights to each data series. The latter implies that for each series it is known whether it is public to everybody, for use in ECA&D, or for use in ENSEMBLES only. The public data are disseminated on the ECA&D website (<http://eca.knmi.nl>). The affiliation of the data provider of each series is also stored in tables of the database.

Countries have been measuring daily variables using different methods and formulae, even though WMO intentions of standardization. Often the time interval for observing minimum and maximum temperatures differs or the time interval for accumulating 24h rainfall differs. When, for instance, at one station the daily precipitation sum is measured as the amount between the morning of the previous day (*e.g.* at 6 am) and the morning of today, and at another station between the morning of today and the morning of the next day, the series will not match. Consequently, problems are expected when these data are interpolated onto grids. In order to deal with this, each data series is identified with an individual element identification in the database that gives information about the measurement procedure. Unfortunately, from most series, the measurement interval is still unknown up till now.

Since not all series are complete or contain data for the most recent years, the database system uses observations from SYNOP and CLIMAT messages to extend and fill gaps in the time series. These data messages are distributed in near real time over the global telecommunication system (GTS) and here we use the database of ECMWF as the source. Besides SYNOP and CLIMAT messages, observations from nearby stations are used for gap filling and updating, provided that these stations are within 25 km distance and 50 m height difference. This process is called blending of data. Data that are flagged as suspicious during the quality control procedure are also filled with useful data by the procedure of blending, provided that SYNOP, CLIMAT or nearby station data are available. In any case, priority is given to series of nearby stations in the database rather than to SYNOP or CLIMAT data for the process of blending. Since SYNOP and CLIMAT data are not verified thoroughly before transmittance through the GTS, the data quality is lower, whereas station series gained from meteorological services normally undergo a more rigorous check before distribution.

3 Quality control

3.1 Introduction

All series are quality checked following the procedures of ECA&D and flagged accordingly. No corrections or adjustments were made. The flags can be used to make selections of the series that are used for the gridding. There are three types of data flags: (0) data is useful, (1) data is suspicious, *i.e.* it did not pass a quality control test, and (9) data is missing. The quality control tests are listed in the sections below and include tests that detect several types of errors: anomalous values, repetitive values, outlying values and incoherencies. All quality control tests are absolute, implying that series are not compared with respect to neighbouring station series to find outliers or suspect data.

Some data sets were quality controlled before delivery and some were in fact adjusted (for instance in EMULATE). Nevertheless, we performed our own quality control and did not adopt the quality control flags of other databases. However, if adjusted time series were available, we implemented these instead of the raw data.

3.2 Quality control tests

3.2.1 Precipitation

The quality control tests for daily precipitation (*rr*) are:

1. *rr* must be positive or zero
2. *rr* must be less than 299.9 mm
3. *rr* must not be repetitive for 10 days when *rr* is larger than 1.0 mm
4. *rr* must not be repetitive for 5 days when *rr* is larger than 5.0 mm
5. a site-specific outlier test to detect incorrect long dry periods*

*We noticed that incorrect long dry periods sometimes occurred at the beginning or ending of a time series and irregularly in the middle of a series. This often happens when precipitation is not recorded and zero values are assigned to daily precipitation instead of missing values. In order to detect these erroneous dry periods, we performed an outlier test. We first calculated for each station and year the largest number of consecutive dry days. Since the distribution of this series is likely not Gaussian and may contain outliers, we used the biweight mean and biweight standard deviation (Lanzante, 1996) of this series to calculate z-scores for each dry period. We then inspected 250 series manually and concluded that all erroneous dry periods exceed a z-score of 1.4. As a result, a z-score of 1.4 was adopted to flag long droughts in the precipitation series as suspicious.

Apart from the above automated tests for precipitation, *rr* was checked by day of week. It turned out that at none of the precipitation stations it was general practice to not observe on Sundays, during weekends, or on specific weekdays and assign a value of zero precipitation to this day instead.

3.2.2 Temperature

The quality control tests related to maximum daily temperature (tx) are:

1. tx must exceed -89.9 °C
2. tx must be less than 60.0 °C
3. tx must exceed or equal tn and tm
4. tx must not be repetitive for 5 days
5. a site-specific outlier test*

The quality control tests related to minimum daily temperature (tn) are:

1. tn must exceed -89.9 °C
2. tn must be less than 60.0 °C
3. tn must be less or equal to daily tx and tm
4. tn must not be repetitive for 5 days
5. a site-specific outlier test*

The quality control tests related to mean daily temperature (tm) are:

1. tm must exceed -89.9 °C
2. tm must be less than 60.0 °C
3. tm must be less or equal to daily tx and exceed or equal tn
4. tm must not be repetitive for 5 days
5. a site-specific outlier test*

*The outlier test adopted here is defined by a temperature range within which temperature values are allowed. For each day in a data record, a z-score is calculated from the climatological mean temperature for the calendar day and standard deviation using a time-window of five days. Days for which the z-score exceeded 5 were flagged as suspicious. Others have used the biweight mean and standard deviation, no time-window or a maximum z-score of 3 or 4 to identify the outliers. Nevertheless, we concluded that the method as described above performs best. It flags most of the erroneous outliers and its false alarm rate is relatively low. The latter is important in this project since its focus is on climate extremes.

Number three of each of the above temperature tests detects so-called incoherencies (e.g. tn should be small or equal to tx). However, if tx, tn and tm are not measured over the same time interval, data can be incorrectly flagged as suspicious. Therefore, these incoherency tests likely overestimate the number of suspect values slightly.

3.2.3 Air pressure

The quality control tests related to air pressure (pp) are:

1. pp must exceed 600.1 hPa
2. pp must be less than 1080.1 hPa
3. pp must not be repetitive for 5 days

Furthermore, station series that were not corrected to sea level because of a station elevation above 1000 m a.s.l. were flagged as suspicious.

3.2.4 Snow depth

The quality control tests related to snow depth (sd) are:

1. sd must exceed or equal 0
2. sd must be less than 15 m

3.3 Quality control results

We appointed flags to the data of all 1456 precipitation series, 970 minimum temperature series, 966 maximum temperature series, 639 mean temperature series, 239 air pressure series and 134 snow depth series in the database according to the above quality control tests. Table 2 gives the percentage of data values flagged for each variable. The percentage of missing values exceeds the percentage of suspicious values by one order of magnitude. Air pressure contains the largest percentage of suspicious data. This is mainly due to 16 air pressure series that were labeled as suspicious because of their station elevation being above 1000 m a.s.l.

Table 2: Percentage of data values flagged as useful (0), suspicious (1) and missing (9) in the database.

Variable	QC=0	QC=1	QC=9
tx	85.1	0.2	14.7
tn	86.4	0.2	13.4
tm	90.4	0.3	9.3
rr	84.9	0.0	15.1
pp	65.5	3.8	30.7
sd	85.4	0.1	14.5

4 Homogeneity

4.1 Introduction

Long climatological time series often contain variations due to non-climatic factors, such as site-relocations, changes in instrumentation or observing practices (e.g. Böhm *et al.*, 2001; Tuomenvirta, 2001; Auer *et al.*, 2005). Such changes can lead to shifts in mean, in variance or spurious trends. As these inhomogeneities can distort or even hide the true climatic signal, the development of homogeneous time series is a key issue for climate change studies.

Numerous methods are in use to evaluate the homogeneity of climatic time series. A comprehensive overview is given in Peterson *et al.* (1998) and WMO guidelines are provided in Aguilar *et al.* (2003). In general homogeneity tests are applied relatively, *i.e.* data are tested with respect to homogeneous neighbouring stations. A combination of statistical methods and analysis of station history is considered to be the most effective approach. While the detection of inhomogeneities is based on monthly or annual values, the calculation of adjustments requires different methods for lower and higher time resolution. Till today only a few methods exist to homogenize daily data (e.g. Della-Marta and Wanner, 2006).

The homogeneity assessment of the daily data set within ENSEMBLES WP 5.1 aims to provide the necessary uncertainty estimates of the data that is input to the gridding process. No adjustments to the daily values have been attempted. The focus is on the establishment of a procedure that is able to detect non-climatic changes in series of a multivariate data set. The large number of series to be tested requires a fully automated procedure since metadata information is generally not available.

Several different possible procedures were evaluated: the quality control procedure VERAQC adaptable for homogenization purposes (Steinacker *et al.*, 2000), the HOCLIS system developed at ZAMG (Auer *et al.*, 2001), the homogenization procedure designed for ECA&D (Wijngaard *et al.*, 2003), THOMAS (Tool for Homogenization of monthly data series) developed at MeteoSwiss (Begert *et al.*, 2005) and a recently presented method of homogenizing daily temperature series (Della-Marta and Wanner, 2006). The evaluation has included estimates of the required amount of manual input as well. Finally the evaluation suggested that a combination of an enhanced version of the absolute homogeneity analysis developed within the framework of ECA&D (Wijngaard *et al.*, 2003) together with a new automated relative method based on the quality control method of Steinacker *et al.* (2000) (VERAQC) optimally meets the requirements.

The use of both methods combines the more powerful approach of relative testing with the ability of absolute tests to deal with simultaneous changes in the series. It also includes a fallback strategy if the performance of one method turns out to be not satisfying in future.

The following sections describe the two methods in detail, discuss advantages and disadvantages of the different approaches and give an overview of their results.

Finally, a combination of the results is proposed to form a basis to select series or parts of series appropriate for the gridding purposes (Section 4.4).

4.2 Absolute testing

4.2.1 The AT method

Within ECA&D, the homogeneity procedure of Wijngaard *et al.* (2003) is used to evaluate the homogeneity of the precipitation and temperature series. For the purpose of ENSEMBLES, Wijngaard's method was adopted and modified. Hereafter, this method will be called the Absolute Test (AT). We expanded the test to two additional variables: air pressure and snow depth, and modified it to detect multiple inhomogeneities in a record instead of only one.

The AT procedure consists of several steps. Initially, four homogeneity tests are applied to five annual testing variables: two testing variables for temperature and one variable for precipitation, air pressure and snow depth each. The testing variables are:

1. the annual mean of the diurnal temperature range (= maximum temperature - minimum temperature)
2. the annual mean of the absolute day-to-day differences of the diurnal temperature range
3. the wet day count (using a threshold 1 mm)
4. the annual mean air pressure
5. the snow depth day count (using a threshold of 0 cm)

Testing variables are exclusively calculated for years that contain at least 362 days of valid data. Data flagged as suspicious (Section 3.1) were omitted. A disadvantage of the testing variables is that minimum, mean and maximum temperatures are not tested separately. Instead, temperature results depend on the first two testing variables only. When a break is found in the series of these variables, it is assumed to be present in all temperature variables, whereas its origin could for instance be in the minimum temperature series only.

The four homogeneity tests applied to the testing variables are:

1. the standard normal homogeneity test (Alexandersson, 1986)
2. the Buishand range test (Buishand, 1982)
3. the Pettitt test (Pettitt, 1979)
4. the Von Neumann ratio test (Von Neumann, 1941).

The first three homogeneity tests are capable of detecting the location of the break year. By contrast, the Von Neumann ratio test is a non-location-specific test. In principal, the starting and final date of a series define the period to which the homogeneity tests are applied. However, some series contain missing data. Therefore, we introduced a threshold: at least 80% of the years of the period to which the tests are applied should contain valid data. If this condition is not fulfilled, a search for the longest possible period in the series with 80% of valid data is carried out and this period is tested accordingly. Series shorter than 10 years are omitted.

After that, the test results are grouped into three classes: useful, doubtful, and suspect. This classification depends on the number of tests rejecting the null

hypothesis (see Table 3). Useful series are thus defined as series for which only one test or none of the tests reject the null hypothesis. These series are regarded as sufficiently homogeneous for trend and variability analysis, and also for gridding. Note that two variables are used for testing the temperature series. Accordingly, the least favorable result out of the two variables is used for each statistical test when grouping the results.

Table 3: Homogeneity classes

Classification	Number of tests rejecting the null hypothesis at the 1% level
useful	0 or 1
doubtful	2
suspect	3 or 4

Since series can exhibit multiple inhomogeneities, AT searches for all useful segments in the data records. The break years of the statistical tests that reject the null hypothesis are used to divide doubtful or suspect series into segments. The most recent break year of the tests that reject the null hypothesis plus one year is used as the starting year for the most recent segment. Likewise, the earliest break year minus one is used as the end year for the segment preceding the break. These two segments are then tested again using the same procedure as described above. The procedure stops if a segment is shorter than 10 years. Finally, a set of periods is created for each data record, each period classified as useful, doubtful, suspect or unknown (when the period length is less than 10 years or less than 80% of the years contain invalid data).

4.2.2 Results

Table 4 presents for the various climate variables, the number of useful segments starting after 1960 subdivided into four length classes. Snow depth and also precipitation are variables with relatively many useful series, 90% and 76% of all series respectively. This is not surprising as inhomogeneities in snow depth and precipitation series are more difficult to detect than in temperature or air pressure series due to the higher year-to-year variability. Fortunately, the majority of the useful periods exceeds 40 years. Still, there are also many short useful periods.

Table 4: Number of stations with homogeneous periods, classified as useful, subdivided into 4 length classes. Only the longest homogeneous period of each station was taken into account. The last row gives the percentage of useful series with regard to the total number of series in the database. *tt* is temperature, *rr* precipitation, *pp* air pressure and *sd* snow depth.

Period length	<i>tt</i>	<i>rr</i>	<i>pp</i>	<i>sd</i>
10-20	144	165	15	4
20-30	150	254	48	8
30-40	118	221	17	4
>40	193	486	102	105
TOTAL	605	1126	182	121
PERCENTAGE	63%*	77%	76%	90%

* calculated with reference the number of minimum temperature series in the database

Figure 5 depicts the stations with useful series. Useful series are found in each country that has provided data. Whereas the station density plot (Figure 1) show a high density in Eastern Europe and Russia, Figure 5 depicts that many of these series contain only short useful periods.

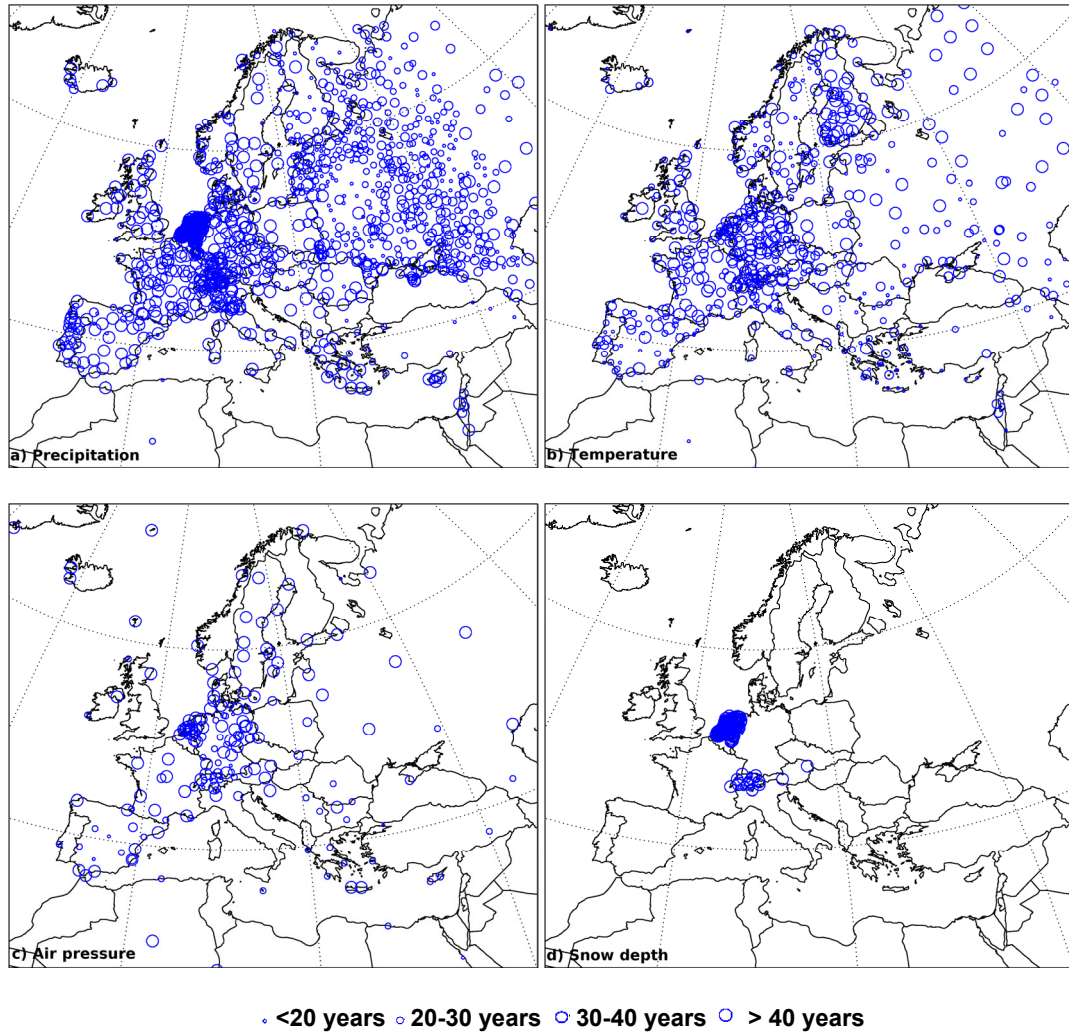


Figure 5: Location of homogeneous series classified as useful and subdivided into 4 classes for the length of the period for the variables precipitation (a), temperature (b), air pressure (c) and snow depth (d). Only the longest homogeneous period of each station was taken into account.

Table 5: Number of stations with homogeneous periods, classified as useful and doubtful, subdivided into 4 length classes. Only the longest homogeneous period of each station was taken into account. The last row gives the percentage of useful series with regard to the total number of series in the database.

Period length	tt	rr	pp	sd
10-20	131	158	13	4
20-30	125	246	45	6
30-40	110	221	16	4
>40	248	503	108	107
TOTAL	614	1128	182	121
PERCENTAGE	64%*	77%	76%	90%

* calculated with reference the number of minimum temperature series in the database

As the scores for homogeneous periods in Table 4 are not high considering the requirement of a dense station network needed for high-resolution gridding, we tested if the scores would increase if, in addition to useful, also doubtful series (see Table 3) are assumed to be homogeneous. The results are given in Table 5. The total number of homogeneous periods increases only slightly for temperature and precipitation when doubtful periods are additionally taken into account. For air pressure and snow depth, there is no difference. However, for all variables the number of homogeneous periods exceeding 40 years increases and the number of

homogeneous periods with a length of 10 to 20 years decreases or remains constant.

4.3 Relative testing

4.3.1 The RT method

“A climatological series is relatively homogeneous with respect to a synchronous series if the differences (or ratios) of pairs of homologous averages constitute a series of random numbers that satisfies the law of errors” (Conrad and Pollak, 1950). Following this idea, relative homogeneity testing is usually based on the comparison of the candidate series with a reference series composed of sufficiently correlated homogeneous series of neighbouring stations. Inhomogeneities in the candidate series are detected in the difference or ratio series (henceforth q-series) using suitable statistical methods. In short: relative homogeneity test procedures can be divided into two main steps: 1) the creation of the reference (q-)series, 2) the appliance of the statistical detection tests.

The procedure developed for relative homogeneity testing in the framework of ENSEMBLES (called RT henceforth) combines a fully automated method to build a q-series with a known statistical homogeneity test:

1. Vienna Enhanced Resolution Analysis Quality Control (VERAQC, Steinacker et al., 2000) is used to build the q-series and
2. Alexandersson’s standard normal homogeneity test (abbr.: SNHT, Alexandersson, 1986) is applied on the q-series to detect the inhomogeneities.

VERAQC was originally designed for quality control of meteorological measurements and runs on an operational basis at MeteoSwiss. The procedure is based on an objective spatial interpolation algorithm. For each candidate station, five surrounding stations are selected by a defined algorithm whereas their locations have to form an ideal pentagon in reference to the central candidate station. A surface with minimized curvature is fitted through the values of the five neighbouring stations using a thin plate spline algorithm. Finally, the difference between the measured value of the central station and the corresponding value on the fitted plane is calculated. An analysis of these differences in time (for a single station) and space (for one time step) is used for quality control purposes. A detailed description of the method is given in Steinacker et al. (2000). The results of its application to routine data are given in Häberli et al. (2004) and an extension to test the relative homogeneity of radiosonde time series in Häberli (2005).

In the context of homogenization the time series of the differences calculated by VERAQC (henceforth deviation series) for a certain station can be regarded as a q-series. Variations due to non-climatic factors appear as shifts or trends in the deviation series and can be detected by common statistical test methods, such as SNHT. SNHT is designed to find shifts in mean and returns the date of one possible break point in the series analyzed. The test is applied iteratively to each input series in order to search for multiple break points. Significant test results on the 95% confidence level serve to divide the deviation series into segments. Each

segment is then investigated separately and the iteration is repeated until the segment is determined as homogenous or falls below a minimum segment length. The optimal minimum segment length was found to be 5 years independent of the variable. It is a compromise between a preferably high hit rate and an acceptable false alarm rate of RT.

In order to enhance the power of the statistical test the variability of the deviation series has been reduced by running VERAQC on monthly values. The calculation of deviation series is reasonable for all variables investigated in ENSEMBLES but snow cover. Therefore, the homogeneity assessment of snow data within ENSEMBLES has to rely on AT only.

For temperature and air pressure, the deviation series can also be used to examine the dimension of a detected shift in the mean by a simple comparison of the deviations in the homogeneous segment before and after the break point. The mean difference indicates the order of magnitude of the shift. The significance of the difference is tested using Student's t-test. In general, significant results of the homogeneity tests will also turn out to be significant using the t-test. However, the mathematics of the homogeneity tests is in principle strictly applicable only to a single break in the input series. Therefore, the test results may be disturbed by multiple break points and have to be considered as indicators only.

For precipitation, the q-series should actually be formed as ratios between a test and a reference station. The deviation series of VERAQC can well be used to detect break points as the homogeneity test results compared to the findings for other variables will proof. But the derivation of adjustments is not feasible.

Two problems associated with the automatically generated reference series using the VERAQC method need to be discussed. First, the algorithm of VERAQC fails if the number of surrounding stations is insufficient. Five surrounding stations were needed for the calculation of the deviation series for a candidate station. Therefore stations along the edge of the region can not be evaluated. This so-called "edge problem", concerns about 10%-20% of the data series (depending on the variable). Second, the procedure is sensitive to changes in the station network in time (in particular for variables which have a high spatial variability, *e.g.* temperature in mountainous regions). On one hand a varying network density leads to incomplete deviation series if a necessary neighbouring station appears or disappears in time. On the other hand incomplete series can be the reason for false breaks being detected by RT because the appearance or disappearance of a neighbouring series can lead to abrupt changes in the deviation series similar to real breaks.

A straight-forward solution to these problems was the reduction of the data set to complete series only. Several other possibilities to reduce the influence of a varying station network have been considered including the use of anomalies according to a standard period. However, the focus on complete series was the only possible solution in the time frame of ENSEMBLES. Exceptions were allowed for missing periods of 2 years (10 years for precipitation). As a consequence a considerable number of series could not be tested (30%-50%, see Table 7).

To assess the performance of RT, comparisons to “manually” homogenized Swiss data series for the period 1961 to 1997 were made. The latter have been homogenized using THOMAS, a tool specially developed for homogenization of climate series (Begert *et al.*, 2003). Because THOMAS needs a substantial amount of manual input, it was not suitable for the homogeneity testing of a large data set like the one in ENSEMBLES.

Only complete series were used for the comparison: 30 rr, 32 tx, 30 tn, 32 tm and 25 pp series. Figure 6 reveals the number of break points per series detected by RT compared to the break points detected by THOMAS. In addition, the number of false alarms (significant break points of RT, no THOMAS indication) and missed shift inhomogeneities (significant break points of THOMAS, no RT indication) are shown. Break points detected within one year by both methods are referred to as hits. Note that the THOMAS procedure includes the full station histories and that the network density for THOMAS was 4 to 20 times higher compared to the data set of ENSEMBLES. A denser network leads to smaller variations in the q-series and homogeneity tests are able to detect smaller (and therefore usually more) inhomogeneities. Similar, a higher network density prevents poorly correlated neighbouring stations to be included in the reference series. The inclusion of such series increases the false alarm rate as they might belong to another climate region. In short, RT can not be expected to reach the performance of THOMAS, especially in a mountainous area like Switzerland.

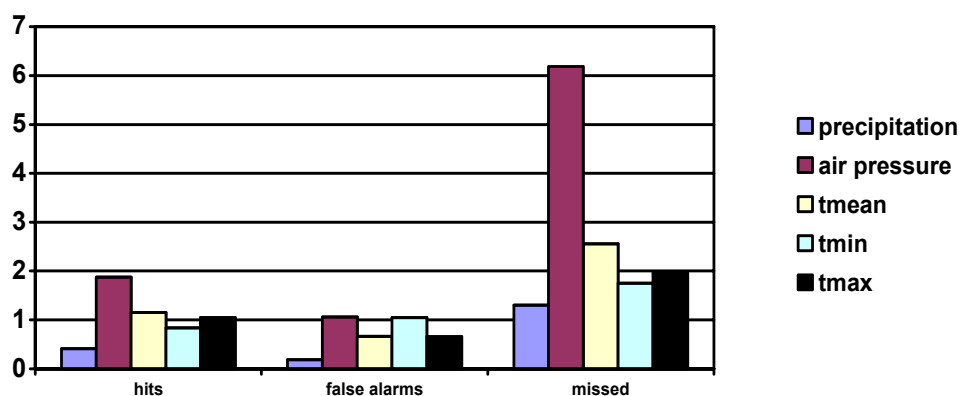


Figure 6: Sum of THOMAS inhomogeneities in the Swiss station series detected by RT, applied on monthly values for the period 1960-2000. The y-axis indicates the sum of break points per series.

Overall, the performance of RT in detecting the THOMAS breaks is rather poor. About half of the shifts detected by RT must be considered as false alarms (a result only weakly depending on the variable examined). The rate of missed shifts is also high, especially for air pressure. As mentioned above it must be kept in mind that this comparison is very strict. Missed break points can be expected due to the implemented minimum segment length and the sparser network density. For instance, the surprisingly high number of missed breaks in air pressure series can be explained by the fact that Swiss air pressure series contain a high number of inhomogeneities in the 1980's because of technical problems. RT is not able to detect breaks that follow each other within less than 5 years. The high number of missed breaks in temperature series on the other hand is due to small breaks that could not be detected by RT. The sparser network density compared to THOMAS causes small shifts to disappear into statistical noise. Concerning the false alarms, there is more than one reason for the rather high rate. The sparser network

density causing series from different climatological regions to be included in the reference series as well as inhomogeneities in series of neighbouring stations might be responsible.

Besides a comparison of “shift dates”, a comparison of “shift dimensions” found by RT and THOMAS gives an idea of the performance of the current configuration of RT. Table 6 compares the number of small ($< 0.5^{\circ}\text{C}$) and large ($\geq 0.5^{\circ}\text{C}$) shifts detected by RT and THOMAS in Swiss temperature series (mean, minimum and maximum series). In general most of the large shifts are characterized as large by RT whereas small shifts are only partly identified as small ones. False alarms and missed breaks in particular are generally small although a substantial number of large shifts is included.

Table 6. Left: number of large ($\geq 0.5^{\circ}\text{C}$) and small ($< 0.5^{\circ}\text{C}$) shifts detected by RT and THOMAS in Swiss temperature series (mean, minimum, maximum). Right: number of large ($\geq 0.5^{\circ}\text{C}$) and small ($< 0.5^{\circ}\text{C}$) false alarms and missed break points of RT (related to THOMAS findings) in Swiss temperature series (mean, minimum, maximum).

		hits (RT)			false alarm (RT)	missed (RT)
		large	small			
THOMAS	large	19	12	large	18	37
	small	7	10	small	40	119

Finally, note that the comparison of the two methods was carried out using series from stations located in the mountainous region of the Alps. As spatial homogeneity is one of the most important factors for successful relative homogeneity testing, a better performance can be expected for less mountainous regions, *i.e.* most parts of Europe. However, a relatively high number of false alarms and missed break points due to poorly correlated neighbouring stations or inhomogeneities in reference series is still expected. A comparison of the RT results with findings for series in other countries needs to be considered in a future study.

As described above, the *limitations of the proposed RT method* as a homogeneity test procedure summarize as follows:

1. The automated method RT is able to detect shift inhomogeneities and estimate their dimensions to a certain degree. The comparison with a sophisticated manual method (THOMAS) in the Alpine region reveals a hit rate between 23% and 35% and a false alarm rate between 31% and 56%, depending on the variable. Improved results can be expected in less mountainous regions, *i.e.* for most parts of Europe.
2. For a given test period only complete data series should be used.
3. Inhomogeneities closer than five years apart can not be detected.
4. Stations located at the boundary can not be tested (edge problem).

4.3.2 Results

The homogeneity of the ENSEMBLES data set was tested using the procedure RT described above. Precipitation, air pressure, mean temperature, maximum

temperature and minimum temperature series were considered. Due to the methodological problems discussed above, the assessment was restricted to the period with the highest number of complete series (1960-2000). The finally tested series were further reduced due to the edge problem. Table 7 gives an overview of the number of series available in the ENSEMBLES database and the mentioned reductions. Because the RT results are based on the data set of November 2005 the number of stations differs from those of Section 2.3.

The results of the homogeneity assessment based on RT are shown in Figure 7. Different colours indicate the number of break points per series. Stations excluded from the analysis due to incompleteness or due to the edge problem are plotted in black. The exact numbers of detected break points and the mean lengths of a homogeneous sub period per variable are given in Tables 8 and 9.

Overall, it is found that air pressure series contain the highest number of break points (Figure 7b, Table 8). This result agrees with findings from other studies (e.g. Begert et al., 2003) and has several reasons. First, air pressure is highly correlated in space resulting in a good performance of relative homogeneity test methods. Smaller shifts can be detected compared to other variables. Second, small changes in the measuring conditions, such as a changing station height, can lead to substantial inhomogeneities. Long homogeneous air pressure series are generally rare.

Table 7: Total number of stations per variable in the ENSEMBLES data set, complete series in the period 1960 to 2000 and number of series tested with RT.

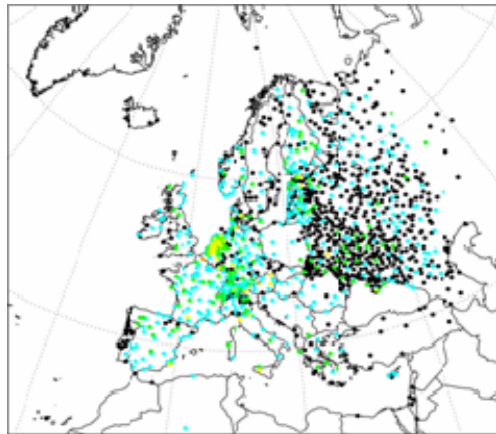
number of stations	tm		tn		tx		rr		pp	
total	631	(100%)	872	(100%)	841	(100%)	1446	(100%)	229	(100%)
complete in 1960 - 2000	463	(73%)	500	(57%)	499	(59%)	730	(50%)	134	(59%)
tested by RT	409	(65%)	446	(51%)	444	(53%)	672	(46%)	107	(47%)

Table 8: Number of series with zero, one, two, three and four or more break points and number of stations with no result from RT.

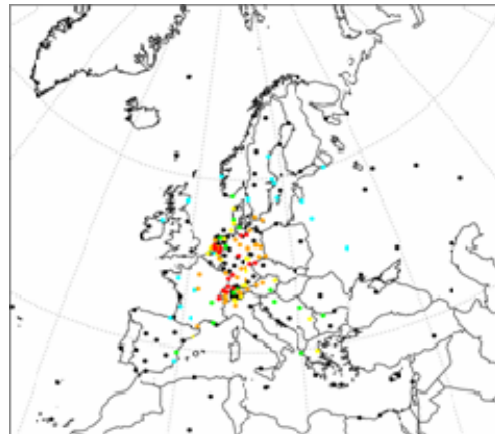
	tm		tn		tx		rr		pp	
homogeneous	145	(23%)	107	(12%)	187	(22%)	471	(32%)	20	(9%)
1 break point	155	(25%)	156	(18%)	156	(19%)	167	(12%)	15	(7%)
2 break points	75	(12%)	122	(14%)	76	(9%)	32	(2%)	24	(10%)
3 break points	29	(5%)	53	(6%)	22	(3%)	2	(<1%)	32	(14%)
4 or more break points	5	(<1%)	8	(1%)	3	(<1%)	0	(0%)	16	(7%)
undefined	222	(35%)	426	(49%)	397	(47%)	774	(54%)	122	(53%)

Table 9: Length of mean homogeneous sub period per variable.

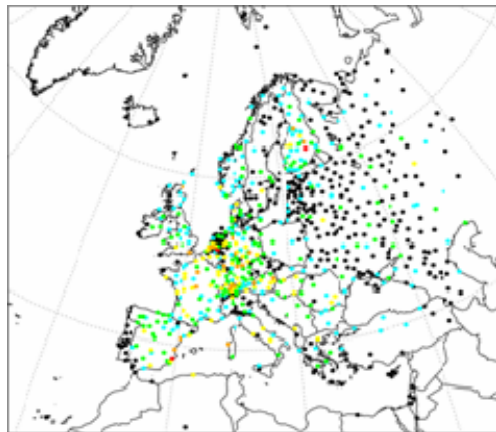
variable	mean homogenous sub period [y]
tm	20.4
tn	17.6
tx	21.9
rr	30.3
pp	13.3



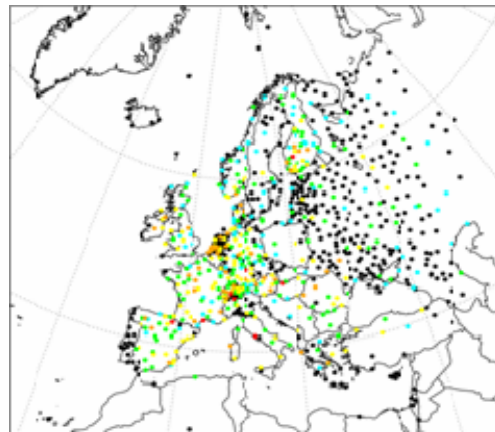
a) precipitation



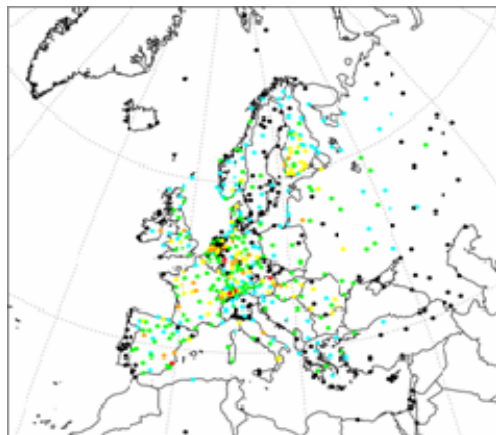
b) air pressure



c) maximum temperature



d) minimum temperature



e) mean temperature

Figure 7: Number of break points detected by RT (period 1960-2000): 0 (●), 1 (●), 2 (●), 3 (●), >4 (●), undefined (●) (i.e. incomplete over 1960-2000 or subject to the edge problem (see text)).

Mean temperature, minimum temperature and maximum temperature series contain similar mean homogeneous sub periods (Table 9) but remarkable differences are found concerning single series. Twice as much maximum temperature series are considered homogeneous compared to minimum series (Table 8). Obviously, changing measuring conditions have larger influence on minimum than on maximum temperature, possibly related to changes in ventilation. Furthermore, the number of break points in the different temperature series of a station often differs (Figures 7c-d). This is valuable additional

information compared to the result of the absolute testing procedure used in ENSEMBLES since no separate information for mean, minimum and maximum temperature is available there.

Precipitation series contain less inhomogeneities than the other variables (Figure 7a, Table 8). Precipitation also shows the longest mean homogeneous sub period (Table 9). Again this result is a combination of the fact that precipitation series are less frequently subject to changes in the measuring conditions and the limited efficiency of relative homogeneity tests due to a higher spatial variability of the variable.

Figure 8 shows the frequency distribution of the shift dimensions in air pressure and temperature series of the ENSEMBLES data set. Most of the shifts in air pressure series (around 80%) are between -1 and 1 hPa. Single shifts of up to -10 hPa and 5 hPa, respectively, occurred. For temperature, the shifts vary between -3 and 4.5 °C. However, large shifts also rarely occurred and most of them lie between -1 and 1°C. A broader distribution can be observed for minimum temperature than for maximum and mean temperature. Shift dimensions as well as such differences in distributions agree with findings for the Swiss temperature series in the period 1961 to 1997 (Begert et al., 2003).

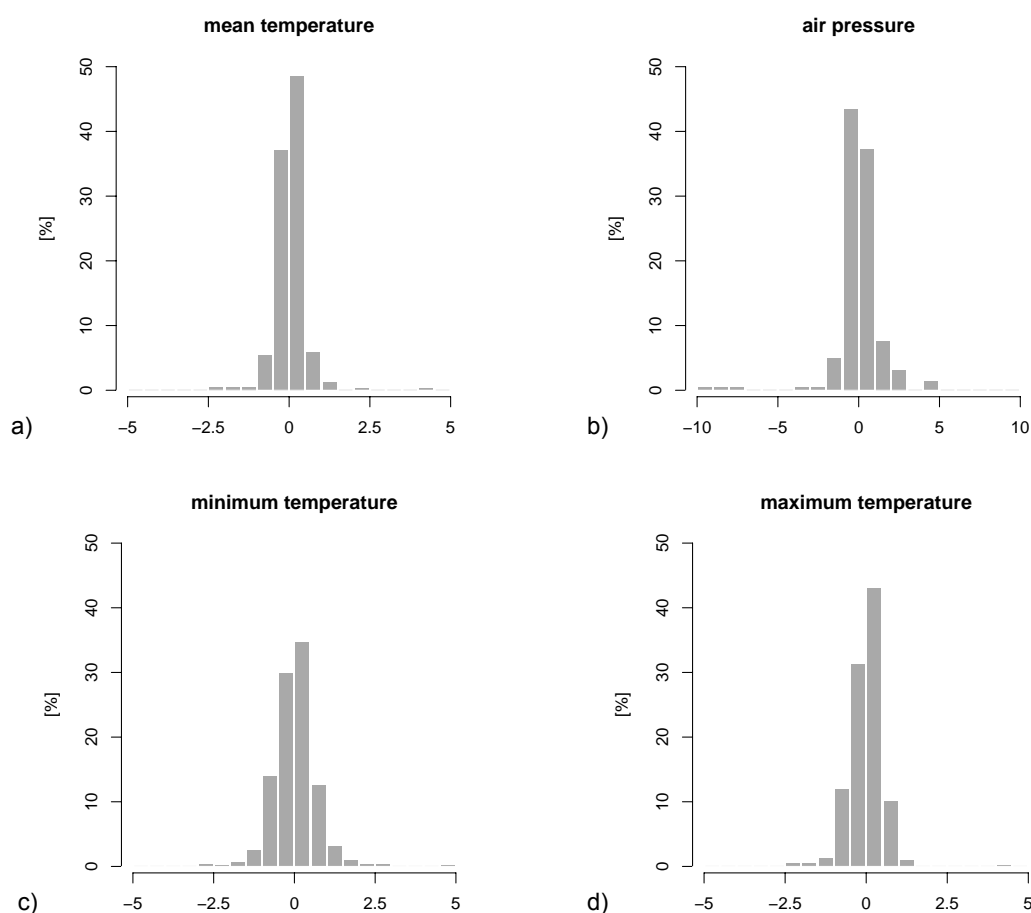


Figure 8: Frequency distribution of shift dimensions for air pressure in classes of 1 hPa (a), for mean temperature in classes of 0.5°C (b), for minimum temperature in classes of 0.5°C (c), and for maximum temperature in classes of 0.5°C (d).

The homogeneity assessment based on RT reveals relatively short homogenous sub periods. However, the calculation of the periods shown in Table 9 includes all the detected significant break points. To enlarge the length of useful periods, shifts smaller than a certain threshold can be neglected. The size of the threshold can be chosen according to the user's objectives.

4.4 Assessment of homogeneity

To assess which series and sub periods are useful for gridding, the results of AT and RT should be combined. We propose the following procedure:

1. Select from the AT results, one useful period for each series. Depending on the user's interest, this could be the most recent, the longest or the oldest useful period. Note that from each series, only one useful period can be selected to avoid inhomogeneities in the analyses!
2. If RT results are available, check for the homogeneous period whether RT detects an additional shift in this period and if so, only use a sub period. Note that RT results are not available for snow depth series. For snow depth, the useful period should be selected from the AT results only.

If we carry out this procedure for the various variables, we find that RT often detects break points in periods that are indicated as homogeneous by AT. For all precipitation series that were tested with both methods, RT finds additional break points (95% confidence level) in 57% of the useful periods detected by AT. For mean, maximum, minimum temperature and air pressure, this number is 30%, 27%, 36% and 22%, respectively.

We selected the longest homogenous period for each series in the database using the above procedure and counted the number of useful periods for each climate variable subdivided into four length classes (Table 10). This table shows that the number of long useful series (>40 years) has decreased significantly compared to using the AT results only (Table 4), while the number of short useful series has increased. The proposed procedure leads to a small decrease in the total number of useful series.

Table 10: Number of stations with homogeneous periods, classified as useful derived from the AT and RT result (see text). The stations are subdivided into 4 length classes. Only the longest homogeneous period of each station was taken into account. The last row gives the percentage of useful series with regard to the total number of series in the database.

Period length	tm	tx	tn	rr	pp	sd
10-20	213	203	239	183	54	4
20-30	178	176	165	336	60	8
30-40	99	109	105	253	17	4
>40	105	109	79	350	47	105
TOTAL	595	597	588	1122	178	121
PERCENTAGE	93%	62%	61%	77%	74%	90%

Apparently, the suggested procedure results in very few data useful for gridding. Therefore, one could decide to include also the doubtful series as found by AT. Additionally, RT's break points smaller than a certain threshold could be neglected. The selection of useful periods from the AT and RT results depends on the purpose of the research. Here, we give two examples:

Example 1

If the focus is on high-quality data of the most recent period for the gridding, the best procedure is to select series that are tested by both AT and RT. Start selecting the most recent useful periods from the AT results (one period for each data record). Then, check if RT detects an additional shift for this homogeneous period and if so, only use a sub period.

Example 2

If the focus is on using as much data as possible of intermediate quality for the gridding, the best procedure is to select the longest useful or doubtful period of each series from the AT results. Then, if RT results are available, check whether RT detects an additional shift larger than a certain threshold for this homogeneous period and if so, only use a sub period.

5 Summary and Conclusions

This report describes the collected station series that will be used for the European high-resolution gridding project of ENSEMBLES (WP 5.1). The data set was quality checked and the homogeneity of the series was tested. The data set includes series of six daily climate variables: minimum, mean and maximum temperature, precipitation sum, snow depth and mean sea level pressure. The report discusses the status of the database at January 2006. At that time, not all data delivered for the project were implemented in the database yet. Besides, several meteorological services confirmed to provide more station series in the near future. It is expected that in summer 2006, the most substantial data collection will be finalized. This is well in time for the gridding process of WP 5.1.

The European Climate Assessment and Data set (ECA&D) served as the starting point for the ENSEMBLES data set and its database structure was used for ENSEMBLES as well. Many additional daily time series have been gained directly from the National Meteorological Services, from various projects, such as STARDEX and EMULATE (two European FP 5 projects) and from existing data sets, such as ECA&D, the GCOS Surface Network, GHCND and MAP. All time series are from stations that reside in Europe or in countries that border the Mediterranean Ocean. The data providers that have contributed to the ENSEMBLES dataset should be acknowledged in the papers in which their data series are used.

Presently, the database stores series of 1670 stations. The station network is densest in the Netherlands and rather sparse in some other areas: southern Italy, Poland, Slovakia, Hungary, the Balkan countries, Turkey and North Africa. However, the number of series counts 1456 for precipitation, 966 for maximum temperature, 970 for minimum temperature, 639 for mean temperature, 239 for air pressure and 134 for snow depth. The latter are concentrated mainly in the Alps, the Netherlands and Bulgaria. Since not all series are complete or contain data for the most recent years, observations from SYNOP and CLIMAT messages are used to extend and fill gaps in the time series. Besides, observations from nearby stations are used for this purpose, provided that the stations are within 25 km distance and 50 m height difference. Depending on the climate variable, between 62% and 92% of the series that are in the database covers the time span 1960-2004.

All series are quality checked following the procedures of ECA&D and flagged accordingly. No corrections or adjustments are made. Three types of data flags exist: (o) data is useful, (1) data is suspicious and (9) data is missing. The homogeneity of each record is evaluated using a combination of an enhanced version of the absolute homogeneity analysis developed within the framework of ECA&D (Wijngaard *et al.*, 2003) together with a new automated relative method based on Steinacker *et al.* (2000) (VERAQC). These two methods are referred to as Absolute Test (AT) and Relative Test (RT), respectively. No adjustments to the daily values have been attempted.

AT is based on four statistical homogeneity tests, which are applied to five testing variables. These testing variables are derived from precipitation, minimum and

maximum temperature, snow depth and air pressure. AT searches for all homogenous segments in each series. The results of AT indicate that 193 temperature series, 486 precipitation series, 102 air pressure series and 121 snow depth series in the data base contain homogenous periods that exceed 40 years and start after 1960.

Depending on the variable, only 46% to 65% of the series in the period 1960 to 2000 could be tested with RT. By far the highest number of inhomogeneities was detected for air pressure, followed by temperature and precipitation. 19% of the air pressure series tested by RT are considered homogenous. The number for minimum temperature, mean temperature, maximum temperature and precipitation is 24%, 35%, 42% and 70% respectively.

RT further shows that that the mean, minimum and maximum temperature series of a single station often differ in number and/or date of break points. This points to a problem of AT where detected inhomogeneities in temperature series can not be assigned to minimum, mean or maximum series.

To assess which series and segments are useful for gridding, the results of AT and RT should be combined. It is suggested to use the homogenous periods as indicated by the AT results as a first step. As a second step, this period should be compared to the RT results, if available. Whenever RT detects an additional shift in this period, only a sub period should be used. When this procedure is followed strictly, only a few data series that contain homogeneous sub periods exceed 40 years: 350 precipitation series, 109 maximum temperature series, 79 minimum temperature series, 105 mean temperature series, 47 air pressure series and 105 snow depth series (Table 10). Therefore, alternative procedures are proposed to relax the criteria and select a set of series of a certain quality from the data set that is closer to the required number for gridding.

NB: The map plots were unfortunately created on old maps of Europe. We apology for not showing the borders of all countries.

References

- Aguilar E, Auer I, Peterson TC and Wieringa J, 2003. Guidelines on climate metadata and homogenization. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization, Geneva.
- Alexandersson H, 1986. A homogeneity test applied to precipitation data. *J. Climatol.*, 6, 661-675.
- Auer I, Böhm R and Schöner W, 2001. Austrian Long-term Climate 1767-2000. Multiple Instrumental Climate Time Series from Central Europe. Österreichische Beiträge zu Meteorologie und Geophysik, Vol. 25, Zentralanstalt für Meteorologie und Geodynamik (ZAMG): Vienna.
- Auer I, Böhm R, Jurkovic A, Orlik A, Potzmann R, Schöner W, Ungersbock M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, J. Moisselin JM, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajic-Capkaj M, Zaninovic K, Majstorovic Z, Szalai S, Szentimrey T, and Mercalli L, 2005. A new instrumental precipitation dataset for the greater alpine region for the period 1800-2002. *Int. J. Climatol.*, 25, 139-166.
- Begert M, Seiz G, Schlegel T, Musa M, Baudraz G and Moesch M, 2003. Homogenisierung von Klimareihen der Schweiz und Bestimmung der Normwerte 1961-1990. Schlussbericht des Projekts NORM90. Veröffentlichung der MeteoSchweiz, Vol.67. MeteoSchweiz, Zürich.
- Begert M, Schlegel T and Kirchhofer W, 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int. J. Climatol.*, 25, 65-80.
- Böhm R, Auer I, Brunetti M, Maugeri M, Nanni T, and Schöner W, 2001. Regional temperature variability in the European Alps 1760-1998 from homogenized instrumental time series. *Int. J. Climatol.*, 21, 1779-1801.
- Caesar J, Alexander L, and Vose R, 2006. Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, 111, D05101, doi:10.1029/2005JD006280.
- Conrad V and Pollak W, 1950. *Methods in climatology*. Second edition. Harvard University Press, Cambridge, MA.
- Della-Marta PM and Wanner H, 2006. A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate* (in print).
- Easterling DR and Peterson TC, 1995. A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.* 15, 369-377.
- Frei C and Schär C, 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, 18, 873-900.
- Häberli C, Groehn I, Steinacker R, Pöttschacher W and Dorninger M, 2004. Performance of the surface observation network during MAP. *Meteorologische Zeitschrift*, Vol. 13, No. 2, 109-121.
- Häberli C, 2005. *The Comprehensive Alpine Radiosonde Dataset (CALRAS)*. Dissertation, Institut für Meteorologie und Geophysik, Universität Wien, Austria.
- Lazante JR, 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde data. *Int. J. Climatol.* 16, 1197-1226.

- Peterson TC, Easterling DR, Karl TR, Groisman P, Auer I, Böhm R, Plummer N, Nicholis N, Torok S, Vincent L, Tuomenvirta H, Salinger J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones P and Parker D, 1998. Homogeneity adjustments of in situ climate data: a review. *Int. J. Climatol.*, 18, 1493-1517.
- Pettit A.N., 1979. A non-parametric approach to the change-point detection. *Appl. Statistics*, 28, 126-135.
- Slonosky VC, Jones PD. and Davies TD, 2001. Instrumental pressure observations and atmospheric circulation from the 17th and 18th centuries: London and Paris. *Int. J. Climatol.*, 21, 285-298.
- Steinacker R, Häberli C, and Pötschacher W, 2000. A transparent method for the analysis and quality evaluation of irregularly distributed and noisy observational data. *Monthly Weather Review*, 128(7), 2303-2316.
- Tuomenvirta H, 2001. Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data. *Int. J. Climatol.*, 21, 495-506.
- Von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Mathematical Stat.*, 13, 367-395.
- Wijngaard JB, Klein Tank AMG and Können GP, 2003. Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.*, 23, 679-692.