Article

An automated procedure to detect discontinuities; performance assessment and application to a large European climate data set

MICHAEL BEGERT[1][*], EVELYN ZENKLUSEN[1], CHRISTIAN HÄBERLI[1], CHRISTOF APPENZELLER[1] and LISETTE KLOK[2]

[1]Federal Office of Meteorology and Climatology (MeteoSwiss), Zürich, Switzerland
[2]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

Abstract

Within the framework of the EU-project ENSEMBLES, an automated procedure to detect shift inhomogeneities in climatological time series was developed and used for the homogeneity assessment of a European data set including values for temperature (min, mean, max), precipitation and air pressure covering at least 45 years. The automated process combines VERAQC (Vienna Enhanced Resolution Analysis Quality Control) output with Alexandersson's Standard Normal Homogeneity Test. VERAQC is designed to find outliers, systematic errors and meteorological biases in a time series comparing the data with concurrent observations of neighbouring stations. The resulting deviations can be analysed as time series for a single station and used as an input for relative homogeneity testing. Shift inhomogeneities are detected in accordance with statistical significance. An iterative procedure ensures that multiple break points are detected. The performance of the automated homogenization method was tested comparing its findings to results of a manually homogenized Swiss data set. The latter is based on a relative homogenization procedure in combination with an in-depth analysis of station history information. It is shown that the new automated procedure is able to detect shift inhomogeneities in temperature, precipitation and air pressure series but several limitations exist. Mainly the number of false alarms as well as missed break points is comparatively high at least in the Swiss Alpine test region with its complex topography. The homogeneity assessment of the ENSEMBLES data set for the period 1960-2004 revealed that 12 % to 59 % of the series were homogeneous depending on the variable. By far the highest number of shifts was detected in air pressure series followed by temperature and precipitation.

Zusammenfassung

Im Rahmen des EU-Projektes ENSEMBLES wurde ein automatisches Verfahren zum Auffinden von Shift-Inhomogenitäten in Klimadatenreihen entwickelt, getestet und auf den ENSEMBLES-Datensatz angewandt. Der Datensatz beinhaltet Messreihen der Temperatur (Mittel, Minimum, Maximum), des Niederschlags und des Luftdrucks von mindestens 45 Jahren Länge. Die entwickelte Homogenisierungsmethode besteht aus einer Kombination von VERAQC (Vienna Enhanced Resolution Analysis Quality Control) mit dem SNHT (Standard Normal Homogeneity Test) nach Alexandersson. VERAQC, entwickelt zur räumlichen Qualitätskontrolle von Daten, vergleicht für jeden Zeitpunkt den Wert einer Teststation mit einem aus umliegenden Nachbarstationen interpolierten Wert. Die Zeitreihe der resultierenden Abweichungen werden mit Hilfe des SNHT auf signifikante Inhomogenitäten geprüft. Das Verfahren wird iterativ angewandt, womit sichergestellt ist, dass mehrere Brüche pro Stationsreihen gefunden werden können. Um die Güte der automatisierten Methode zu ermitteln, wurden die Resultate mit solchen eines manuell homogenisierten schweizerischen Datensatzes verglichen. Hierbei zeigte sich, dass der automatische Homogenitätstest in der Lage ist, Shift-Inhomogenitäten in klimatologischen Zeitreihen korrekt zu detektieren. Zumindest in komplexer Topographie, wie den Schweizer Alpen, treten jedoch eine relativ hohe Anzahl falscher Alarme und nicht detektierter Inhomogenitäten auf. Die Ergebnisse der Homogenitätsprüfung des ENSEMBLES-Datensatzes für die Periode 1960-2004 zeigen, dass je nach Variable, 12 % bis 59 % der Klimadatenreihen homogen sind. Die grösste Anzahl der Brüche fanden sich in den Luftdruckreihen gefolgt von denen für Temperatur und Niederschlag.

1 Introduction

Climate model verification and climate monitoring require quality controlled, high resolution data sets ideally covering some previous decades. Such a European scale gridded daily data set is under development within the ENSEMBLES project (www.ensembles-eu.org; HEWITT and GRIGGS, 2004). In this paper the homogeneity of the European station series used as input data is assessed applying a new automated procedure. 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 re-

---

[*]Corresponding author: Michael Begert, MeteoSchweiz, Postfach 514, CH-8044 Zürich, Switzerland, e-mail: michael.begert@meteoswiss.ch

spect to homogeneous neighbouring stations. A combination of statistical methods and analysis of station history is considered to be the most effective approach.

The homogeneity assessment of the data set developed within ENSEMBLES (referred to as ENSEMBLES data set) aimed to provide the necessary uncertainty estimates of the data that are the base for the following gridding process. The large number of series to be tested required a fully automated procedure based on statistical tests only since metadata information was generally not available or only existent in native languages. HÄBERLI (2005) has demonstrated that a procedure combining the quality control method VERAQC (STEINACKER et al., 2000) with a conventional homogeneity test can meet the requirements. Following this idea we combined VERAQC with the standard normal homogeneity test developed by ALEXANDERSSON (1986) and automated the procedure. The test method is based on the analysis of monthly time series and focuses on the detection of shifts in mean since these are known to be the most common problems in climatological time series. Knowledge on existing shifts is useful for data gridding and support the decision whether to omit a certain series completely due to inhomogeneities or only use part of it.

This paper describes the developed automated detection procedure, its performance compared to a high quality homogeneous Swiss data set and its application to temperature, precipitation and air pressure series of the European data set collated in ENSEMBLES. After a short introduction of the data set chapters 3 and 4 explain the details and performance of the automated homogeneity test procedure. Chapter 5 shows the results of the method applied to the ENSEMBLES European data set and finally chapter 6 concludes and summarizes the experiences made during this study.

2 European daily data set

The daily data set of KLOK and KLEIN TANK (submitted), which updates the ECA data set of KLEIN TANK et al. (2002), will be used for the development of the ENSEMBLES gridded data sets. It contains series of about 2000 stations and 9 variables: minimum, mean and maximum temperature, precipitation, sea level air pressure, snow depth, sunshine duration, relative humidity, and cloud cover. Only, the first six variables will be used to create the ENSEMBLES gridded data sets. The homogeneity of these series is assessed in this paper, except for snow depth (see Chapter 3). Most of the stations in the data set contain precipitation series (90 %) or temperature series (60–67 %), whereas a smaller number holds air pressure (13 %) and snow depth data (8 %). Figure 1 shows the temporal data coverage for each climate variable. The best coverage is achieved after 1960. The strong decline in the precipitation series over the last 15 years is mainly caused by the precipitation series from the former Soviet Union, which cease in the

early nineties. The spatial distribution of the stations is depicted in the plots showing the number of breakpoints for each series (Figures 5 to 9). Data sparse areas are mainly found in Eastern Europe and Northern Africa.

All data in the data set have been automatically quality checked and flagged accordingly (KLOK et al., 2006). No corrections or adjustments were made to the time series. There are three types of data quality flags: (0) useful, (1) suspicious, i.e. the data value does not pass the quality control test and (9) missing. The quality control tests are absolute, implying that the data are not compared with respect to neighbouring station series. In total, the data set contains 84 % of useful data, 1 % of suspicious data and 15 % of missing data. For the homogeneity assessment, only useful data values were selected.

3 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, the relative homogeneity testing is usually based on the comparison of the candidate series with a reference series. Inhomogeneities in the candidate series are detected in the difference or ratio series (henceforth q-series) using suitable statistical methods. Most of the variables used for the ENSEMBLES gridded data set comply with the requirements for relative homogeneity testing. However, due to poor spatial and temporal correlation the concept does not work for snow depth. This variable was therefore excluded from the homogeneity assessment.

The relative homogeneity test procedures in general can be divided into two main steps: 1) the creation of the reference (q-) series, 2) the appliance of the statistical detection tests. As a fully automated method to build a q-series the Vienna Enhanced Resolution Analysis Quality Control (VERAQC) procedure was used. VERAQC was originally designed for quality control of meteorological measurements and runs on an operational basis at MeteoSwiss. VERAQC is based on an objective spatial interpolation algorithm. For each candidate station five surrounding stations are selected in a way in which their locations form an ideal pentagon around the central candidate station. A surface with minimized curvature is fitted trough 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 surface 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 MAP (Mesoscale Alpine Programme) data are given in HÄBERLI et al. (2004) and an
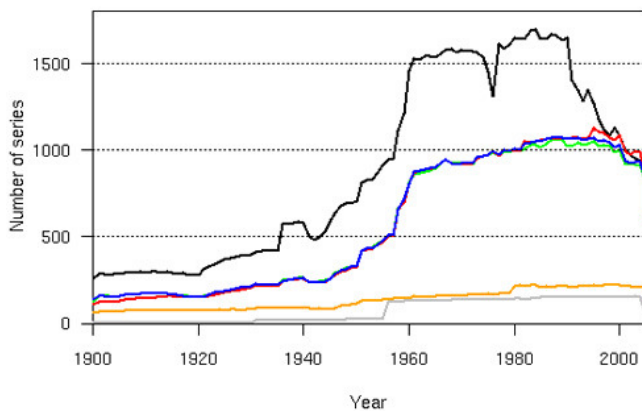
Figure 1: Number of stations for each year with series for precipitation (black), mean temperature (green), maximum temperature (red), minimum temperature (blue), air pressure (orange), and snow depth (grey). A year was counted when at least 292 days contain useful data (flag=0, see Section 2).

extension to test the relative homogeneity of radiosonde time series for the Alpine region in HÄBERLI (2005).

In the context of homogenization the time series of the differences calculated by VERAQC (henceforth deviation series) can be regarded as a q-series, as long as the inter-station distance is less than the decorrelation distance of the variable in question (SCHEIFINGER et al., 2003; AUER et al., 2005).

The ENSEMBLES data set does fulfil this requirement within the examined period 1960–2004. Variations due to non-climatic factors appear as shifts or trends in the deviation series and can be detected by statistical test methods. In order to enhance the power of the statistical tests the variability of the deviation series has been reduced by running VERAQC on monthly values. In addition, anomalies calculated from a common period were used as input values to avoid influences of a varying station network. Otherwise the appearance or disappearance of a neighbouring series can lead to abrupt changes in the deviation series similar to real breaks. Note that inhomogeneities in reference series may result in false alarms (HANSSEN-BAUER and FORLAND, 1994; MENNE and WILLIAMS, 2005). Although the setup of VERAQC minimizes the influence of inhomogeneities in single series, this drawback has to be accounted for when analyzing the homogeneity results of the method.

Alexandersson's standard normal homogeneity test (abbr.: SNHT, ALEXANDERSSON, 1986) has been used in the second step of the homogeneity test procedure. The SNHT method is designed to find shifts in mean and returns the date of one possible break point. 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 each

segment is determined as homogeneous or falls below a minimum length. Figure 2 gives an example of a deviation series and the according homogeneity test results for maximum temperature at the station Basel.

For temperature and air pressure the deviation series can also be used to examine the dimension of a detected shift in 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. Note that the method does not account for possible non-zero slopes in the segments that can lead to over- or underestimation of the magnitude (PIELKE et al., 2007). But magnitudes are only used to separate small from large shifts in the performance assessment of VERHOM (see chapter 4). 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 ratio series instead of difference series are commonly used in the homogenization process. Tests showed that the performance of VERHOM decreased when using ratios. As a consequence the difference series was also used for precipitation. This test procedure developed in the framework of ENSEMBLES is referred to as VERHOM (VERAQC Relative Homogenization Method) henceforth.

4 Performance assessment and limitations of the relative homogeneity test procedure (VERHOM)

The performance assessment was based on a comparison with high quality homogeneous Swiss series for the period 1961 to 1997. The latter have been homogenized using THOMAS, a tool specially developed for homogenization of climate series (BEGERT et al. 2003, 2005). Because THOMAS needs manual input it was not suitable for the homogeneity testing of a large data set like the one in ENSEMBLES. 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 serve as reference series. The inclusion of such series increases the false alarm rate as they might belong to another climate region. In short VERHOM cannot be expected to reach the performance of THOMAS, especially in a mountainous area
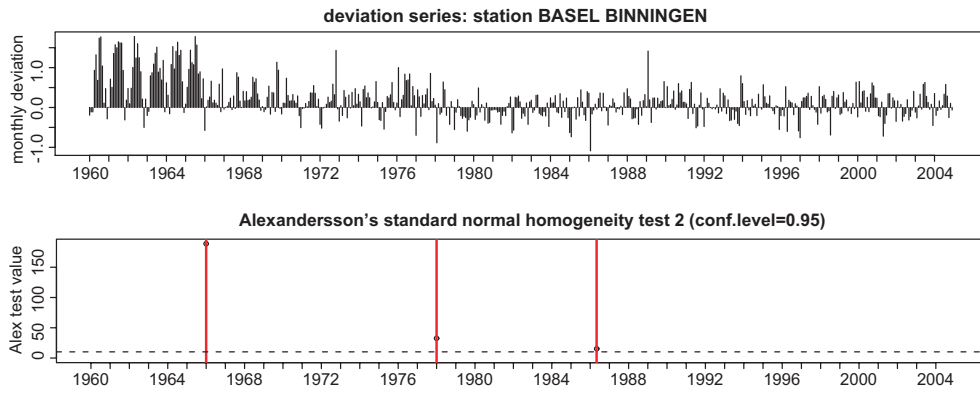
Figure 2: Deviation series (top) and homogeneity test results (bottom) for the maximum temperature series of Basel, Switzerland between 1960 and 2004. Small circles indicate the test values and the dashed line represents the 95 % significance level. Red lines mark the significant inhomogeneities.
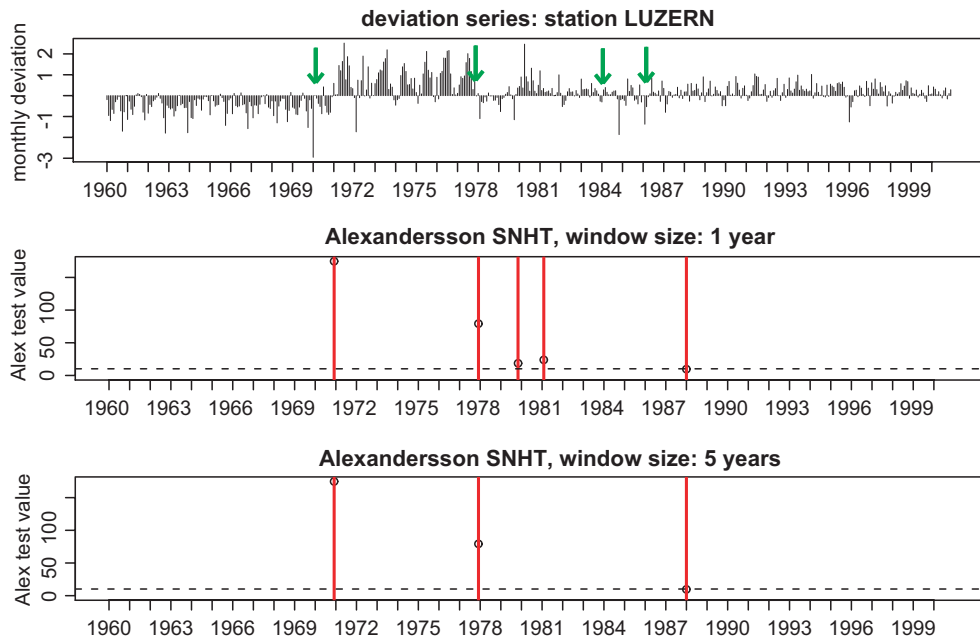

Figure 3: Test results SNHT for the maximum temperature series of Luzern, Switzerland between 1961 and 2000 using a minimum segment length of 1 year (middle) and 5 years (bottom). Findings of THOMAS are indicated in the deviation series (top) with arrows. Small circles show the test values and the dashed line represents the 95 % significance level. Red lines stress significant inhomogeneities.

Table 1: Left: number of large ($>= 0.5 \,^\circ$C) and small ($< 0.5^\circ$C) shifts detected by VERHOM and THOMAS in Swiss temperature series (mean, minimum, maximum). Right: number of large ($>= 0.5^\circ$C) and small ($< 0.5^\circ$C) false alarms and missed break points of VERHOM in Swiss temperature series (mean, minimum, maximum). In total VERHOM has detected 171 and THOMAS 259 breakpoints for the period 1961–1997.

		hits VERHOM	
		large	small
break points THOMAS	large	28	16
	small	9	19

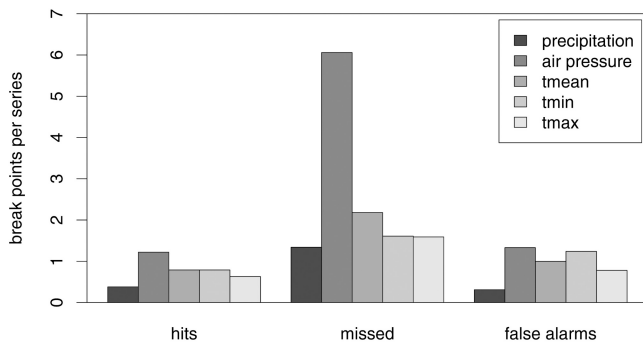|		VERHOM	
		false alarms	missed
large		19	49
small		80	138

Figure 4: Sum of THOMAS inhomogeneities in the Swiss station series detected by VERHOM, applied on monthly values for the period 1960–1997.

like Switzerland. However, the comparison gives an idea of the performance of VERHOM in the context of the homogeneity assessment of the large ENSEMBLES data set.

In a first step the THOMAS results were used to find the best setup for VERHOM whereas the THOMAS results were regarded as the truth. Hit rates and false alarm rates of both methods were compared in order to determine the most suitable minimum segment length and the adequate significance level for VERHOM. The choice of the minimum segment length turned out to be a compromise between a preferably high hit rate and an acceptable false alarm rate. The false alarm rate in particular was sensitive to the segment length. The optimal length was found to be 5 years independent of the variable. This result is in accordance with the recommendation for the SNHT (ALEXANDERSSON, 1986). As a consequence inhomogeneities closer than 5 years can not be detected. Figure 3 exemplarily shows the test results of SNHT using 1 and 5 year minimum segment lengths.

In a second step the results calculated automatically with the optimal setup of VERHOM were compared to the findings of THOMAS. 33 mean temperature, 33 minimum temperature, 32 maximum temperature, 32 precipitation and 18 air pressure series were finally used for the comparison. Figure 4 reveals the number of break points per series detected by VERHOM compared to the break points detected by THOMAS. In addition, the number of false alarms (significant break points of VERHOM, no THOMAS indication) and missed shift inhomogeneities (significant break points of THOMAS, no VERHOM indication) are shown. Break points detected within one year by both methods are referred to as hits. Overall the performance of VERHOM in detecting the THOMAS breaks is rather poor. About half of the shifts detected by VERHOM 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 and has to be interpreted with care. Missed break points can

be expected due to the implemented minimum segment length and the substantially sparser network density. The surprisingly high number of missed breaks in air pressure series for instance can be explained by the fact that Swiss air pressure series contain a high number of inhomogeneities in the 1980s because of technical problems. VERHOM is not able to detect breaks that follow each other within less than 5 years while there is no restriction in THOMAS because inhomogeneities are detected according to test results and station history analysis. The high number of missed breaks in temperature series on the other hand is due to small breaks that could not be detected by VERHOM. 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. In addition, SNHT tends to break up possible short trends in the deviation series into many small steps (REEVES et al., 2007). Although we introduced a minimum segment length to be tested this tendency might still contribute to the high false alarm rate.

Beside a comparison between "shift dates" a comparison of "shift dimensions" gives an idea of the performance of the current setup of VERHOM. Table 1 compares the number of small ($< 0.5°C$) and large ($\geq 0.5°C$) shifts detected by VERHOM and THOMAS in Swiss temperature series (mean, minimum and maximum series). In general most of the large shifts are characterized as large by VERHOM 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.

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 correlation 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 other 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 must still be expected. A comparison of the VERHOM results with findings for series in other countries would be a valuable completion of the current performance assessment.

Overall the new automated method VERHOM is able to detect shift inhomogeneities and estimate their dimensions. However, some limitations exist:

i. Inhomogeneities closer than 5 years apart can not be detected, which leads to missed break points.

ii. Stations located at the boundary of the investigated region cannot be tested, as the algorithm of VERAQC fails if the number of surrounding stations is

Table 2: Number of series with zero, one, two, three and four or more break points and number of stations with no result from VERHOM.

	mean temp		minimum temp		maximum temp		precipitation		air pressure	
homogeneous	245	(20 %)	168	(12 %)	231	(17 %)	1071	(59 %)	36	(12 %)
1 break point	316	(26 %)	323	(23 %)	300	(22 %)	406	(23 %)	45	(15 %)
2 break points	238	(19 %)	269	(20 %)	264	(19 %)	94	(5 %)	65	(21 %)
3 break points	104	(8 %)	147	(11 %)	128	(9 %)	21	(1 %)	47	(15 %)
4 or more break points	31	(3 %)	47	(3 %)	30	(2 %)	0	(0 %)	30	(10 %)
undefined	299	(24 %)	423	(31 %)	420	(31 %)	209	(12 %)	83	(27 %)

0(●), 1(●), 2(●), 3(●), >4(●), undefined (●)
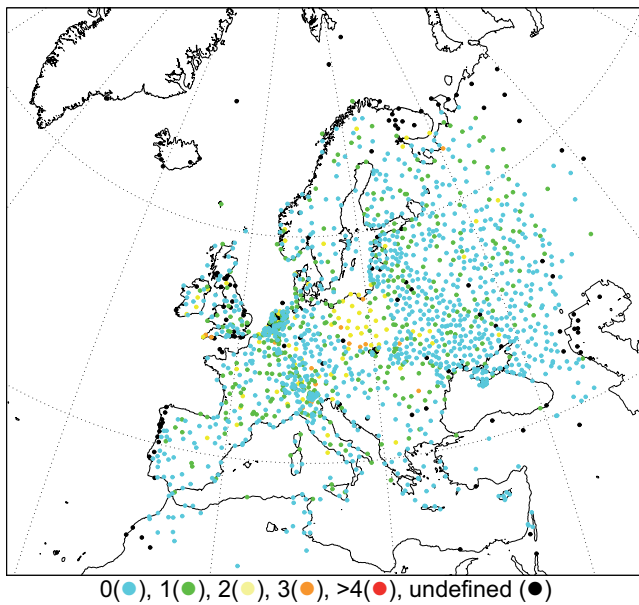
Figure 5: Precipitation, 1960–2004. Number of breakpoints detected:

Table 3: Length of mean homogeneous sub-period per parameter.

parameter	mean homogenous sub period [y]
mean temperature	17.7
minimum temperature	16.0
maximum temperature	17.1
precipitation	29.0
air pressure	15.5

insufficient. 10 % to 20 % of the series in the EN-SEMBLES data base are affected depending on the variable.

iii. As consequence of limitation ii., variations in the station network can cause incomplete deviation series close to the boundary if a neighbouring station appears or disappears in time. This can cause missed break points.
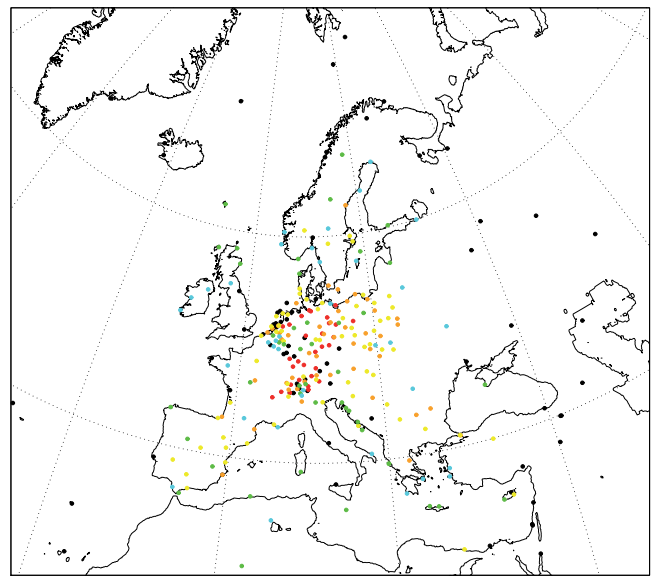
Figure 6: As Figure 5 but for air pressure.

5   Homogeneity assessment of the ENSEMBLES data set with VERHOM for the period 1960–2004

The homogeneity of the ENSEMBLES data set was tested using the procedure VERHOM described above. Precipitation, air pressure, mean temperature, maximum temperature and minimum temperature series were considered.

The results of the homogeneity assessment based on VERHOM are shown in Figures 5 to 9. 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 parameter are given in Tables 2 and 3.

Overall it is found that air pressure series contain the highest number of break points (Figure 6, Table 2). 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
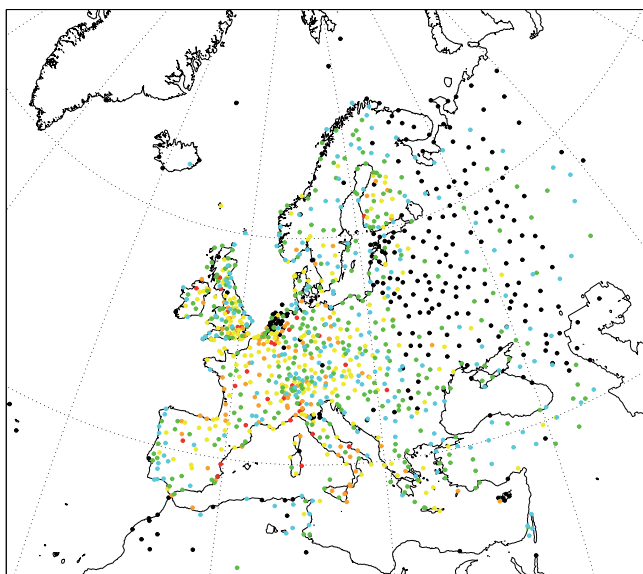
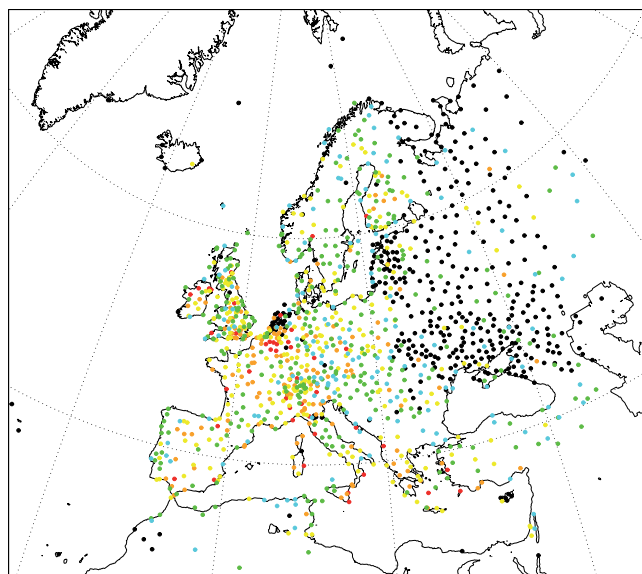**Figure 7:** As Figure 5 but for mean temperature.

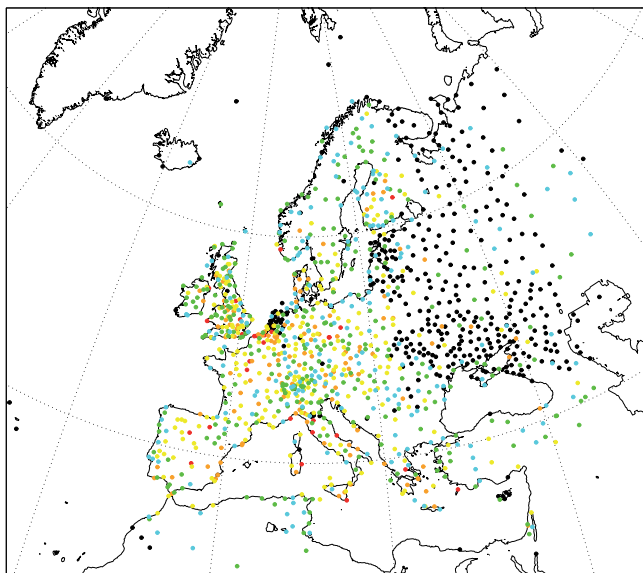**Figure 9:** As Figure 5 but for minimum temperature.

Figure 8: As Figure 5 but for maximum temperature.

good performance of relative homogeneity test methods. Smaller shifts can be detected compared to other parameters. 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.

For temperature series mean homogeneous subperiods of around 20 years are present (Table 3). Minimum temperature series contain the highest number of shifts followed by mean and maximum temperature (Table 2). In addition, the number of break points in the different temperature series of a station often differs (Figures 7–9). As seen in Swiss series the higher number of inhomogeneities in minimum temperature series might be a result of the fact that minimum temperature mea-

surements are more sensitive to disturbances in the measuring conditions (e.g. relocations) than the other temperature variables.

Precipitation series contain less inhomogeneities than the other parameters (Figure 5 to 9, Table 2). Precipitation also shows the longest mean homogeneous subperiods (Table 3). 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 and temporal variability of the variable. Figures 10 and 11 show 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 %) lie 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 3°C. Large shifts rarely occurred but most of the temperature shifts 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 found differences in distributions agree with findings for the Swiss temperature series in the period 1961 to 1997 (BEGERT et al., 2003).

6   Summary and conclusions

With VERHOM an automated relative homogenization procedure has been developed, tested and used for the homogeneity assessment of a large European data set collated in the framework of the EU-project ENSEMBLES. VERHOM combines VERAQC (Vienna Enhanced Resolution Analysis Quality Control) output with Alexandersson's standard normal homogeneity test. VERAQC uses a modified thin-plate spline interpolation to calculate a reference value for a station to be
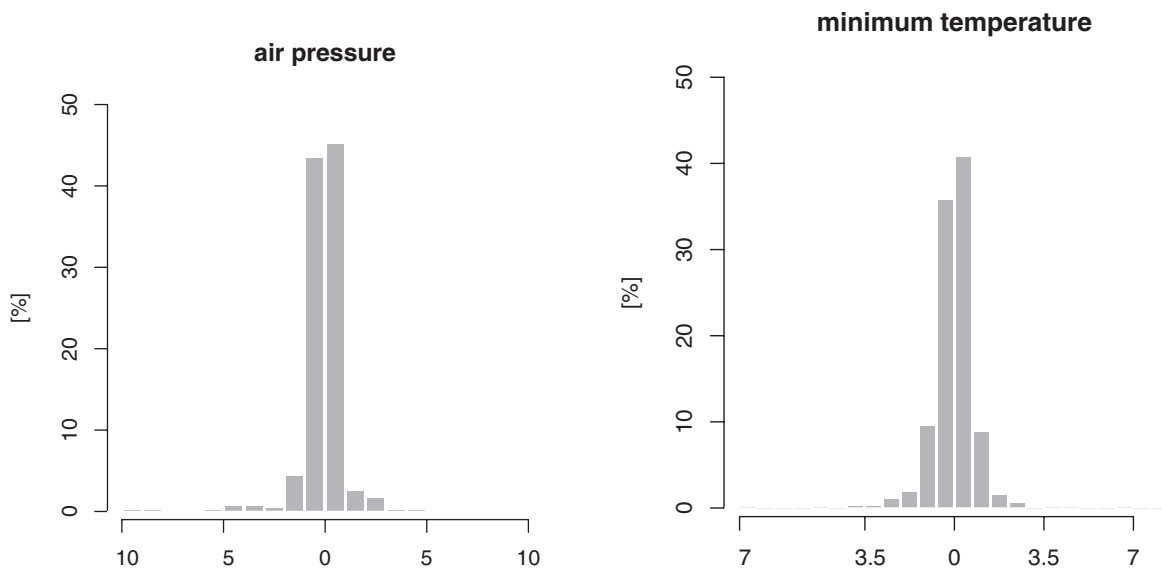
air pressure

minimum temperature

Figure 10: Left: frequency distribution of shift dimensions for air pressure in classes of 1 hPa. Right: frequency distribution of shift dimensions for mean temperature in classes of 0.5°C.
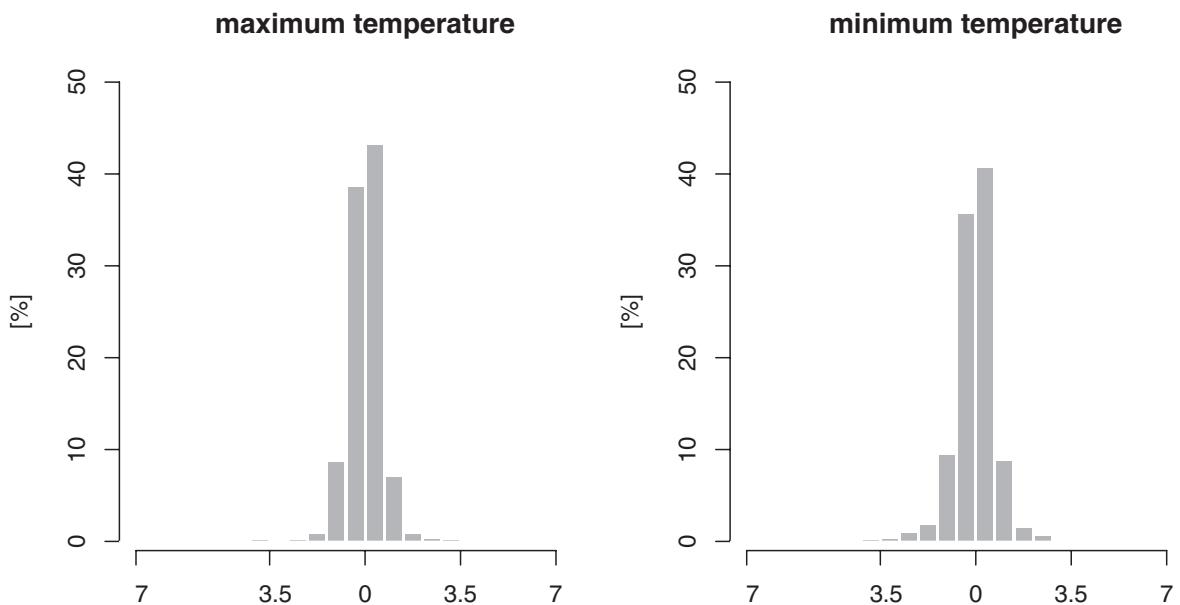
maximum temperature

minimum temperature

Figure 11: Left: frequency distribution of shift dimensions for maximum temperature in classes of 0.5°C. Right: frequency distribution of shift dimensions for minimum temperature in classes of 0.5°C.

tested. The resulting differences between measurements and calculated values can be analysed as time series for a single station and used as input for relative homogeneity tests. VERHOM is designed to detect shifts in mean in climatological time series.

To assess the performance of the automated procedure the findings were compared to results of a carefully homogenized data set in the Swiss Alpine region. The latter is based on the relative homogenization procedure THOMAS combining statistical methods with an in-depth analysis of the station history information. Overall the proposed new method VERHOM is able to

detect shift inhomogeneities in temperature, precipitation and air pressure series although the number of false alarms as well as missed break points is comparatively high. Depending on the variable 45 % to 60 % of the indicated shifts were false alarms and only 40 % to 55 % of the inhomogeneities found by THOMAS were detected. Since VERHOM was tested in a region with complex topography a better performance can be expected for most parts of Europe outside the Alpine region. The high false alarm rate for small shifts might also be due to inhomogeneities in the reference series built from surrounding stations. Addressing this problem would re-

quire additional strategies (e.g. GONZALES-ROUCO et al., 2001 or MENNE and WILLIAMS, 2005). About 10–20 % of all series could not be tested due to the fact that VERHOM is not able to process stations along the edge of the region covered by the observations. The assessment of the data homogeneity of the ENSEMBLES data set for the period 1960-2004 revealed that 59 % of the precipitation, 20 % of the mean temperature, 17 % of the maximum temperature, 12 % of the minimum temperature and air pressure series were homogeneous. By far the highest number of shift inhomogeneities was detected in air pressure series, followed by temperature and precipitation. The mean homogeneous sub-periods vary between 15.5 (air pressure) and 29 (precipitation) years.

List of acronyms

ENSEMBLES Research project supported by the European Comission under the 6th Framework Programme 2002–2006, Priority Global Change and Ecosystems. The focus is on Ensemble based predictions of climate changes and their impacts. www.ensembles-eu.org

SNHT Standard Normal Homogeneity Test

THOMAS Tool for Homogenization of Monthly Data Series

VERAQC Vienna Enhanced Resolution Analysis Quality Control

VERHOM Homogenization procedure which combines VERAQC with a relative homogeneity test.

Acknowledgments

References

AGUILAR E., I. AUER, T.C. PETERSON, J. WIERINGA, 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., R. BOEHM, A. JURKOVIC, A. ORLIK, R. POTZMANN, W. SCHÖNER, M. UNGERSBÖCK, M. BRUNETTI, T. NANNI, M. MAUGERI, K. BRIFFA, P. JONES, D. EFTHYMIADIS, O. MESTRE, J.M. MOISSELIN, M. BEGERT, R. BRAZDIL, O. BOCHNICEK, T. CEGNAR, M. GAJI-APKA, K. ZANINOVI, S. MAJSTOROVI, S. SZALAI, T. SZENTIMREY, L. PERCALLI, 2005: A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002. – Int. J. Climatol. **25**, 139–166.

BEGERT M., G. SEIZ, T. SCHLEGEL, M. MUSA, G. BAUDRAZ, M. MOESCH, 2003: Homogenisierung von Klimareihen der Schweiz und Bestimmung der Normwerte 1961–1990. – Schlussbericht des Projekts NORM90. Veröffentlichung der MeteoSchweiz **67**, MeteoSchweiz, Zürich.

BEGERT M., T. SCHLEGEL, W. KIRCHHOFER, 2005: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. – Int. J. Climatol. **25**, 65–80.

CONRAD V., W. POLLAK, 1950. Methods in climatology. Second edition. – Harvard University Press, Cambridge, MA.

GONZALEZ-ROUCO J.F., J.L. JIMENEZ, V. QUESADA, F. VALERO, 2001: Quality control and homogeneity of precipitation data in the southwest of Europe. – J. Climate **14**, 964–978.

HÄBERLI C., 2005: The Comprehensive Alpine Radiosonde Data set (CALRAS): Contribution to the diagnostic based on upper-air soundings 1957–1999. – Wien: Facultas, 2006. – VIII, 297 S. Wiener Meteorologische Schriften; 4. Zugl: Wien, Univ., Dissertation.

HÄBERLI C., I. GROEHN, R. STEINACKER, W. PÖTTSCHACHER, M. DORNINGER, 2004: Performance of the surface observation network during MAP. – Meteorol. Z. **13**, 109–121.

HANSSEN-BAUER I., E.J. FORLAND, 1994: Homogenizing long Norwegian precipitation series. – J. Climate **7**, 1001–1013.

HEWITT C.D., D.J. GRIGGS, 2004: Ensembles-Based Predictions of Climate Changes and their Impacts. – Eos Trans. AGU **85**(52), 566, 10.1029/2004EO520005.

KLEIN TANK A.M.G., J.B. WIJNGAARD, G.P. KÖNNEN, R. BÖHM, G. DEMARÉE, A. GOCHEVA, M. MILETA, S. PASHIARDIS, L. HEJKRLIK, C. KERN-HANSEN, R. HEINO, P. BESSEMOULIN, G. MÜLLER-WESTERMEIER, M. TZANAKOU, S. SZALAI, T. PÁLSDÓTTIR, D. FITZGERALD, S. RUBIN, M. CAPALDO, M. MAUGERI, A. LEITASS, A. BUKANTIS, R. ABERFELD, A.F.V. VAN ENGELEN, E. FORLAND, M. MIETUS, F. COELHO, C. MARES, V. RAZUVAEV, E. NIEPLOVA, T. CEGNAR, J. ANTONIO LÓPEZ, B. DAHLSTRÖM, A. MOBERG, W. KIRCHHOFER, A. CEYLAN, O. PACHALIUK, L.V. ALEXANDER, P. PETROVIC, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. – Int. J. Climatol. **22**, 1441–1453.

KLOK E.J., A.M.G. KLEIN TANK, submitted: Updated and extended European dataset of daily climate observations. – Int. J. Climatol.

KLOK E.J., A.M.G. KLEIN TANK, M. BEGERT, E. ZENKLUSEN, 2006: Assessment of the available station density for the gridding and daily data quality/homogeneity. – ENSEMBLES project report D 5.8.

MENNE M.J., C.N. WILLIAMS, 2005. Detection of undocumented changepoints using multiple test statistics and composite reference series. – J. Climate 18, 4271–4286.

PETERSON T.C., D.R. EASTERLING, T.R. KARL, P. GROISMAN, I. AUER, R. BÖHM, N. PLUMMER, N. NICHOLIS, S. TOROK, L. VINCENT, H. TUOMENVIRTA, J. SALINGER, E.J. FØRLAND, I. HANSSEN-BAUER, H. ALEXANDERSSON, P. JONES, D. PARKER, 1998: Homogeneity adjustments of in situ climate data: a review. – Int. J. Climatol. **18**, 1493–1517.

PIELKE R.A., C. DAVEY, J. ANGEL, O. BLISS, N. DOESKEN, M. CAI, S. FALL, D. NIYOGI, K. GALLO, R. HALE, K. HUBBARD, X. LIN, H. LI, J. NIELSEN-GAMMON, S. RAMAN, 2007: Documentation of bias associated with surface temperature measurement sites for climate change assessment. – Bull. Amer. Meteor. Soc. **88**, 913–928.

REEVES J., J. CHEN, X.L. WANG, R. LUND, Q.Q. LU, 2007: A review and comparison of changepoint detection techniques for climate data. – J. Appl. Meteor. Climatol. **46**, 900–914.

SCHEIFINGER H., R. BÖHM, I. AUER, 2003: Räumliche Dekorrelation von Klimazeitreihen unterschiedlicher zeitlicher Auflösung und ihre Bedeutung für ihre Homogenisierbarkeit und die Repräsentativität von Ergebnis-

sen. – 6. Deutsche Klimatagung, Klimavariabilität 2003, 22–25, September 2003, Potsdam. Terra Nostra, Schriftenreihe der Alfred–Wegener–Stiftung 2003/6: 375–379.

STEINACKER R., C. HÄBERLI, W. PÖTTSCHACHER, 2000: A transparent method for the analysis and quality evaluation of irregularly distributed and noisy observational data. – Mon. Wea. Rev. **128**, 2303–2316.