For more information, please contact:
World Meteorological Organization

Observing and Information Systems Department
Tel.: +41 (0) 22 730 82 68 – Fax: +41 (0) 22 730 80 21
E-mail: wcdmp@wmo.int

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland
www.wmo.int

World Meteorological Organization
Weather • Climate • Water

Climate Data and Monitoring
WCDMP-No. 78

Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the Meeting of COST ES0601 (HOME) Action MC Meeting

(Budapest, Hungary, 24-27 October 2011)

© World Meteorological Organization, 2011

The right of publication in print, electronic and any other form and in any language is reserved by WMO. Short extracts from WMO publications may be reproduced without authorization, provided that the complete source is clearly indicated. Editorial correspondence and requests to publish, reproduce or translate this publication in part or in whole should be addressed to:

Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 84 03
Fax: +41 (0) 22 730 80 40
E-mail: Publications@wmo.int

NOTE

The designations employed in WMO publications and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of WMO concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Opinions expressed in WMO publications are those of the authors and do not necessarily reflect those of WMO. The mention of specific companies or products does not imply that they are endorsed or recommended by WMO in preference to others of a similar nature which are not mentioned or advertised.

SEVENTH SEMINAR
FOR HOMOGENIZATION AND QUALITY
CONTROL IN CLIMATOLOGICAL DATABASES

Organized by Hungarian Meteorological Service (HMS)

jointly with the

MEETING OF

COST-ES0601 (HOME) ACTION
MANAGEMENT COMMITTEE

Budapest, Hungary, 24 – 27 October 2011

Organized by the Hungarian Meteorological Service

Supported by EU/COST-ES0601 Action, WMO, HMS

Edited by Mónika Lakatos, Tamás Szentimrey, Enikő Vincze

PREFACE

Long term instrumental climate records are the basis of climate research. However, these series are usually affected by inhomogeneities (artificial shifts), due to changes in the measurement conditions (relocations, instrumentation). As the artificial shifts often have the same magnitude as the climate signal, such as long-term variations, trends, or cycles, a direct analysis of the raw data series can lead to wrong conclusions about climate change.

The Seminars for Homogenization and Quality Control in Climatological Databases are traditionally held in Budapest and hosted by the Hungarian Meteorological Service since 1996. The 7th Homogenization Seminar was organized together the final meeting of the Management Committee of the COST Action ES0601: Advances in Homogenization Methods of Climate Series: an integrated approach (HOME). The COST HOME Action ended in October 2011 and its main objective was to review and improve common homogenization methods, and to assess their impact on climate time series. As one of the high importance achievements of the Action a benchmark dataset was generated for comparing monthly homogenization algorithms. The main results of this examination were published in the journal Climate of the Past.

The jointly organized 7th Seminar and the final MC meeting was a good occasion for conversation between the participants of the HOME Action and other researchers of the homogenization community. During this meeting, publishing a special issue of the COST HOME Action was suggested. This publication has been realized at the Quarterly Journal of the Hungarian Meteorological Service: IDŐJÁRÁS, Vol. 117, No.1. 2013, as a Special Issue of the Action.

The meeting was supported by EU/COST-ES0601 Action, WMO and OMSZ, and we hope that the series of Homogenization Seminars can co-operate with different other initiative for development of data quality with special regards to homogenization.

The Organizers

CONTENTS

PREFACE .....

CONTENTS 1

THEORETICAL ASPECTS OF HOMOGENIZATION .....

Tamás Szentimrey 2

INFLUENCE OF NETWORK DENSITY ON HOMOGENIZATION PERFORMANCE .....

José A. Guijarro 11

EFFICIENCIES OF HOMOGENISATION METHODS: OUR PRESENT KNOWLEDGE AND ITS LIMITATION .....

Peter Domonkos, Victor Venema and Olivier Mestre 19

ACMANT: WHY IS IT EFFICIENT? .....

Peter Domonkos 33

THEORETICAL QUESTIONS OF DAILY DATA HOMOGENIZATION AND DAILY BENCHMARK RESULTS FOR MASH .....

Tamás Szentimrey 45

APPLICATION OF THE REDISTRIBUTION METHOD ON CUSTOM DATA SUBSETS ..

Predrag Petrović 56

DETECTION AND CORRECTION OF BREAKPOINTS IN LONG-TERM GERMAN PRECIPITATION SERIES .....

E. Rustemeier, A. Kapala, V. Venema and C. Simmer 62

PORTUGUESE TEMPERATURE DATASET HOMOGENEITY WITH HOME R .....

L. Freitas, M. G. Pereira, L. Amorim, L. Caramelo, M. Mendes, Luís F. Nunes 72

CONSIDERATIONS ON SNOW AND TEMPERATURES TIME SERIES IN WESTERN ITALIAN ALPS .....

Fratianni S., Terzago S., Acquaotta F. 85

DATA HOMOGENIZATION IN THE CARPATCLIM PROJECT .....

Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S. 95

HOMOGENIZATION OF LONG MONTHLY AIR TEMPERATURE TIME SERIES COLLECTED IN UKRAINE .....

Osadchy V., Nabyvanets Y., Skrynyk O., Krakovska S. and Skrynyk O. 100

PROGRAMME 107

LIST OF PARTICIPANTS 110

THEORETICAL ASPECTS OF HOMOGENIZATION

Tamás Szentimrey
Hungarian Meteorological Service, P.O. Box 38, H-1525 Budapest, Hungary
E-mail: szentimrey.t@met.hu

1. INTRODUCTION

The theme of homogenization can be divided into two subgroups, such as monthly and daily data series homogenization. These subjects are in strong connection with each other of course, for example the monthly results can be used for the homogenization of daily data. This paper considers some various theoretical aspects of monthly series homogenization.
The main topics discussed are as follows.

- In case of relative methods the statistical spatiotemporal modelling of the series is a fundamental question. The adequate comparison, break point detection and correction procedures are depending on the chosen statistical model.
- Adequate validation statistics are necessary for the benchmarking of homogenization algorithms and software.
- Automation of the homogenization algorithms and software with special respect to the automatic usage of metadata.

2. STATISTICAL SPATIOTEMPORAL MODELLING OF MONTHLY SERIES

The aim of the homogenization procedures is to detect the inhomogeneities and to correct the series. In practice there are absolute and relative methods applied for this purpose. However the main problem of the application of absolute methods is that the separation of climate change signal and the inhomogeneity is essentially impossible. Relative methods can be applied if there are more station series given, which can be compared mutually. In this case the statistical spatiotemporal modelling of the series is a fundamental question. The adequate comparison, break point detection and correction procedures are depending on the chosen statistical model.

2.1. General structure of additive models

In case of relative methods a general form of additive model for more monthly series belonging to the same month in a small climate region can be written as follows,

$$X_j(t) = \mu(t) + E_j + IH_j(t) + \varepsilon_j(t) \qquad \left(j = 1,2,\ldots,N\,;\; t = 1,2,\ldots,n \right), \tag{1}$$

where $\mu(t)$ is the common and unknown climate change signal, E_j are the spatial expected values, $IH_j(t)$ are the inhomogeneity signals and $\varepsilon_j(t)$ are normal white noise series. Depending on model the signal $\mu(t)$ can be a fixed parameter or random variable. The type of inhomogeneity $IH(t)$ is in general a 'step-like function' with unknown break points T and

shifts $IH(T)-IH(T+1)\neq 0$, and $IH(n)=0$ is assumed in general. The inhomogeneity can be written as linear function of the shifts, $IH_j(t)=\mathbf{T}_j^{\mathrm{T}}(t)\mathbf{v}_j$, so from model (1) the following form can be obtained,

$$X_j(t)=\mu(t)+E_j+\mathbf{T}_j^{\mathrm{T}}(t)\mathbf{v}_j+\varepsilon_j(t) \quad (j=1,2,\ldots,N;\ t=1,2,\ldots,n), \tag{2}$$

where $\mathbf{T}_j(t)$ is the vector of one break point functions, and $\mathbf{v}_j$ is the vector of shifts at the break points.

THE ABOVE ADDITIVE (LINEAR) MODEL (2) MAY BE WRITTEN ALSO IN VECTOR FORM,

$$\mathbf{X}(t)=\mu(t)\mathbf{1}+\mathbf{E}+\mathbf{T}(t)\mathbf{v}+\boldsymbol{\varepsilon}(t) \qquad (t=1,\ldots,n) \tag{3}$$

where $\mathbf{X}^{\mathrm{T}}(t)=[X_1(t),\ldots\ldots,X_N(t)]$, $\mathbf{E}^{\mathrm{T}}=[E_1,\ldots\ldots,E_N]$, $\mathbf{T}(t)=[\mathbf{e}_1\mathbf{T}_1^{\mathrm{T}}(t),\ldots,\mathbf{e}_N\mathbf{T}_N^{\mathrm{T}}(t)]$, $\mathbf{v}=[\mathbf{v}_1^{\mathrm{T}},\ldots,\mathbf{v}_N^{\mathrm{T}}]^{\mathrm{T}}$, vector $\mathbf{1}$ is identically one, and the normal distributed vector variables $\boldsymbol{\varepsilon}(t)=[\varepsilon_1(t),\ldots,\varepsilon_N(t)]^{\mathrm{T}}\in N(\mathbf{0},\mathbf{C})$ $(t=1,\ldots,n)$ are totally independent in time. The spatial covariance matrix $\mathbf{C}$ describes the spatial structure of the series.

2.2 Statistical model assumed at MASH method

According to the assumption of MASH (Multiple Analysis of Series for Homogenization, *Szentimrey*, 1999, 2011), using the model (3) the signal $\mu(t)$ is a fixed parameter without any assumption about the shape, i.e. expected values are $\mathrm{E}(X_j(t))=\mu(t)+E_j+\mathbf{T}_j^{\mathrm{T}}(t)\mathbf{v}_j$. The covariance matrix $\mathbf{C}$ is also arbitrary.

2.3 Mixed Linear Model (MLM)

The Mixed Linear Model (MLM) is often applied statistical model for segmentation in biology and medicine (*Picard et al.*, 2007). This model can be also formulated according to the model (3) where the signal $\mu(t)$ $(t=1,\ldots,n)$ is assumed to be the random effect which models the correlations among the series, namely $\boldsymbol{\mu}=[\mu(1),\ldots\ldots,\mu(n)]^{\mathrm{T}}\in N(\mathbf{0},\mathbf{G})$. Further assumptions are that the covariance matrix $\mathbf{C}$ is a diagonal matrix again and the vector variables $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}=[\boldsymbol{\varepsilon}^{\mathrm{T}}(1),\ldots\ldots,\boldsymbol{\varepsilon}^{\mathrm{T}}(n)]^{\mathrm{T}}$ are independent. However the assumptions $\boldsymbol{\mu}\in N(\mathbf{0},\mathbf{G})$ and the diagonality of matrix $\mathbf{C}$ are unacceptable in case of climate series.

(i), (Temporal problem). If $\boldsymbol{\mu}\in N(\mathbf{0},\mathbf{G})$ were true then the expected values would be $\mathrm{E}(\mu(t))=0$ $(t=1,\ldots,n)$ consequently we exclude the climate change in mean implicitly.

(ii), Since vector variables $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ are independent therefore the model of fixed parameter $\boldsymbol{\mu}$ also can be applied where $\boldsymbol{\mu}$ as fixed parameter is the realization of random variable

$\boldsymbol{\mu} \in N(\mathbf{0}, \mathbf{G})$. However the fixed parameter $\boldsymbol{\mu}$ without any assumption about the shape is more general and more adequate model for climate series than the realization of random variable $\boldsymbol{\mu} \in N(\mathbf{0}, \mathbf{G})$.

III, (SPATIAL PROBLEM). ACCORDING TO LEMMA1 IN CASE OF CLIMATE SERIES $X_j(t)$ $(j = 1,2,\ldots,N)$ *AND FOR ALL THE SUBSET* $X_{j_1}(t), X_{j_2}(t), X_{j_3}(t), X_{j_4}(t)$ $(j_1 \neq j_3,\ j_2 \neq j_4)$,

$$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) \leq \operatorname{var}\left(X_{j_1}(t) - X_{j_3}(t)\right) + \operatorname{var}\left(X_{j_2}(t) - X_{j_4}(t)\right),$$

if $\mathbf{C}$ is a diagonal matrix.

But let us assume, $\left\|\mathbf{s}_{j_1} - \mathbf{s}_{j_3}\right\| << \left\|\mathbf{s}_{j_1} - \mathbf{s}_{j_2}\right\|$ and $\left\|\mathbf{s}_{j_2} - \mathbf{s}_{j_4}\right\| << \left\|\mathbf{s}_{j_1} - \mathbf{s}_{j_2}\right\|$,

where $\mathbf{s}_{j_1}, \mathbf{s}_{j_2}, \mathbf{s}_{j_3}, \mathbf{s}_{j_4}$ are the corresponding location vectors.
Then just the opposite inequality is expected in case of climate series, i.e.

$$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) > \operatorname{var}\left(X_{j_1}(t) - X_{j_3}(t)\right) + \operatorname{var}\left(X_{j_2}(t) - X_{j_4}(t)\right)$$

so the Mixed Linear Model (MLM) and the climate are in contradiction!

LEMMA 1 *ASSUMING MODEL (2), (3) IF* $\mathbf{C}$ *IS A DIAGONAL MATRIX, THEN THE FOLLOWING STATEMENTS ARE TRUE FOR THE VARIANCES,*

(i) $$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) = \operatorname{var}\left(\varepsilon_{j_1}(t)\right) + \operatorname{var}\left(\varepsilon_{j_2}(t)\right) \quad (j_1 \neq j_2)$$

(ii) $$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) \leq \operatorname{var}\left(X_{j_1}(t) - X_{j_3}(t)\right) + \operatorname{var}\left(X_{j_2}(t) - X_{j_4}(t)\right)$$
$$(j_1 \neq j_3,\ j_2 \neq j_4)$$

Proof.
(i)

$$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) = \operatorname{var}\left(\varepsilon_{j_1}(t) - \varepsilon_{j_2}(t)\right) =$$
$$= \operatorname{var}\left(\varepsilon_{j_1}(t)\right) - 2\operatorname{cov}\left(\varepsilon_{j_1}(t), \varepsilon_{j_2}(t)\right) + \operatorname{var}\left(\varepsilon_{j_2}(t)\right) = \operatorname{var}\left(\varepsilon_{j_1}(t)\right) + \operatorname{var}\left(\varepsilon_{j_2}(t)\right)$$

since $\operatorname{cov}\left(\varepsilon_{j_1}(t), \varepsilon_{j_2}(t)\right) = 0$, because $\mathbf{C}$ is a diagonal matrix.

(ii), As it was proved at i,

$$\operatorname{var}\left(X_{j_1}(t) - X_{j_2}(t)\right) = \operatorname{var}\left(\varepsilon_{j_1}(t)\right) + \operatorname{var}\left(\varepsilon_{j_2}(t)\right) \leq$$
$$\leq \operatorname{var}\left(\varepsilon_{j_1}(t)\right) + \operatorname{var}\left(\varepsilon_{j_2}(t)\right) + \operatorname{var}\left(\varepsilon_{j_3}(t)\right) + \operatorname{var}\left(\varepsilon_{j_4}(t)\right) =$$
$$= \left(\operatorname{var}\left(\varepsilon_{j_1}(t)\right) + \operatorname{var}\left(\varepsilon_{j_3}(t)\right)\right) + \left(\operatorname{var}\left(\varepsilon_{j_2}(t)\right) + \operatorname{var}\left(\varepsilon_{j_4}(t)\right)\right) =$$

and using again i, we obtain,

$$= \operatorname{var}\left(X_{j_1}(t) - X_{j_3}(t)\right) + \operatorname{var}\left(X_{j_2}(t) - X_{j_4}(t)\right)$$

2.4 Conlcusion

The adequate model for inhomogeneous climate series can be formulated by (2), (3) where the signal $\mu(t)$ is a fixed parameter without any assumption about the shape and the covariance matrix $\mathbf{C}$ is also arbitrary. The MLM can be taken such a special case of this model which is not enough general to model the climate series.

3. SERIES COMPARISON, BREAK POINT DETECTION, CORRECTION

3.1 Methodology for comparison of series

The problem of comparison of series is related to the following questions: reference series creation, difference series constitution, multiple comparisons of series etc. This topic is very important for detection as well as for correction, because the efficient series comparison can increase both the significance and the power. The development of efficient comparison methods can be based on the examination of the spatial covariance structure of data series. The examined series $X_j(t)\ (j=1,...,N)$ have to be taken as candidate and reference series alike, furthermore the reference series are not assumed to be homogeneous at the correct examinations!

The main problem arises from the fact that the shape of climate change signal is unknown. Therefore so-called difference series are examined in order to filter out the climate change signal $\mu(t)$. The simple difference series between pairs are $Z(t)=X_j(t)-X_i(t)$. However the difference series constitution can be formulated in more general way as well. Assuming that $X_j(t)$ is the candidate series and the other ones are the reference series, then the difference series belonging to the candidate series can be constituted as,

$$Z_j(t)=X_j(t)-\sum_{i\neq j}\lambda_{ji}X_i(t)=IH_j(t)-\sum_{i\neq j}\lambda_{ji}IH_i(t)+\varepsilon_{Z_j}(t) \tag{4}$$

with condition of $\sum_{i\neq j}\lambda_{ji}=1$ for the weighting factors. As a result of the last condition, the unknown climate change signal $\mu(t)$ has been filtered out. Consequently the inhomogeneities can be detected by the examination of the above difference series. In addition if we want to increase the signal to noise ratio in order to increase the power of detection then we have to minimize the variance of noise term $\varepsilon_{Z_j}(t)$.

The covariance matrix $\mathbf{C}$ uniquely determines the optimum weighting factors that minimize the variance, and the optimal difference series created in this manner can be applied efficiently for the detection and correction procedures (MASH, *Szentimrey*, 1999, 2011). We mention that in case of using the generalized-least-squares estimation for the unknown climate change signal $\mu(t)$, also the optimal difference series is obtained with minimal variance. We have to examine more difference series in order to separate the appropriate detected inhomogeneities for the candidate series. More difference series created without common reference series and with minimal variances can be defined as optimal difference series system (MASH).

3.2 Methodology for break point (changepoint) detection

One of the basic tasks of the homogenization is the examination of more difference series in order to detect the break points and to attribute them for the candidate series. The key question of the homogenization software is to develop automatic procedures for this attribution problem!!!

The scheme of the break point detection is as follows. Let $Z(t)$ be a difference series according to the formula (4), that is

$$Z(t) = IH_Z(t) + \varepsilon_Z(t) \qquad (t = 1,..,n), \tag{5}$$

where $IH_Z(t)$ is a mixed inhomogeneity of difference series $Z(t)$ with K break points $T_1 < T_2 < ..... < T_K$. In general the number K and the position of the multiple break points $T_1 < T_2 < ..... < T_K$ are unknown, furthermore the noise variables $\varepsilon_Z(t) \in N(E_Z, \sigma_Z^2)$ $(t = 1,..., n)$ are totally independent in time. The basic types of the detection procedures are the stepwise and the multiple break points detection. Let us have the following notation of the estimates: $\hat{K}$; $\hat{T}_1 < \hat{T}_2 < ..... < \hat{T}_{\hat{K}}$.

The more sophisticated multiple break points detection procedures were developed for joint estimation of the break points. There may be different principles of these methods that are classical ways in mathematical statistics.

3.2.1 Break point detection based on Bayesian Approach

The methods based on Bayesian model selection are the penalized likelihood methods. These methods are different in the penalty terms or criterions e.g. Akaike criterion, Schwarz criterion, Caussinus-Lyazrhi criterion.

The PRODIGE procedure (*Caussinus* and *Mestre*, 2004) based on the Caussinus-Lyazrhi criterion is an example for the penalized likelihood methods.

3.2.2 Break point detection based on Test of Hypothesis

Another possibility is to use hypothesis test methods for the detection of break points. At the MASH method (*Szentimrey*, 1999, 2011) a hypothesis test procedure has been developed, as we want to avoid the type one error that is the damage of data series. The essence of this multiple multiple break points detection procedure based on test of hypothesis on a given significance level is as follows.

If the detected break points of $Z(t)$ are $\hat{K}$; $\hat{T}_1 < \hat{T}_2 < ..... < \hat{T}_{\hat{K}}$, then on the given significance level p (e.g.: p=0.1):

(i) $Z(t)$ is not homogeneous above the intervals $(\hat{T}_{k-1}, \hat{T}_{k+1}]$ because,

$$\mathrm{P}\left(\exists(\hat{T}_{k\text{-}1},\hat{T}_{k+1}] \text{ above that: } Z(t) \text{ homogeneous} \right) = p$$

Consequently the detected break points $\hat{T}_k$ are not superfluous.

This means there is no serious type one error.

(ii) $Z(t)$ can be accepted to be homogeneous above the intervals $(\hat{T}_{k\text{-}1},\hat{T}_k]$.

This means there is no serious type two error.

Remark

Confidence intervals are also given for the break points beside the point estimation at the method MASH (*Szentimrey*, 1999, 2011).

3.3 Methodology for correction of series

Beside the detection another basic task of the homogenization is the correction of series. Calculating of correction factors can be based on the examination of difference series for estimation of shifts $IH(\hat{T}_k) - IH(\hat{T}_k + 1)$ $(k = 1,..., \hat{K})$ at the detected break points.

Almost all the methods use point estimation for the correction factors at the detected break points. For example the PRODIGE method (*Caussinus* and *Mestre*, 2004) uses the standard least squares technique to estimate the correction factors. Probably the generalized least squares estimation technique based on spatial covariance structure would be more efficient.

The MASH procedure (*Szentimrey*, 1999, 2011) is an exception because the correction factors are estimated on the basis of confidence intervals. The confidence intervals given for the break points and shifts make possible also the automatic usage of metadata at MASH!

3.4 Automation of methods and software

One of the fundamental problems of homogenization procedures is the relation of the manual versus interactive or automatic methods. In the practice the simple manual methods (e.g. Craddock method) are very popular however these ones are unusable for the real climate observation networks. We have to understand the fact that a lot of stations must be examined together in general! The reality for the number of stations per network is more than 100 instead of 10-15 used at COST HOME benchmark dataset.

Therefore the key questions of the homogenization methods and software are,

- firstly, the quality of homogenized data,
- secondly, the quantity of stations.

If we want to fulfill both respects it is necessary to develop automatic procedures.
Further necessary conditions required for automation of methods and software are,

- ability for automatic attribution of break points for the candidate series,
- and automatic usage of metadata.

To solve the above problems without advanced mathematics is impossible!!!

4. BENCHMARK FOR METHODS

The COST Action ES0601 (HOME) has executed a blind intercomparison and validation study for monthly homogenization methods. The methods were tested on a realistic benchmark dataset. The benchmark contained simulated data with inserted inhomogeneities (*Venema et al.*, 2012). Testing the methods on a generated benchmark dataset seems to be an objective validation procedure however we have to know also the limits of such type of examinations.

The interpretation of benchmark results is not a trivial problem, since these are depending on different factors, such as:

- tested methods (quality, manual or automatic),
- testing benchmark dataset (quality, adequacy),
- testers (skilled or unskilled),
- methodology of evaluation (validation statistics).

Moreover if want to obtain a real image of the methods, then the theoretical evaluation of their mathematical basis is also indispensible.
We do not intend to detail the above problems but some of them are presented in the following sections.

4.1 Comparison of manual and automatic methods

The question of the comparison of manual methods to automatic ones seems similar to the comparison of handmade and factory products. Or how can we compare the results of a manual time consuming method with a skilled tester versus the results of an interactive method with an unskilled tester. The method or the user is tested if we evaluate the test results?

4.2 Evaluation methodology and validation statistics

The methodology of evaluation and the derivation of validation statistics is not a trivial task at all. The following example is a tricky problem connected with the definition of reference period what is still unsolved in the practice.

Notations (additive model)
Month: $k = 1,2,\ldots,12$, year: $t = 1,2,\ldots,n$
Original homogeneous series (orig): $X_H^{(k)}(t)$
Inhomogeneous series (inhom): $X_{IH}^{(k)}(t) = X_H^{(k)}(t) + IH^{(k)}(t)$
Estimated inhomogeneity series: $I\hat{H}^{(k)}(t)$
Homogenized series (h0): $\hat{X}_H^{(k)}(t) = X_{IH}^{(k)}(t) - I\hat{H}^{(k)}(t)$

In the practice the generally used definition is that the reference time is *n*, i.e.,
$I\hat{H}^{(k)}(n) = 0$ or $\hat{X}_H^{(k)}(n) = X_{IH}^{(k)}(n)$

However in this case certain contradictions can be obtained.

Problem 1
Let us assume *n*-1 is the only one detected break, i.e.,
$\exists k:\ I\hat{H}^{(k)}(n-1) \neq I\hat{H}^{(k)}(n)$
Then the more intelligent methods do not change all the values for $t = 1,\ldots,n-1$. In this case the reference time is rather *n*-1 than *n*, and the last values are taken as outliers, since we do not want to damage all the series!
Consequently the homogenized series: $\hat{X}_H^{(k)}(t) = X_{IH}^{(k)}(t)\ (t = 1,.,n-1)$, $\hat{X}_H^{(k)}(n) = X_{IH}^{(k)}(n) - \hat{O}_k$,
where $\hat{O}_k$ are the estimation of the outliers at time *n*.

Problem 2
Let us assume that the benchmark data series was also defined with only one break at *n*-1,
$IH^{(k)}(t) \equiv C_k\ (t = 1,.,n-1),\ IH^{(k)}(n) = 0$
Then the difference series of the homogenized and original series is:
$Z^{(k)}(t) = \hat{X}_H^{(k)}(t) - X_H^{(k)}(t) \equiv C_k\ \ (t = 1,.,n-1)$ and $Z^{(k)}(n) = \hat{X}_H^{(k)}(n) - X_H^{(k)}(n) = -\hat{O}_k$

In this case the validation statistics as $RMSE(Z)$, $CRMSE(Z)$ may be very large, opposite to the fact, that the "homogenization" was quite good. The contradiction is arising from the non trivial problem of the definition, interpretation of the reference period.

5. SOFTWARE MASH

The most important properties of MASHv3.03
(Multiple Analysis of Series for Homogenization; *Szentimrey* 1999, 2008, 2011)

Homogenization of monthly series:
- Relative homogeneity test procedure.
- Step by step iteration procedure: the role of series (candidate, reference) changes step by step in the course of the procedure.
- Additive (e.g. temperature) or multiplicative (e.g. precipitation) model can be used depending on the climate elements.
- Including quality control and missing data completion.
- Providing the homogeneity of the seasonal and annual series as well.
- Metadata (probable dates of break points) can be used automatically.
- The homogenization results and the metadata can be verified.

Homogenization of daily series:
- Based on the detected monthly inhomogeneities.
- Including quality control and missing data completion for daily data.

New developments of version MASHv3.03
- COST ES0601 (HOME) format also can be used at I/O
- developments for automation

Some MASH specialty

- use of spatial covariance for comparison of series
- automatic attribution of break points for the candidate series
- automatic use of metadata

Software **MASHv3.03 c**an be downloaded from:
http://owww.met.hu/pages/seminars/seeera/downloads.htm

References

Caussinus, H, Mestre, O. 2004: Detection and correction of artificial shifts in climate series, Appl. Statist., 53, Part 3, pp. 405-425.

Picard, F., Lebarbier, E., Budinska, E., Robin, S. 2007: Joint segmentation of multivariate Gaussian processes using mixed linear models, Research resport 5, Statistics for Systems Biology Group, Jouy-en-Josas, France.

Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH), Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2008: Development of MASH homogenization procedure for daily data, Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary, 2006; WCDMP-No. 68, WMO-TD NO. 1434, 2008, pp. 116-125.

Szentimrey, T., 2011: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, pp. 64.

Venema, V. K. C. et al., 2012: Benchmarking homogenization algorithms for monthly data, Climate of the Past, 8, 89-115

INFLUENCE OF NETWORK DENSITY ON HOMOGENIZATION PERFORMANCE

José A. Guijarro
State Meteorological Agency, Palma de Mallorca, Spain (jguijarrop AT aemet.es)

Abstract

The comparison of homogenization methods has proven in COST Action ES0601 to be a tough and tricky issue: Many methods involve much work, and they must be applied to different types of series and inhomogeneities for a better assessment of their capabilities. Yet the benchmark has consisted of long and mostly complete series, while there are many shorter series in real data-sets whose information is disregarded when only the longer ones are used. To test the influence of using all these series for homogenization purposes, an R routine has been developed that generates networks of data with different degrees of correlation and completeness, introduce abrupt shifts □2°C to the first five series, and applies several homogenization methods (which must be able to run without any human intervention) to adjust them. Results show the poor performance of the absolute homogenization methods, with an overall good agreement between the other, and that the benefit of using short series is limited.

1. INTRODUCTION

It is quite frequent that the observing conditions of climatological stations suffer changes along their history, due to changes in instrumentation, relocations, changes in the environment of the station, etc. This changes normally affect the measurements, perturbing them and making difficult the study of the climate variability. Climatologist have been aware of this problem since long, and many methodologies have been proposed to detect and correct this inhomogeneities of the series. Reviews of the different approaches can be seen in PETERSON *et al*. (1998) and AGUILAR *et al.* (2003). Researchers had to implement their preferred methodologies by themselves in the past, but now there are a bunch of homogenization computer packages available in the Internet.

In the frame of COST Action ES0601, an extensive inter-comparison of homogenization methods was performed, comparing their ability to correct the variety of jumps, trends and outliers included in a complex realistic benchmark data-set of surrogate and synthetic series of monthly temperature and precipitation values. The details of this comparison exercise can be seen in Venema et al. (2012), where a wealth of conclusions are extracted and discussed about the different methodologies. Yet the complexity of the problem data-set prevented the comparison from being exhaustive: only participants using fully automatic algorithms could return complete results, while some time-consuming manual methods only could homogenize a reduced number of series. Moreover, some results produced by automatic methods were affected by errors not coming from the homogenization algorithms themselves, but rather from processes aimed at providing the results in the required output formats.

Other problems aroused from the complex realism with which the benchmark data-set was implemented. Surrogate series were developed mimicking the statistical properties of real climatological series assumed homogeneous, but as some inhomogeneities were still present in these example series, their surrogates exhibited some inhomogeneities as well. Moreover, as the magnitude of the shifts introduced in the problem series were derived from a normal distribution, many small shifts were accompanying the bigger ones, and hence the different

detection thresholds set in the compared methodologies introduced variations in the results not directly related to the algorithm performances.

On the other hand, and as an outcome of the meetings and discussions promoted by the COST Action HOME, the tested methodologies have improved after this first inter-comparison exercise. Yet implementing a new version of the benchmark data-set and repeating the homogenization exercise is not likely to take place in a near future due to the amount of work involved. Therefore, automatic benchmarking of the homogenization packages is an alternative worth to be explored, and this work is a first attempt in this line. The drawback is that only fully automatic methods can be tested in this way, but these can be foreseen as the more needed in operational climatology, where National Meteorological and Hydrological Services have to homogenize their data-bases repeatedly as they grow with the inclusion of new data. Manual methods may still be used in research works involving a limited number of selected long climatological series, but are impractical for the homogenization of whole dense networks.

Both manual and automatic methods are normally addressed to long selected series, with no or relatively few number of missing data. However, operational observation networks are often populated with stations with short recording periods, whose information in usually neglected in homogenization studies. The potential advantage of using this complementary information will be tested also in this first automatic inter-comparison exercise through the Climatol package, thanks to its high tolerance to missing data.

2. METHODOLOGY

Temperature is a key climatological variable in the context of present concern about the variability of the Earth Climate, and this first benchmarking exercise will focus on it, generating three basic thermometric networks of 100 locations randomly distributed on a 4x3° lon-lat area. The seasonal cycle was taken from the monthly averages of maximum daily temperatures of 53 stations from the central area of the Duero river basin, Spain (Table 1).

Table 1. Monthly averages of maximum daily temperatures (TA, °C) used for the generation of the basic networks

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TA	7.9	10.5	14.2	16.3	20.6	25.8	29.9	29.3	25.2	18.7	12.2	8.3

For the first location, a set of 12 monthly series of 60 terms each was randomly derived from normal distributions with the referred monthly means and a fixed standard deviation of 1.5°C(similar to the observed values). The closest location to this first simulated station had their data computed adding to the reference series a random normal value with zero mean and the same 1.5 standard deviation, multiplied by a factor of 0.20. Data for the other 98 stations were generated similarly, in order of minimum distance to any station whose data were already computed.

Two other basic networks were generated in this way, but using factors of 0.40 and 0.80, and the three networks were named TA20, TA40 and TA80. A final step consisted in adjusting the temperatures for altitude, varying the mean annual range of the stations in a random amount between -20 and 20%, and adding a constant trend of 2°C/Century to all series. In this way we obtained three basic networks with different degrees of inter-station correlations, as can be seen in Figure 1.

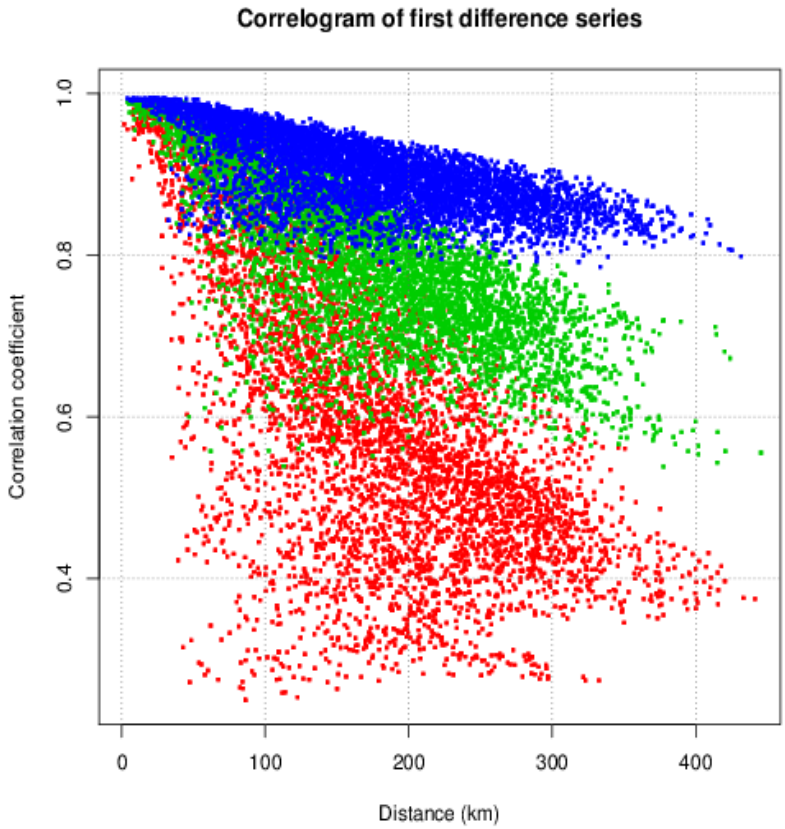


Figure 1. Correlogram of first differenced series of the three basic networks TA20 (very well correlated, in blue), TA40 (with intermediate correlations, in green) and TA80 (poorly correlated, in red)

For each of the three basic networks, 100 homogenization runs were performed, according to the following procedure:

1. 40 stations were randomly selected from the 100 available (Figure 2).
2. Shifts in the mean of □2°C were imposed on the first 5 stations (problem inhomogeneous series): two of them at fixed positions in series 1 to 3, and only one, but with random sign and location, in series 4 and 5. The last 10 years were always kept unaltered (Figure 3).
3. Stations 6 to 10 remained unchanged (complete homogeneous references).
4. Stations 11 to 40 were partially blanked, leaving only from 10 to 30 years of data (17 to 50 %), with random location of the data period (incomplete homogeneous references).

In each homogenization run, the problem series (the first five selected stations) were homogenized by some of the homogenization packages tested in Venema *et al.* (2012):

1. ACMANT (Acm).
2. Climatol: Applied to 10, 20 and 40 stations, with constant (cl1, cl2, cl4) and variable (Cl1, Cl2, Cl4) corrections.
3. RHTestV3: Absolute (a, A) and relative (r, R) homogenization strategies, adjusting the mean (a, r) or the quantile frequencies (A, R). (RHa, RHA, RHr, RHR).
4. HOMER: Pairwise detection + Joint detection + Correction (Hom).

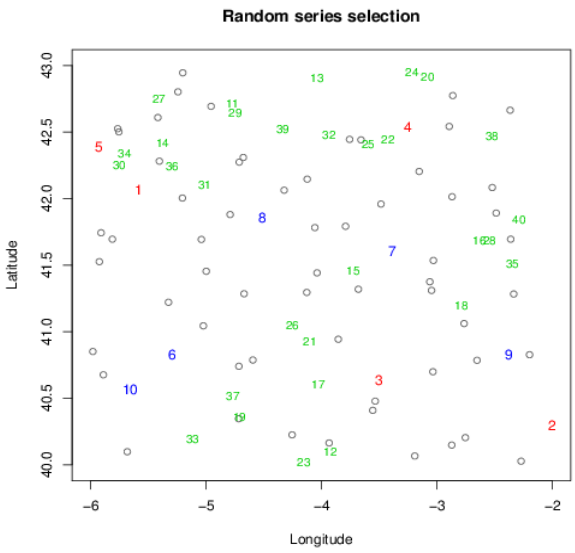


Figure 2. Example series: 40 locations (numbered) are randomly selected from the 100 base network. Series 1 to 5 will be inhomogenized (problem stations, in red), 6 to 10 will serve as complete homogeneous references (in blue), and series 11 to 40 will also be homogeneous, but only retaining 17 to 50% of their data (in green).

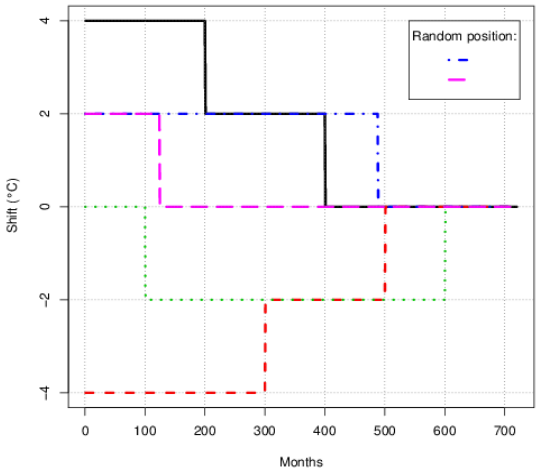


Figure 3. Inhomogeneities applied to series 1 to 5. Two fixed shifts for the first three series (in black, red and green), and only one, but randomly chosen, for series 4 and 5.

The general procedure was to give the first 10 complete series (5 highly inhomogeneous and 5 homogeneous) to every package and ask them to provide the solutions, adjusting the series to the last homogeneous period. One exception was RHTestV3, that cannot currently built a references series autonomously. In this case the nearest homogeneous series was provided for each of the five problem series. (This must be accounted for at the time of discussing the results, since the other methodologies have to build their reference series by themselves). The other exception was with Climatol, which in four applications could use the information of 10 (in cl2 and Cl2) and 30 (in cl4 and Cl4) additional stations with many missing data. (The other packages could not use this fragmentary information because of their limited missing data tolerance).

The whole process was implemented in an R script running on a Linux PC, and therefore ACMANT, being a Windows executable, had to be run through the Linux utility „wine". Other two homogenization packages running on Windows (MASH and AnClim) were tried

with the same strategy, but these attempts have been unsuccessful so far. Another homogenization package could not be tested (USHCN_v52d), but in this case the problem aroused from errors encountered when trying to compile its FORTRAN sources. Further efforts to incorporate these packages in this testing scheme will be done in the future.

3. RESULTS AND DISCUSSION

Four statistical parameters were computed for comparing the inhomogeneous problem series and the solutions provided for them with the original homogeneous series:

1. RMSE (main general measure of homogenization performance)
2. Difference of trends (relevant for climate change detection)
3. Difference of means (important for climate mapping)
4. Difference of standard deviations (useful for climate variability studies, although data at the daily scale are preferred for these)

The results for RMSE and trend differences are presented in the box-plots of Figure 4, where each box-plot summarizes 500 values (5 problem series and 100 runs). The scale of the graphs has been set constant for an easier interpretation, although some outliers lye outside the plot margins in the worst results. We can see that, in general, all the tested homogenization methodologies improve the problem inhomogeneous series (Inh), as we could expect given the high degree of inhomogeneity that was imposed to them: RMS errors are (mostly) lower than before the homogenization, and trend differences are unbiased even in the more dispersed results.

It is also clear how the lower correlations of networks TA40 and T80 make the homogenization problem increasingly difficult. This happens also in the case of the absolute homogenizations RHa and RHA, in spite of not using reference stations, because TA40 and TA80 series display higher variability as well, reducing the signal to noise ratio and hence compromising the detection of the shifts. Therefore, absolute homogenization should be avoided whenever possible, and efforts to get reference series, even from proxy variables, are always advisable.

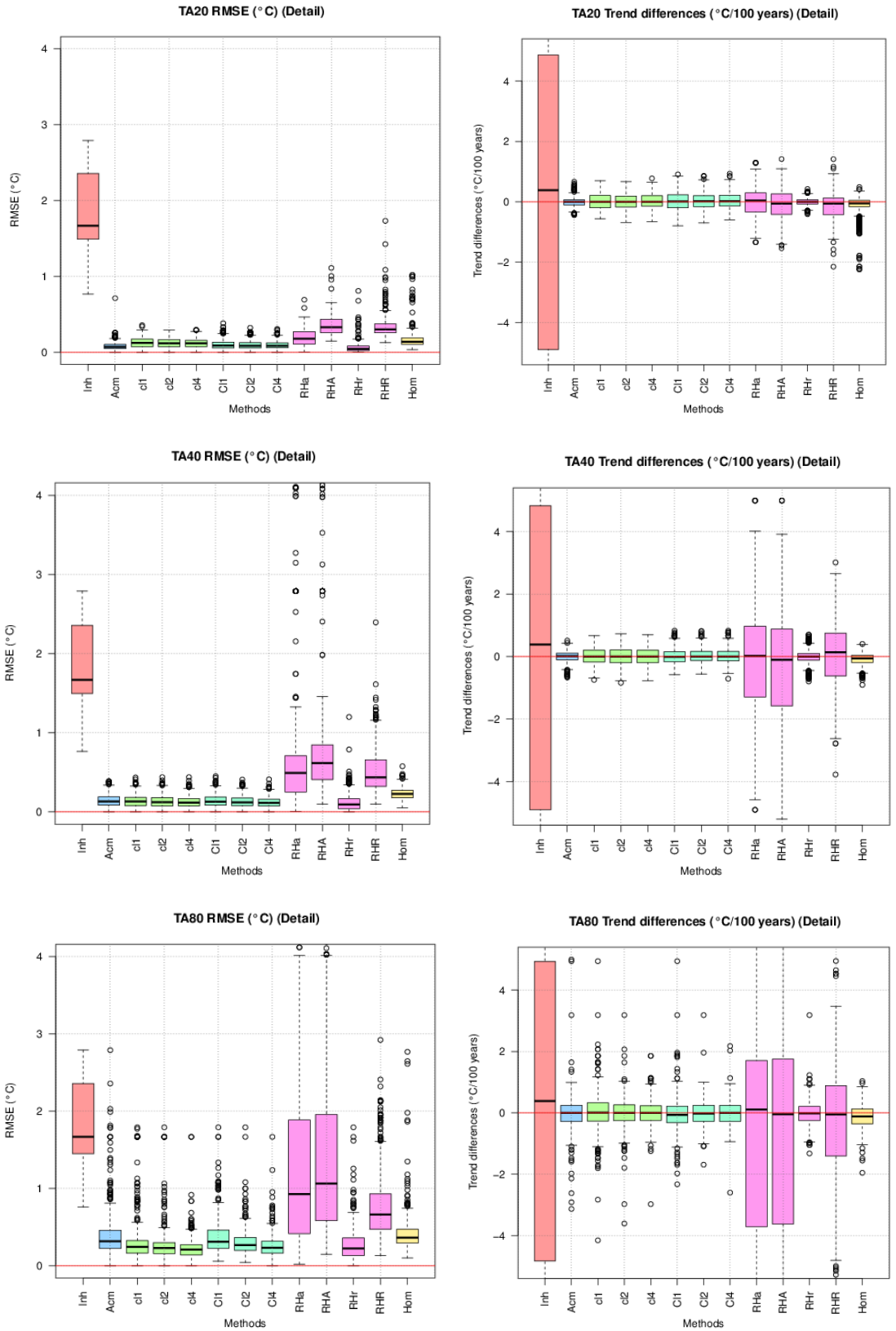


Figure 4. RMSE (left) and trend differences (right) of the problem series before (Inh) and after their homogenization by the tested methods. (Top to bottom plots: networks with good, intermediate and poor correlations)

All other results come from relative homogenization applications and, apart from RHR, give comparable results. The worse results of RHR come from the quantile matching correction, which also gave bad results in the previous benchmark comparison (Venema *et al.*, 2012). The other RHTestV3 result, RHr, yields much better results, as could be expected from the fact that the problem series have constant shifts. Actually, it ranks in the first places or very near, although the differences with other methods are subtle, and the fact that the best reference series was provided to it by the testing script may be unfair to the rest of the methods, that had to build their references themselves.

With respect to the use of additional information from fragmentary series, its advantage is only noticeable in the TA80 network, which has the lower inter-station correlations. Therefore, the ability of Climatol to manage series with a high rate of missing data will be more an advantage for these series (that will get their data gaps filled) than for the longer, more complete series. Applications Cl1, Cl2 and Cl4 differ from their counterparts cl1, cl2 and cl4 in the correction method, since the latter apply a constant correction term to all values, while the former adjust also the standard deviation. As we know that the problem only need a constant correction to be solved, it is satisfactory to see that the variable corrections do not worsen the results very much, as happened with the RHR solution.

The ranking of the results for trend differences is similar: once rejected the absolute homogenization approach and the RHR application, the remaining methods show very similar performance, with minor spread differences. The differences of means (not shown) exhibit the same results as the differences of trends, while the differences of standard deviations (not shown) have special characteristics: The constant correction methods (cl1, cl2, cl3 and RHr) yield the smaller (best) values, while the quantile adjustments applied by RHA and RHR give the worse results (biased and spread). Cl1 also produces dispersed values, that improve with the use of additional information in Cl2 andCl4 and approach their performance to the intermediate results achieved by Hom and Acm.

4. CONCLUSIONS AND FUTURE WORK

It is clear that absolute homogenization methods are to be avoided in favor of relative methodologies. Most of the other packages tested here showed similar performance, and the ability of Climatol to use series with many missing data has a limited effect on the improvement of its results.

Automatic benchmarking has proven a useful method for repeated validation and comparison of the results of automatic homogenization packages, and therefore further exercises are envisaged to expand the scope of the results in:

1. Including other packages in the comparison.
2. Introduce random and seasonally dependent shifts in the problem series.
3. Simulate variables with different seasonality and biased distribution (e.g.: precipitation).
4. Study the effect of shortening the last homogenous period.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J (2003): Guidelines on climate metadata and homogenization. World Meteorological Organization, WMO-TD No. 1186, WCDMP No. 53, Geneva, Switzerland, 55 pp.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salingeri J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998): Homogeneity adjustments of in situ atmospheric climate data: A review. Int. J. Climatol., 18, 1493–1517, .

Venema V, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Müller-Westermeier G, Lakatos M, Williams CN, Menne M, Lindau R, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquaotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P and Brandsma T (2012): Benchmarking homogenization algorithms for monthly data. Clim. Past, 8:89-115.

EFFICIENCIES OF HOMOGENISATION METHODS: OUR PRESENT KNOWLEDGE AND ITS LIMITATION

Peter Domonkos[1], Victor Venema[2] and Olivier Mestre[3]
[1]Center for Climate Change, Univ. Rovira i Virgili, Campus Terres de l'Ebre, Tortosa, Spain
[2]Meteorological institute of the University of Bonn, Germany
[3]Meteo France, Ecole Nationale de la Meteorologie, Toulouse, France
peter.domonkos@urv.cat

Abstract

In the recent years new methodological tools have been developed for measuring the efficiencies of homogenisation methods. This paper describes these new tools and presents the efficiency of some widely applied homogenisation methods focusing on the homogenisation of monthly temperature datasets. Conclusions are based on the synthesis of different testing methods. The results show that the best homogenisation methods are the PRODIGE, MASH, ACMANT and the Craddock test.

1. INTRODUCTION

A large range of homogenisation methods (HM) is available for climatologists to improve the quality of observational time series. However, selecting the best performing HM is not trivial. The core of the problem is that the efficiency depends on the properties of the time series the method is applied to. These properties of real networks are diverse, thus it is hard to provide generally valid settings for validation datasets. In addition, the practical efficiency depends on the purpose of homogenisation. For instance, the accuracy of linear trend estimation is usually more important in climatology than the accuracy of change-point estimations. Until recently, efficiency examination of homogenisation methods typically meant the computation of the detection skill in test datasets, composed of an arbitrary number of change-points and white noise. Although such examinations may serve important information for the method developers, the detection skills obtained in this way is not a relevant metric for the users of homogenised climatic data.

This study presents the latest results of the methodological development in measuring efficiencies of HMs, together with observed efficiencies for some widely used HMs. Three kinds of examinations are presented: i) Tests of detection parts only, in sets of relative time series (i.e. the difference between candidate series and reference series), ii) Blind tests of full homogenisation methods in the COST ES0601 (COST HOME) action with simulated datasets of networks, iii) The performance of ANOVA correction method with sets of detection results from various HMs. By comparing the results of these three kinds of examinations, the quality of the main components of HMs can be studied.

2. METHODS

In the first group of examinations (G1), the simulated test series are considered to be relative time series and the imaginary reference series are considered to be perfect. Consequently, there is no error due to imperfections in the reference time series used for detection and correction. Iterations or supplementary elements (e.g. metadata-use) are excluded. In this way

the only source of the errors is from the detection part of HMs. In the second group of examinations (G2) the blind test results with the COST HOME benchmark dataset (hereafter: Benchmark) are evaluated. In these experiments full networks are simulated and complete HMs are tested. In the third group of examinations (G3) the performance of ANOVA correction method is shown when it is coupled with the change-point detection results from various HMs.

The properties of test datasets for G1 are very different from those of G2, while for G3 the input field does not include test series. The examined HMs also differ for G1 and G2, partly because only detection parts can be tested with G1, and partly because not all HMS tested in G1 also participated in the openly announced benchmarking exercise (G2).

2.1. Input data

In G1 the simulated series are relative time series, and the inhomogeneities in them fully belong to the candidate series. The length of time series is always 100 years, the time-resolution is annual, and the background noise is white noise. The frequency, shape, and size-distribution of inhomogeneities differ in the three versions used in this paper. The mean shift-magnitude (m) is expressed as a ratio to the standard deviation of the white noise.

- Dataset A: One change-point is included in each time series. Its timing is year = 40 or year = 60. The shift-size is constant, $m = 3$.
- Dataset B: Five change-points per 100 years are included on average, but the actual number of change-points varies in the series of the dataset. The break sizes are normally distributed with a mean of zero, $m = 3.5$.
- Dataset C: This dataset contains a rather complex structure with randomly distributed inhomogeneities (IHs) of different types (change-points, platform-shaped changes and trends) and magnitudes, possessing similar statistical properties to those of detected IHs in relative time series derived from observed temperature time series in Hungary (Domonkos, 2011a). In this dataset the number of IHs is high, and short-term platforms are particularly frequent. This dataset was derived in a way that the statistical properties of detected inhomogeneities from simulated datasets made to be similar to those from true observed temperature datasets in Hungary through an iterative development of test datasets. Due to the way of its derivation the inhomogeneity-properties of this dataset are likely the closest to the real-world properties. The mean frequency of change-points is 31.1 per 100 years and $m = 1.2$.
- In G2 the Benchmark is used. In this dataset monthly temperatures of complete observational networks are simulated. The statistical properties of the temperature data in the Benchmark (moments, spatial correlations, seasonality, low frequency fluctuations) mimic the properties of true European observational temperature networks (Venema et al., 2011). The frequency, size-distribution and seasonality of inhomogeneities were set by an expert-team of the COST Action HOME, and finally the properties are close to those of dataset B in most respects (the mean frequency is 5 per 100 years and $m = 0.8$°C).
- In G3 the input data are not test series, but the list of timings of detected change-points and outliers in the public homogenisation of the Benchmark. These data are available for the public (Venema, 2011).

2.2. Homogenisation methods

In G1 nine HMs are examined. All of them are widely used, objective methods that can be applied automatically. In this paper the abbreviations of the HMs for G1 have always three letters, thus they may deviate somewhat from their widely used acronyms. The Bayes method (Bay, Ducré-Robitaille et al., 2003), Caussinus - Mestre method (C-M, also known as PRODIGE, Caussinus and Mestre, 2004), Easterling – Peterson method (E-P, also known as FTP, Easterling and Peterson, 1995), Multiple Analysis of Series for Homogenisation (MAS, Szentimrey, 1999), Multiple Linear Regression (MLR, Vincent, 1998), Standard Normal Homogeneity Test for shifts only (SNH, Alexandersson, 1986), Standard Normal Homogeneity Test for shifts and trends (SNT, Alexandersson and Moberg, 1997), t-test (tts, Ducré-Robitaille et al., 2003) and Wilcoxon Rank Sum test (WRS, Wilcoxon, 1945) are examined in G1.

Searching-algorithms for detecting multiple IHs, as well as parameterization for selecting significant IHs are usually the same as in the referred sources, but there are some deviations. The significance thresholds recommended by the constructors to ensure the 0.05 rate first type error in pure white noise processes are applied generally, but with some exceptions: The Caussinus – Lyazrhi criterion (Caussinus and Lyazrhi, 1997) is applied for Bay, while for MLR and tts the thresholds are based on the authors' Monte-Carlo simulations. In SNT the detected IHs are always trends when the estimated duration of change is at least 5 years, and always change-points in the reverse case. The version of MLR used in this study slightly differs from the original one, but it has no considerable effect on the detection results (not shown). Partly deviating from the original content of HMs, the cutting algorithm (Easterling and Peterson, 1995) is included in Bay, MLR, SNH, SNT and WRS.

A uniform pre-filtering of outliers is applied before the use of any HM, and the correction is also uniform, ensuring that any difference between the results is due to the different skills in the detection process.

In G2 and G3 the true contributions of the public homogenization of the Benchmark are examined. In this study only complete contributions with 15 homogenised networks from the surrogated temperature dataset are considered, except for the Craddock-test, which homogenised seven networks. The AnClim contribution was complete, but unfortunately had to be removed from the G3 examination because of discrepancies between applied and reported breaks. MASH detection has not been examined with ANOVA either, because it homogenizes every month separately (Venema et al. 2011).

From different versions of the same HM often the one with the best performance is selected only. The following HMs are analysed here: PRODIGE monthly (PROD), ACMANT late (ACMA, Domonkos, 2011b), MASH main (MASH), Craddock test by Vertacnik (Crad, Craddock, 1979), USHCN 52x (USHC, Menne and Williams, 2009), C3SNHT (SNHT), Climatol 2.1a (Clim, Guijarro, 2011), PMTred rel (PMT, Wang et al., 2007).

2.3. Efficiency measures

In this study efficiency (E) of homogenisation is usually expressed as the percentage of ceased root mean squared error (RMSE) relative to the RMSE in raw data (eq 1).

$$E = \frac{\mathrm{RMSE}_{\mathrm{raw}} - \mathrm{RMSE}_{\mathrm{hom}}}{\mathrm{RMSE}_{\mathrm{raw}}} \cdot 100\% \tag{1}$$

In some examinations centred RMSE (CRMSE) is applied. It is the RMSE of the anomalies from the mean. Let true values (homogenisation estimations) be denoted by t (x) in n-year long time series, then the CRMSE is formulated by (2).

$$\mathrm{CRMSE} = \sqrt{\frac{1}{n}\left(x_i - t_i - \frac{1}{n}\sum_{j=1}^{n}\left(x_j - t_j\right)\right)^2} \tag{2}$$

The advantage of CRMSE is that RMSE depends on the chosen fix point of the homogenisation, i.e. on a value or section of the time series that is considered to be free of errors, hence all the other values are adjusted to that. The drawback of CRMSE is that distortions of true spatial coherence in networks are not considered by that.

The CRMSE is chosen for evaluating E of annual value estimations in the Benchmark homogenisation (denotation: E_{CA}), because there was no fix point accepted by all contributors. The RMSE is applied for calculating the E for i) annual values in G1 (E_A), ii) trend-slopes for individual annual time series (E_T), iii) network-mean trend-slopes for the whole Benchmark-period (1900-1999, E_{TNL}), iv) network-mean trend-slopes for the second half of the Benchmark-period (1950-1999, E_{TNS}).

In the early part of Benchmark the ratio of missing data is high, therefore it has to be taken into account that the expected value of subset-means deviates from the expected value of the mean of all time series in network. Let the number of available stations be denoted by K for year j, then the subset-mean (denoted by X with upper stroke) is calculated by (3).

$$\overline{X_{j,K}} = \frac{1}{K}\sum_{k=1}^{K} x_{j,k} \tag{3}$$

When K equals the total number of stations the following notation is used:

$$\overline{X_j} = \overline{X_{j,k}} \tag{4}$$

Then the unbiased estimation of network-mean (Y) for year j is shown by (5).

$$Y_j = \overline{X_{j,k}} + \frac{1}{m}\sum_{i=1}^{m}\left(\overline{X_i} - \overline{X_{i,k}}\right) \tag{5}$$

In eq. 5 m denotes the number of years with available data in each station.

Detection skill is also applied. Let the number of right detections, that of false detections and that of all change-points be denoted by S_R, S_F and S, respectively. Then the detection skill (E_D) is calculated by (6).

$$E_D = \frac{S_R - S_F}{S} \tag{6}$$

The parameterisation of S_R, S_F and S is the same as in Domonkos (2011a).

2.4. ANOVA

The ANOVA is based on the minimisation of variance of homogenised data under the following criterions: i) The climate signal is the same for each time series, ii) the station-effect is always constant between two adjacent change-points of a time series. The minimum variance can be calculated by an equation system in which the time-dependent climate effects and the site-effects are the variables (Caussinus and Mestre, 2004).

The ANOVA provides the optimal solution of homogenisation-task when the input data meets with the written criterions. In practice, there is hardly any problem with the uniformity of climate signal, since networks are expected to be formed for a region of the same climate. The second criterion is more problematic, because the detected timings of change-points by HMs are usually neither complete nor accurate.

3. RESULTS

3.1. Tests of detection parts (G1)

Fig. 1 shows the E_A and E_T for Dataset A. Throughout this paper, the colour blue means C-M or C-M based HM, red means MAS and green means SNH or SNH based HM. It can be seen that in case of 1 change-point per time series the change-points are detected well with high certainty and with an ideally good reference series the homogenisation could be almost perfect, i.e. most efficiencies are above 90% and for E_A the values are often above 95%. E_A is always higher than E_T. E-P (tts) has slightly (markedly) poorer performance than the other HMs. The differences among the performances of other HMs are very small.

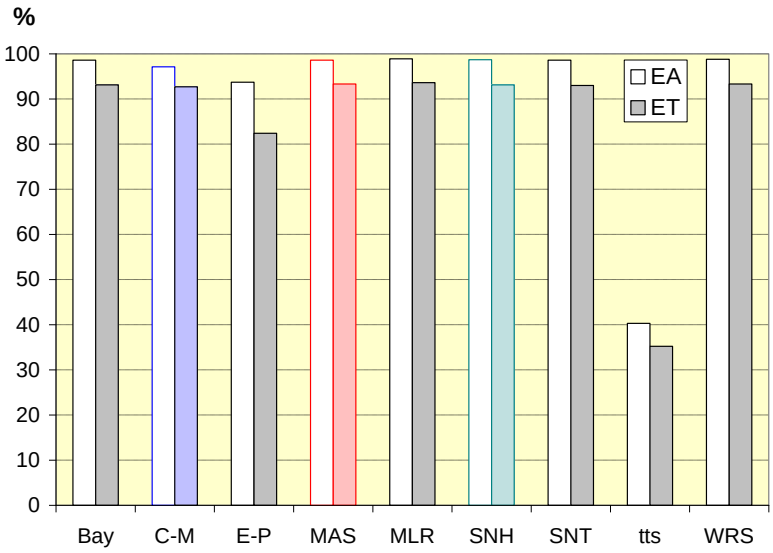


Figure 1. Efficiencies of nine detection methods in RMSE of annual values (E_A) and individual trends (E_T) for Dataset A

Fig. 2 presents the same kind results for Dataset B. In case of five change-points per time series on average, the homogenisation results are slightly poorer than in case of 1 change-point only, but the performances are generally still very good, the efficiencies are mostly above 90%. E_A is always higher than E_T again. Similarly as with Dataset A, E-P and tts show clearly poorer performance than the other HMs. Considering the mean of E_A and E_T, the rank order of the best HMs is C-M, Bay, MAS and SNH, but with insignificant differences in the performances (95.4%, 95.1%, 94.8% and 94.7%, respectively).

Fig. 3 presents the results for Dataset C. These results show that the presence of large number of small-size and platform-like inhomogeneities raises the uncertainty of inhomogeneity detection. The E-P and tts have very poor performances, while the other HMs produce rather similar results, around 70-75% efficiency. In this experiment E_T is sometimes

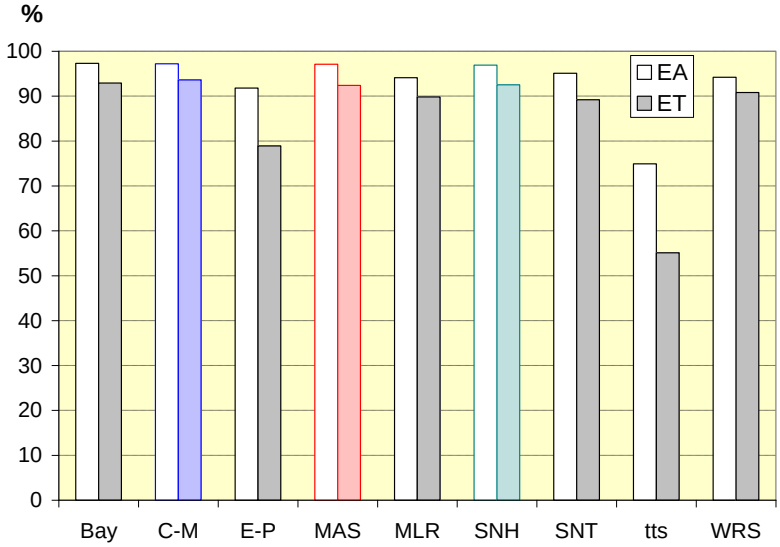


Figure 2. The same as Fig. 1, but for Dataset B

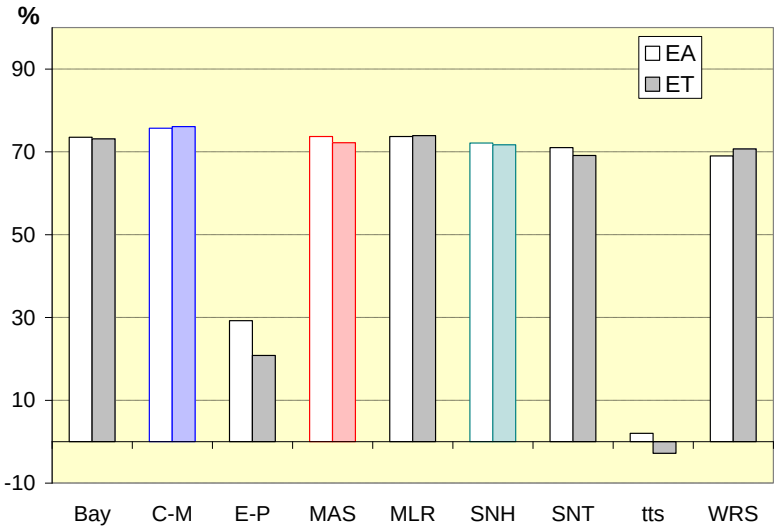


Figure 3. The same as Fig. 1, but for Dataset C

higher than E_A. Focusing on the rank order of the performances of the best methods, two important differences appear relative to the results of Dataset B: i) The lead of C-M is slightly larger, ii) The MLR has the second best performance here. The rank order: C-M (75.9%),

MLR (73.8%), Bay (73.3%), MAS (73.0%), SNH (71.9%). For Dataset C the detection skills are also calculated (Fig. 4). Here the C-M is the best again, but otherwise the picture is quite different compared to Fig. 3. The main differences are: i) Only the MAS has comparably good performance with C-M, ii) The performance of E-P and tts is not markedly poorer than that of the other HMs, moreover, counting with E_D the E-P would be the third best HM, iii) The most popular HMs (SNH, SNT) have markedly poorer performances than the best methods.

With the joint evaluation of E_A, E_T and E_D we can prove that the traditional detection skill examinations could not give a reliable picture about the true performance of HMs. The E-P seems to be one of the best methods when only the E_D is considered, but the reconstruction of true annual means and trends is more important in climatology than the ratio

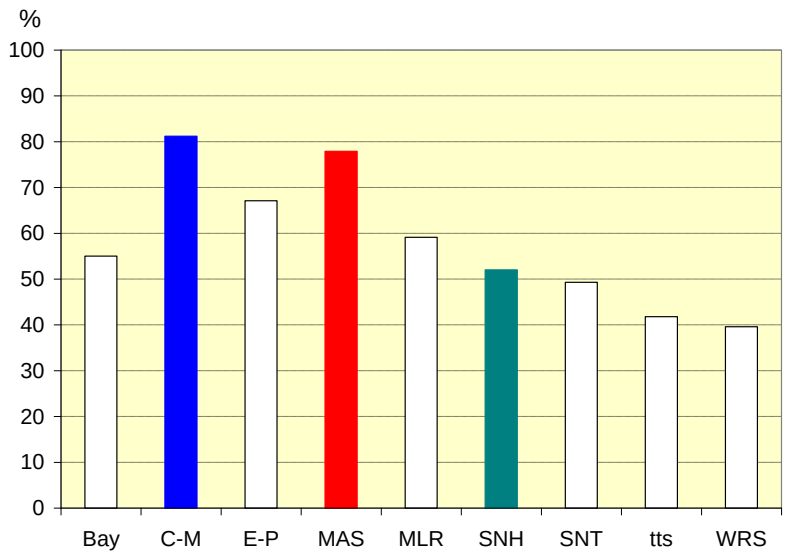


Figure 4. Detection skills (E_D) of nine homogenisation methods for Dataset C

of accurately detected change-points. On the other hand, when HMs of similar performances in E_A and E_T are examined, the differences in E_D is a valuable piece of information because in contrast with the simplified examinations in G1, the full homogenisation procedures often contain step-by-step elements, and for this reason the detection-errors in a certain step may affect the accuracy of estimations in later steps. All in all, the results of G1 show that C-M is the best detection method, followed by MAS. The other detection tools are less powerful, although in many cases the deficiencies from the best HMs are small.

3.2. Blind test results of the COST Action HOME (G2)

Fig. 5 presents the E_{CA} and E_T results for eight contributions which homogenised the Benchmark. In constructing this figure, HMs that are based on C-M, MAS or SNH detection are selected, as well as Crad that is included due to its outstanding performance.

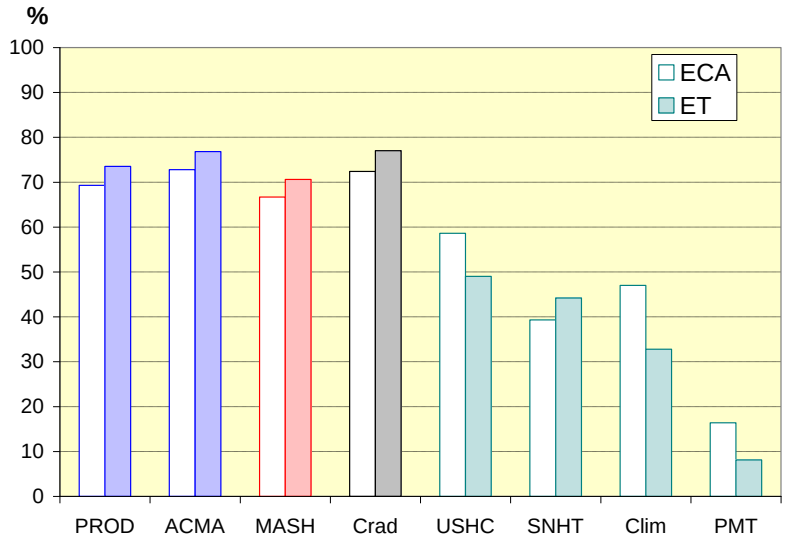


Figure 5. Efficiencies of various homogenisation methods in the Benchmark experiments in the CRMSE of annual values (E_{CA}) and RMSE of individual trends (E_T)

The main findings are as follows: i) The performances are much poorer than in the G1 experiment with Dataset B, which indicates that most homogenisation errors are not due to detection errors. ii) The efficiencies are positive what proves that the homogenised time series have better quality than before homogenisation. iii) The best four methods (including PROD and MASH) have a markedly better performance than SNH-based HMs. iv) ACMA and Crad have even better performance than PROD and MASH.

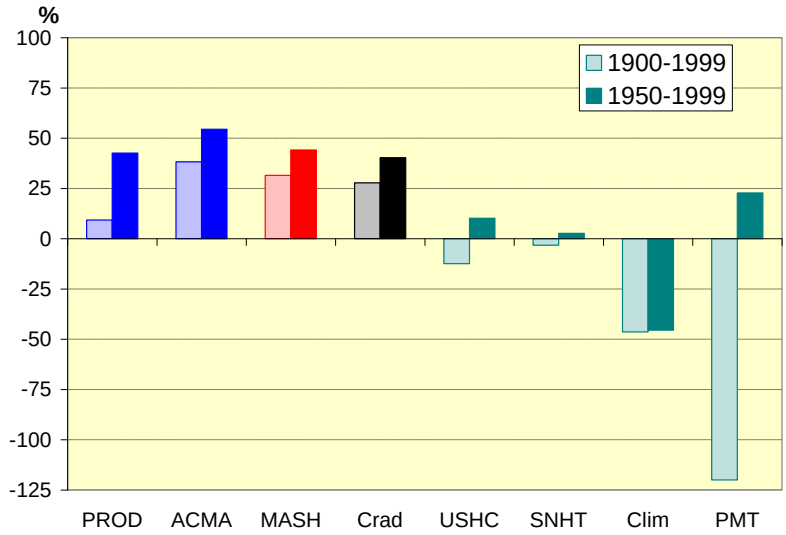


Figure 6. The same as Fig. 5, but for the RMSE of network-mean trends between 1900-1999 (E_{NL}) and between 1950-1999 (E_{NS})

Fig. 6 presents the performances in network-mean trend estimations. The mean bias of network mean trends in raw data is 0.39°C/100yr (0.69°C/100yr) for 1900-1999 (1950-1999). The results show that it is a difficult task to reduce these trend errors by homogenisation. Particularly E_{NL} is generally low or often even negative. Its likely explanation is that in the early section of the time series the ratio of missing data is very high. The rank order of the best HMs is somewhat different here than in Fig. 5. The best is the ACMA, which is followed

by the MASH, Crad and PROD. Note that due to the small sample-size the stochastic noise is very high in these characteristics.

3.3. Unified evaluation of G1 and G2

First it has to be clarified that the possibility of reliable comparisons between the results of G1 and G2 is limited, because the full detection process of HMs is often different from the basic method. For instance, the detection part of ACMA is a modified version of C-M-detection. In addition, ACMA is applicable only for monthly data, and it is recommendable only when station-effects have monthly cycle (series of monthly temperatures). Similarly, the detection parts of USHC, Clim and PMT are not exactly identical with SNH, there are differences in the selection of significant change-points and sometimes also in other details. Crad, being subjective, has no counterpart in the G1 examinations. Note that objective versions of accumulated anomaly based detection methods, similar to that of Crad, exist (Buishand, 1982) and they have been tested in G1-like examinations (Domonkos, 2008). According to those examinations the accumulated anomaly based detection methods have similar performance to that of WRS, which means that its deficiency from the best methods is moderate. Note also that the detection part of PMT was also tested (Domonkos, 2011a) and its performance was almost exactly the same as that of the SNH.

One interesting finding is that for G2, the performance of PROD and MASH relative to the SNH-based HMs is much stronger than in G1 results. Its explanation is the consequent mathematical structures in the whole homogenisation processes of PROD and MASH. While in G1 results the differences between performances are often very small, in G2 examinations only two HMs have comparable or better results than PROD and MASH. These two methods are the ACMA and Crad. ACMA is an improved version of PROD, and the results show that the improvement has been successful. The good performance of Crad has a very different explanation. The Crad incorporates the unique features of human intelligence that are, according to our results, difficult to convert perfectly into automatic or semi-automatic HMs.

3.4. ANOVA with various detection results

This section examines the performance of ANOVA correction method, when it is applied to the detection results of various HMs that originally did not apply ANOVA for correction. The results are presented in Fig. 7 and 8. Fig. 7a (7b) shows the impact of ANOVA on E_{CA} (E_T).

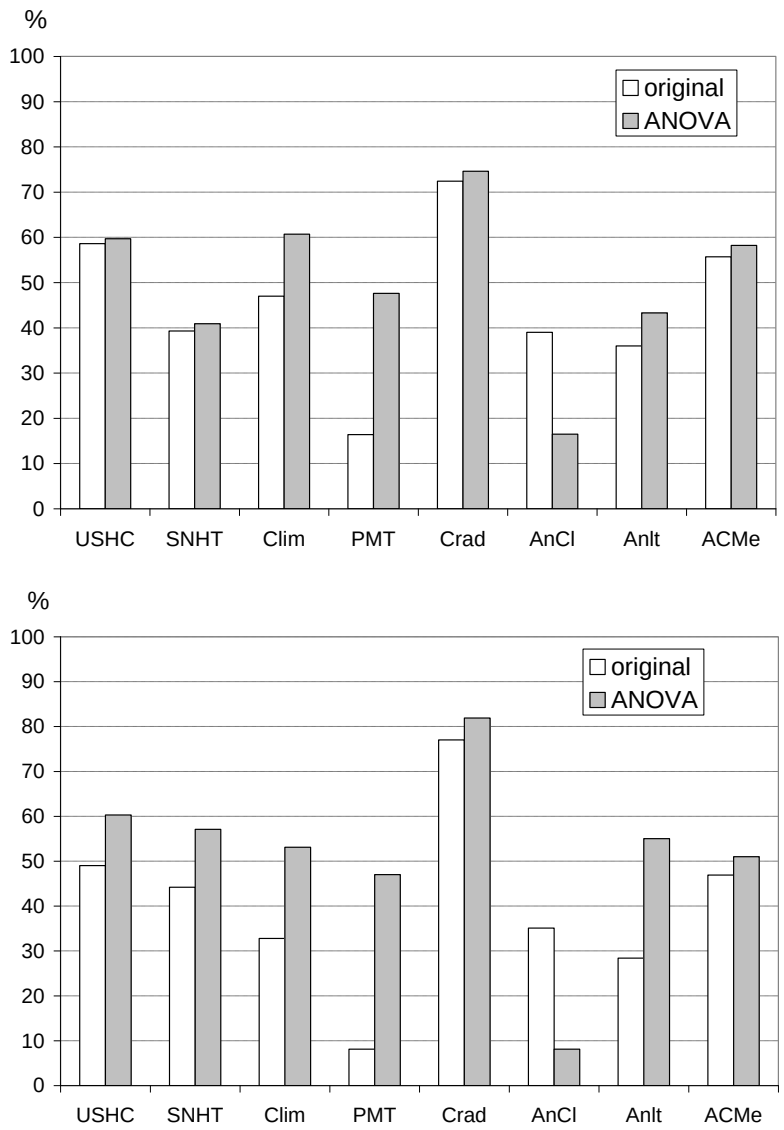


Figure 7. Efficiencies of various homogenisation methods in the Benchmark experiments with and without ANOVA application. a) (upper) CRMSE of annual values (E_{CA}) b) (bottom) RMSE of individual trends (E_T)

It can be seen that the application of ANOVA generally raises the efficiencies. The improvement is bigger in E_T than in E_{CA}. Interestingly, the ANOVA is even beneficially for the already very accurate contribution Crad, and this combination shows the highest efficiencies (except for E_{NS}, see later) for all the HMs examined in this study.

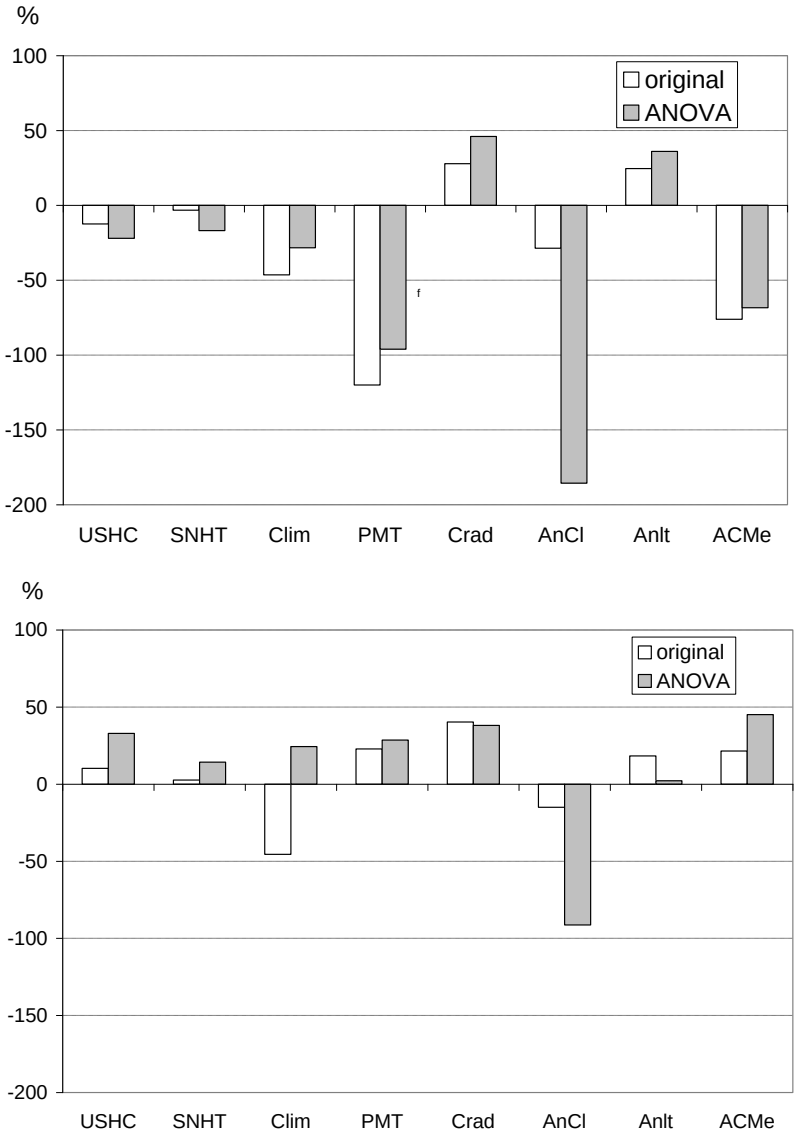


Figure 8. The same as fig. 7, but for RMSE of network-mean trends a) (upper) 1900-1999 (E_{NL}), b) (bottom) 1950-1999 (E_{NS})

The general picture for network-mean trends is very different than for annual CRMSE and individual trends (Fig. 8), i.e. the effect of ANOVA is sometimes negative, particularly for E_{NL}. Note that the E_{NL} values are often negative both without ANOVA and together with that. These results tend to show that when original HMs yield negative efficiencies, the ANOVA might not be able to do improvement, perhaps due to the detection errors on which the network-mean trends are more sensitive than the annual CRMSE and individual trends.

4. DISCUSSION AND CONCLUSIONS

One important finding of this study is that apart from some poor detection methods the main source of the inaccuracy in homogenised time series is often not the detection part of HMs. In a complex homogenisation procedure, the time series comparison, the treatment of missing data and outliers, and the calculation of adjustment factors all together influence the final performance. For long the central question in homogenisation was the search for more

accurate detection methods, while relatively little attention was paid to the influence of other parts of HMs. These results suggest that simple detection examinations, such as G1 are not suitable to reveal the true properties of complete homogenisation methods.

Another notable finding is that efficiencies are more often negative than it was thought earlier (Peterson et al. 1998, Auer et al. 2005, etc.). That overestimation was based on the fact that in the homogenisation results of real datasets the common bias of trends in networks does not appear at all, it can be seen only when the evaluation is made on simulated data with exactly known properties of artificially set inhomogeneities (Venema et al., 2011). On the other hand we note that in the true world the use of metadata may substantially help in achieving effective homogenisation results, while metadata was not simulated for the Benchmark homogenisation.

The reconstruction of network-mean trends, that is a crucially important task in climatology, seems to be more sensitive to the weak points of HMs. In the examination of the Benchmark homogenisation the network-mean trends for 1900-1999 often have larger errors after homogenisation than in the raw data. Note that the lengths of time series in the Benchmark are different, and only 3 time series start from 1900 in each network, while the other time series start later, and the low number of comparable time series significantly increases the uncertainty of the homogenisation results. One might think that to reconstruct the network-mean trends with these conditions is an unfairly hard task, but in fact, climatologists often have to face with similar problems, so the experience of negative efficiencies should be taken seriously. Other characteristics than the E_{NL} rarely show negative efficiencies in this study, and for the best HMs (ACMA, Crad, PROD, MASH) all the efficiencies are positive. However, one has to take into account at this point that real datasets could be less favourably for achieving good performances than the Benchmark.

The difference between Dataset B and Dataset C indicates that the Benchmark is likely easier for HMs than the true datasets (see more discussions about this in Domonkos, 2011a). Even if the Benchmark represents well a true class of real datasets, there are surely many others for which the homogenisation task is an even bigger challenge. For keeping low the chance of negative efficiency and its effect on climate variability analyses we have two main recommendations. First, moderately effective but popular HMs should be replaced with HMs of the best performance in climatological studies. Second, more benchmarking studies are needed, and ones with particular attention to the search of threshold conditions until which the statistical homogenisation is beneficial would be essentially useful.

Not correcting some detected breaks may well sometimes lead to more accurate data. To explain more the last idea, we mention that an indication of inhomogeneity e.g. by a homogeneity test does not mean necessarily that homogenisation-adjustments will provide more reliable and more accurate data in comparison with the raw data. Note that this idea is not new, and in the USHCN it has already been applied (Menne and Williams, 2009).

The examinations of ANOVA with the detection results of various HMs show that the ANOVA often produces higher efficiencies than the original results of the examined HMs. To understand the success of ANOVA it has to be noted that in ANOVA all inhomogeneity-caused biases and their mutual influences are treated together in one equation system. By contrast, when the correction-factors are computed for every break individually, errors, for instance due to undetected IH, may accumulate.

A tentative conclusion of G3 experiments could be that the ANOVA should be included in HMs that have not included ANOVA until now. However, some examples show that the performance of ANOVA is poor when the quality of detection results is insufficient. Therefore our final conclusion is that HMs with effective detection method and correction method together should be applied in the homogenisation of climatic time series. The SNHT and its modern versions (PMT, Climatol, USHCN) may function quite well in certain cases, but both their detection parts and correction parts are poorer than that of the best HMs. The same also refers to other HMs that are based on the single change-point detection and corrections to one or more reference series or reference sections. According to our present knowledge the best HMs are the ACMANT, Craddock, Craddock + ANOVA, PRODIGE and MASH. Note that the performance of Craddock is strongly user-dependent, and this HM cannot be tested in large datasets. Our conclusions refer primarily to the homogenisation of temperature datasets.

Acknowledgements

The research was supported by the European projects COST ES0601, EURO4M FP7-SPACE-2009-1 /242093 and by the Spanish project "Tourism and Environment" ECO 2010-18158.

References

Alexandersson, H. 1986: A homogeneity test applied to precipitation data. J. Climatol. 6, 661-675.

Alexandersson, H. and Moberg, A. 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. Int. J. Climatol. 17, 25-34.

Auer, I., Böhm, R. and 23 coauthors, 2005: A new instrumental precipitation dataset for the greater Alpine region for the period 1800-2002. Int. J. Climatol. 25, 139-166. DOI: 10.1002/joc.1135.

Buishand, T.A. 1982: Some methods for testing the homogeneity of rainfall records. J. Hydrology 58, 11-27.

Caussinus, H, and Lyazrhi, F. 1997: Choosing a linear model with a random number of change-points and outliers. Ann. Inst. Statist. Math. 49/4, 761-775.

Caussinus, H. and Mestre, O. 2004: Detection and correction of artificial shifts in climate series. J. Roy. Stat. Soc. Series C 53, 405-425.

Craddock, J.M. 1979: Methods of comparing annual rainfall records for climatic purposes. Weather 34: 332-346.

Domonkos, P. 2008: Testing of homogenisation methods: purposes, tools and problems of implementation. Proceedings of the 5th Seminar and Quality Control in Climatological Databases. (Ed. Lakatos, M., Szentimrey, T., Bihari, Z. and Szalai, S.),WCDMP-No. 71, WMO/TD-NO. 1493, 126-145.

Domonkos, P. 2011a: Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods. Theor. Appl. Climatol., 105, 455-467, DOI: 10.1007/s00704-011-0399-7.

Domonkos, P. 2011b: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). Int. J. Geosci., 2, 293-309, DOI: 10.4236/ijg.2011.23032.

Ducré-Robitaille, J-F., Vincent, L.A. and Boulet, G. 2003: Comparison of techniques for detection of discontinuities in temperature series. Int. J. Climatol. 23, 1087-1101. DOI: 10.1002/joc.924.

Easterling, D.R. and Peterson, T.C. 1995: A new method for detecting undocumented discontinuities in climatological time series. Int. J. Climatol. 15, 369-377.

Guijarro, J.A. 2011: User's guide to climatol. An R contributed package for homogenization of climatological series. Report, State Meteorological Agency, Balearic Islands Office, Spain, http://webs.ono.com/climatol/climatol.html, 2011.

Menne, M.J. and Williams Jr, C.N. 2009: Homogenization of temperature series via pairwise comparisons. J. Climate, 22, 1700-1717. DOI: 10.1175/2008JCLI2263.1.

Peterson, T.C., Easterling, D.E. and 19 co-authors, 1998: Homogeneity adjustments of in situ atmospheric climate data: a review. Int. J. Climatol. 18, 1493-1517.

Štěpánek, P., Zahradniéek, P. and Skalák, P. 2009: Data quality control and homogenization of the air temperature and precipitation series in the Czech Republic in the period 1961–2007, Adv. Sci. Res. 3, 23–26.

Szentimrey, T. 1999: Multiple Analysis of Series for Homogenization (MASH). Second Seminar for Homogenization of Surface Climatological Data. WCDMP 41, WMO-TD 962, WMO, Geneva, 27-46.

Venema, V., Mestre, O. and the COST HOME team, 2012: Benchmarking monthly homogenization algorithms. Climate of the Past, in press.
Venema, V. 2011: ftp://ftp.meteo.uni-bonn.de/pub/victor/costhome/homogenized_monthly_benchmark.
Vincent, L.A. 1998: A technique for the identification of inhomogeneities in Canadian temperature series. J. Climate 11, 1094-1104.
Wang, X. L., Wen, Q.H. and Wu, Y. 2007: Penalized maximal *t* test for detecting undocumented mean change in climate data series. J. Appl. Meteor. Climatol. 46 (No. 6), 916-931. DOI:10.1175/JAM2504.1
Wilcoxon, F. 1945: Individual comparisons by ranking methods. Biometrics Bull., 1, 80-83.

ACMANT: WHY IS IT EFFICIENT?

Peter Domonkos
Center for Climate Change, Univ. Rovira i Virgili, Campus Terres de l'Ebre, Tortosa, Spain
peter.domonkos@urv.cat

Abstract

In the COST ES0601 project (COST HOME) a benchmark temperature and precipitation dataset (Benchmark) was developed for assessing the efficiencies of homogenisation methods via blind test experiments. In homogenising temperature data, the best performance was achieved with ACMANT. However, ACMANT was developed with the use of the Benchmark, thus its observed performance is not fully comparable with the blind tests results of other methods. Consequently, our knowledge about the true performance of ACMANT is limited. This study includes a brief analysis of the theoretical properties of ACMANT, as well as presents new experimental results. The aim of the study is to provide more information for the objective evaluation of ACMANT.

1. INTRODUCTION

ACMANT (Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series) is a recently developed homogenisation method. It is applicable for monthly temperature series. In the blind test experiments of COST HOME an early version of ACMANT was tested. The early version produced outstanding results regarding the RMSE of monthly values, but it was rather poor in reconstructing true climatic trends (Venema et al. 2012). Later the ACMANTv1.2 (http://www.c3.urv.cat/members/pdomonkos.html) was developed, it is referred as ACMANT late in Venema et al. (2012) and referred as ACMANT in this study.

The efficiency results that are obtained with homogenising Benchmark could be affected by the fact that Benchmark was used in the development of the new method. The treatment of seasonal changes in inhomogeneity caused biases (hereafter: station effects) is particularly criticised from the point of view that the properties of true observed data might considerably differ from the model that applied both in Benchmark and by ACMANT. Unfortunately, at present there is no opportunity to perform new blind tests for ACMANT that are comparative with other homogenisation methods. In this paper the arguments and evidences are intended to be collected in order to obtain the evaluation of ACMANT as objective as possible. The organisation of the paper deviates somewhat from the traditional form. In the next section the theoretical properties of ACMANT will be discussed, then a distinct section will be devoted to analyse the seasonal changes of station effects in temperature series. Studies about real data homogenisation, as well as a test experiment with data of uni-seasonal station effects are discussed there. In section 4 more test results with ACMANT will be presented and analysed, while the last (fifth) section is for synthesising the results and drawing the conclusions.

2. DESCRIPTION OF ACMANT

2.1. General characterisation

In this section a brief description of ACMANT is provided, the full description can be found in Domonkos (2011a). ACMANT is a fully automatic homogenisation method what means that after inputting the raw data and some characteristics of the network (number of stations, length of time series, etc.) the execution does not need human assistance as far as the final homogenised data are produced by the software.

The most important characteristics of ACMANT are a) The harmonisation of examinations in different time-scales (i.e. in annual and monthly scales), b) The use of the optimal segmentation and Caussinus-Lyazrhi criterion in the detection of inhomogeneities, c) The use of ANOVA for the final corrections of inhomogeneities.

ACMANT has four main parts.

I) Preparation. This part contains initial calculations (anomalies, spatial correlations, etc.), outlier filtering, as well as filling the data gaps caused by missing data and outliers.

II) Pre-homogenisation. The purpose of pre-homogenisation is to filter the largest errors from the reference-composites of the final homogenisation. In the pre-homogenisation temporary adjustments are applied to reduce the station effects. Then outlier-filtering and interpolations are performed again using the improved data.

III) Homogenisation. First the long-term biases are searched in annual scale (Main Detection), then with further calculations the timings of the change-points are determined in monthly scale. The remaining station effects are checked on monthly scale with Secondary Detection. Once the detection is finished, ANOVA is applied for calculating the correction terms. ANOVA also provides the final calculations for filling the gaps caused by missing data and outliers.

IV) Final adjustments. In this step change-points with insignificant shift-sizes are excluded from the list of change-points, and ANOVA is applied again with the reduced set of change-points.

2.2. Selected segments of ACMANT

a) *Building reference series from composites*

ACMANT performs relative homogenisation. It means that before homogenisation relative time series are derived from the original series. A relative time series (**T**) is the arithmetic difference between the candidate series (**A**) and a reference series (**F**). The traditional way of creating reference series (Peterson and Easterling, 1994) is applied with a specific parameterisation in ACMANT. For a candidate series (g) all the other series in network (s = 1…S) are used as reference composites when the spatial correlation (r) with g exceeds a preset threshold. More precisely, r stands for the spatial correlation between the first difference (increment) series. The reference-composites are weighted according to the squared correlations (r^2).

$$\mathbf{F_{g[j_1,j_2]}} = \frac{\sum_{s=1}^{S} r_{g,s}^2 \mathbf{A_{s[j_1,j_2]}}}{\sum_{s=1}^{S} r_{g,s}^2} \tag{1}$$

In eq. 1 [j1,j2] represents a section between years *j*1 and *j*2, i.e. eq. 1 can be applied to any period of the time series. In ACMANT the threshold *r* is 0.4, but for at least two composites *r* ≥ 0.5 is expected. Note that these thresholds are low relative to some recommendations (Alexandersson and Moberg, 1997, Auer et al. 2005, etc.).

b) <u>*Creation of multiple relative time series*</u>
ACMANT uses multiple relative time series, because the number of available reference composites often varies according to the sections of the time series. The minimum length of **T** series is 30 years. Three **T** series are always constructed: (i) with the highest possible Σr^2, (ii) with the earliest starting year, (iii) with the latest ending year. However, as criterions (i), (ii) and (iii) might be satisfied by the same series, the true number of **T** series can be lower than 3. On the other hand, more than three **T** series can be involved according to the changes in Σr^2 for different sections of the candidate series (Domonkos, 2011a).

c) <u>*Main Detection*</u>
Detection of inhomogeneities is always performed one-by-one for different candidate series in ACMANT. Optimal step-function is fitted to two annual variables, namely to the annual mean (**TM**) and amplitude of seasonal cycle (**TD**), and the common change-points of them are searched. The minimum distance between two change-points is 3 years. In other respects Main Detection is the same as the detection of PRODIGE (Caussinus and Mestre, 2004).

Eqs. 2 and 3 show the calculation of *tm* and *td* for year *j*.

$$tm_j = \frac{\sum_{m=1}^{12} t_{j,m}}{12}, \qquad td_j = \frac{t_{j,5} + t_{j,6} + t_{j,7} + 0.5t_{j,8} - t_{j,11} - t_{j,12} - t_{j,1} - 0.5t_{j,2}}{3.5} \tag{2), (3}$$

Then the optimal segmentation of an *L* year long period into *K* + 1 segments is given by eq. 4.

$$\min_{[j_1, j_2, \ldots j_K]} \left\{ \sum_{k=0}^{K} \sum_{i=j_k+1}^{j_{k+1}} (tm_i - \overline{\mathbf{TM_k}})^2 + c_0^2 (td_i - \overline{\mathbf{TD_k}})^2 \right\} \tag{4}$$

$$j_0 = 0, \qquad j_{K+1} = L$$

Upper stroke denotes the time average for segment *k*, c_0 is constant, its value in ACMANTv1.2 is 1.414.

The number of segments is optimised by the Caussinus-Lyazrhi criterion (eq. 5, Caussinus and Lyazrhi, 1997).

$$\ln\left\{1-\frac{\sum_{k=0}^{K}(j_{k+1}-j_k)\cdot\left[(\overline{\mathbf{TM_k}}-\overline{\mathbf{TM}})^2+c_0^2(\overline{\mathbf{TD_k}}-\overline{\mathbf{TD}})^2\right]}{\sum_{i=1}^{L}(tm_i-\overline{\mathbf{TM}})^2+c_0^2(td_i-\overline{\mathbf{TD}})^2}\right\}+\frac{2K}{L-1}\ln(L) \tag{5}$$

d) *Precision of the timings of detected IHs*

Main Detection works on relatively long time-scale. After Main Detection more precise timings of change-points are searched applying 48-month symmetric windows around the pre-estimated timings of the change-points. Note that within such a window only 1 change-point was detected, since the minimum length of segments is 3 years in Main Detection. In a window, two-phase harmonic functions are fitted to the values, and the optimum fitting is searched. Phase-change in one of the 25 central months is accepted only. The timing of the phase-change in the optimum fitting is the final timing of the detected change-point.

e) *Secondary Detection*

When after the adjustments that applied according to Main Detection results, accumulated anomalies still exceed some predefined thresholds, Secondary Detection is applied.

In Secondary Detection 60-month long sub-series of monthly values around the maximum of accumulated anomalies are examined. At this step the optimal segmentation is applied for the time averages of monthly values, but the section-means are substituted with harmonic functions of annual cycle for sections of minimum 10 months, and the number of segments is maximised by 3 for a sub-series.

f) *Correction: ANOVA*

In ACMANT, ANOVA is applied for calculating the final correction-terms. This procedure minimises the standard deviation of the homogenised data. It can be proved that ANOVA provides the optimal estimation of correction terms when the following two conditions exist: i) the climate signal is the same in the network, ii) the station effect is constant between two adjacent known change-points (Caussinus and Mestre, 2004).

g) *Pre-homogenisation*

In ACMANT each time series are pre-homogenised in a way that in the calculation of the adjustment-terms for series *s*, series *g* is excluded from the process, when *s* is prepared to be a reference-composite in the final segmentation of series *g*. As *s* usually takes part in the homogenisation of all the other time series of the network, usually *N*-1 different pre-homogenisations are performed for an individual *s* in a network of *N* stations.

First the order of the candidate series is set, it is from the series of estimated poorest quality to the series of estimated best homogeneity. The determination of the order is based on the estimation of maximal station effects. This step necessarily contains some arbitrary criterions (Domonkos, 2011a).

During the pre-homogenisation ANOVA is not applied, because the repetition of ANOVA would overuse the spatial connections among data. In the pre-homogenisation temporary adjustment terms are applied. These terms are calculated with the help of unified relative time series.

h) *Temporary adjustments: Unified relative time series*
A unified relative time series has the following properties: i) It covers the whole period for which the homogenisation of the candidate series can be performed, ii) It includes the relative time series unchanged for which Σr^2 is the highest (it is often shorter than the whole period), it is the principal section, iii) The principal section is completed with other relative time series to cover the whole period defined by i), the latter series are complementary series, iv) The complementary series are adjusted before completion in a way that systematic biases due to differences of the spatial means of station effects for differing subsets of stations are intended to be elaborated. With other words, as the regional average of all stations (N) does not equal to the average of some subsets of Q stations ($Q < N$), this source of bias has to be treated in the creation of unified relative time series. In the early version of ACMANT that was tested in the blind test experiments with the Benchmark (Venema et al. 2012), the main source of trend-errors with ACMANT was the lack of harmonisation among relative time series.

2.3. Evaluation of the theoretical properties of ACMANT

I) Factors explaining the high efficiency

- Optimal segmentation with the Caussinus-Lyazrhi criterion. Earlier efficiency tests of detection methods proved that this method performs best among the inhomogeneity-detection methods used in climatology (Domonkos, 2008, 2011b). The joint segmentation (Picard et al., 2011) could have similar or even better performance, but it has not been proved yet with tests.
- Harmonisation of the examinations of different time-scales in the detection process. The signal to noise ratio is higher in low time resolution than in high resolution, but to find the precise timings and the occurrences of short-term biases the examinations in high (monthly) resolution is also necessary. This kind of harmonisation is unique in ACMANT. A particularly valuable novelty is the bivariate detection that yields results in monthly scale in spite of the examinations are made in annual scale. Note that the latter is also criticised sometimes, because the true annual cycles of station effects are unknown.
- The application of ANOVA makes the correction terms to be as accurate as possible. See also Domonkos et al. (2012).
- Pre-homogenisation is applied in a way that multiple use of the same spatial connection is not allowed, thus the accumulation of errors due to the repeated inclusion of a noise-term or a non-revealed inhomogeneity is excluded.
- Outlier filtering and gap filling are repeated using data of higher and higher quality during the homogenisation procedure.

II) Weak points and doubts around ACMANT

- ACMANT was tested with Benchmark, and in Benchmark the station effects have harmonic annual cycles with maximum biases in winter and summer, thus this feature of Benchmark favours to ACMANT. It could be questioned how the Benchmark-model is applicable for real data with respect to the annual cycle of station effects. See its analysis in Sect. 3.
- ACMANT uses some arbitrary parameters. These parameters were set with the help of Benchmark-experiments, thus the performance of ACMANT on datasets of markedly different properties from Benchmark is unclear. See more discussion about it in Sect. 4.2.
- The use of unified relative time series for calculating adjustment-factors may be suboptimal, since the use of homogeneous sections in pairwise comparisons could likely produce more accurate results. However, it is a challenge to find a good automatic

subprogram of pairwise comparisons (although such subprogram has already been created by Menne and Williams, 2009). Note that the unified relative time series are applied only for temporary adjustments, thus the undesired effect to the final efficiency is supposed to be little if any.

- The use of multiple comparison might be better in the detection part than the use of one reference series from composites. It is a question that cannot be decided in a theoretical way, and the results of the Benchmark experiment are also inadequate to clarify this point, since the effect of the chosen method in time series comparison cannot be examined separately from the other characteristics of homogenisation methods.

3. THE SEASONAL CYCLE OF STATION EFFECTS AND ITS IMPACT ON THE PERFORMANCE OF ACMANT

ACMANT presumes harmonic annual cycle of station effects with extremes of biases in mid-summer and mid-winter, therefore the adequacy of Benchmark to this respect has to be evaluated.

In Benchmark the station effects have annual cycles of definitely harmonic shape and modes in winter and summer. The size of the maximum deviation has a standard normal distribution with 0 expected value and 0.4 °C standard deviation. Although the modes are always in winter and summer, the phases still have substantial variation: the modes can be in any month of summer and winter, and they more often occur in the beginning or ending months of seasons than in the middle month (Venema et al. 2012).

In studies of the homogenisation of true observational temperature series more inhomogeneities were reported for summer series than for winter series (Moberg and Alexandersson, 1997; Drogue et al. 2005; Domonkos, 2006, Domonkos and Štěpánek 2009). Moberg and Alexandersson (1997) pointed on the main cause of this seasonal difference: changes in radiation-effects due to technical changes of the temperature observations are larger in summer than in winter. Czech and Hungarian temperature series were homogenized with 16 homogenization methods (Domonkos and Štěpánek 2009), and the results show that both the frequency and the magnitude of station effects are the smallest in winter and the largest in summer. On the other hand, Moberg and Alexandersson (1997) found the largest temperature-shifts in winter (but the highest shift-frequency in summer), Štěpánek reported the largest station effect for September (personal information, 2010), and according to some assessments, the mean magnitude of seasonal cycles is larger in Benchmark than in true observational series (Venema et al. 2012). Note that synoptic climatological factors might cause annual cycle of station effects in other way than with summer and winter modes.

Considering that the detection results provide only estimations of the true properties, and the seasonal cycles of Benchmark have some non-natural irregularity, it is hard to estimate if Benchmark or the true observational data favours more ACMANT regarding to the seasonal cycle of station effects included in them. Anyhow, it is useful to know the performance of ACMANT when there is no seasonal cycle of station effects at all. For this reason a special experiment was made with Benchmark: The monthly anomalies (from the monthly mean of a given series) of a given year were randomly re-ordered. The way of reordering was the same for each station-series and for both the homogeneous and inhomogeneous ("raw") data. The re-ordering was changed randomly year-by-year. In this way a dataset was created in which

all the annual values are the same as in Benchmark, but the seasonal cycle of station effects is ceased. The performance of ACMANT ("ACMANTx" in Table 1) was tested in this dataset.

Table 1. Efficiencies in reducing the RMSE / CRMSE errors of the row data applying various homogenisation methods (1 = perfect homogenisation). CRMSE = centered RMSE (Venema et al, 2012), m = monthly, a = annual, t = trend slope, nt1 = network mean trend slope for 1900-1999, nt2 = network mean trend slope 1950-1999, the description of the homogenisation methods is in Venema et al. (2011)

Method	CRMSE(m)	CRMSE(a)	RMSE(t)	RMSE(nt1)	RMSE(nt2)
ACMANTv1.2	0.563	0.728	0.768	0.382	0.544
ACMANTx	0.412	0.662	0.739	0.272	0.634
PRODIGE main	0.416	0.676	0.724	0.042	0.436
PRODIGE trendy	0.413	0.679	0.728	0.031	0.435
PRODIGE monthly	0.431	0.693	0.735	0.093	0.426
MASH main	0.398	0.667	0.706	0.315	0.441
Craddock Vertacnik	0.461	0.724	0.770	0.278	0.403
USHCN 52x	0.382	0.586	0.490	-0.124	0.102

In accordance with the expectations, the efficiencies are lower than with the true Benchmark, but the decline turned out to be not too large. The observed efficiencies are in the range of the efficiencies of PRODIGE and MASH with the true Benchmark, and the estimation of network-mean trend bias is even better. The efficiency for reconstructing the network-mean trends of 1950-1999 has increased relative to the original experiment with ACMANT, but it is likely a random effect due to the small sample size. The highest decline of efficiency occurred with the CRMSE of monthly values (from 0.56 to 0.41) but the 0.41 is still not lower than the efficiency of the other best methods, except Craddock-test. Note that the observed efficiencies of the Craddock-test are based on a partial contribution (i.e. 7 networks were homogenised from the available 15), therefore they are not fully appropriate for making comparisons with the results of full Benchmark experiments.

When comparisons are made between the performance of ACMANT and the performance of other best homogenisation methods in this study, the "other best methods" are composed by PRODIGE main, PRODIGE trendy, PRODIGE monthly and MASH main. Although the group of the best methods is wider, it includes also Craddock-test, USHCN and the newly developed HOMER of the COST HOME team, good test results of full Benchmark experiments are available only for PRODIGE and MASH. There are full contributions also with USHCN, but the USHCN belongs to the best methods for other reason than the observed efficiencies (that are not too high), i.e. the USHCN has a stably low false alarm rate, which is an important positive feature, but out of the analyses of this study.

4. FURTHER TEST EXPERIMENTS WITH ACMANT

4.1. Tests with the "Big Benchmark"

Big Benchmark is another (bigger) test dataset than the official Benchmark, and its creator is also Victor Venema. The original reason of its creation was to test the fully automatic USHCN software with a test dataset that is similar to the true observational temperature dataset in the United States. The statistical properties of Big Benchmark are similar to those of Benchmark with two important exceptions: (i) The Big Benchmark is much bigger than Benchmark, it contains 200 surrogated and 200 synthetic network, (ii) In Big Benchmark networks may contain 9 or 15 time series, but never only 5 time series. Note that the Benchmark surrogated temperature dataset consists of 15 networks, and 9 of them contains only 5 time series. For this reason, unfortunately, test results with Big Benchmark are not directly comparable with the Benchmark result, although they still can be interesting.

ACMANT was subdued to a blind test with the first 100 surrogated networks of Big Benchmark. The efficiency turned out to be higher than with the official Benchmark (Table 2), likely due to the denser networks of Big Benchmark. For making fairer comparisons, characteristics for subsets of fixed network-sizes are also shown. Note however, that the number of networks in the official Benchmark with 9 (15) time series is only 4 (2), therefore the opportunity to draw profound conclusions from these results is very limited. Due to the small sample-size in Benchmark, network-mean errors have not been calculated for subsets of fixed network-sizes.

Table 2. Efficiency characteristics with ACMANT for the official Benchmark (OB) and Big Benchmark (BB). nt = network, denotations in the headline are the same as in Table 1

Test-data	CRMSE(m)	CRMSE(a)	RMSE(t)	RMSE(nt1)	RMSE(nt2)
OB, 9 series / nt	56.5	72.1	74.2		
BB, 9 series / nt	56.2	74.1	71.6		
OB, 15 series /nt	58.4	78.0	84.0		
BB, 15 series /nt	62.2	80.9	85.0		
OB, all networks	56.3	72.8	76.8	38.2	54.4
BB, all networks	60.1	78.5	79.6	40.5	55.5

The results with the Big Benchmark experiment indicate that the performance of ACMANT is stably high for dense networks of data with high spatial correlations. The observable efficiency characteristics here are higher than those were achieved with the other best homogenisation methods during the Benchmark experiment, but for making correct comparisons with them, further common blind test experiments would be needed.

4.2. ACMANT with changing parameters

The second unit of the Caussinus-Lyazrhi criterion (sect. 2.2, eq. 5) is a penalty-term. Its effect is that higher number of change-points (K) is allowed only when the fitting of the step function to the data becomes substantially better with the increase of K. In the present experiment this penalty-term is supplied with coefficient p (eq. 6), and it is varied between 1 and 7 in the pre-homogenisation part of the procedure. (In the final homogenisation p always equals 1.)

$$\ln\left\{1-\frac{\sum_{k=0}^{K}(j_{k+1}-j_k)\cdot\left[(\overline{\mathbf{TM_k}}-\overline{\mathbf{TM}})^2+c_0^2(\overline{\mathbf{TD_k}}-\overline{\mathbf{TD}})^2\right]}{\sum_{i=1}^{L}(tm_i-\overline{\mathbf{TM}})^2+c_0^2(td_i-\overline{\mathbf{TD}})^2}\right\}+p\frac{2K}{L-1}\ln(L) \qquad (6)$$

The effect of increasing p is that only the most significant inhomogeneities can be filtered out during the pre-homogenisation. As the aim of the pre-homogenisation is just to eliminate the largest biases of the raw data, the application of $p > 1$ in the pre-homogenisation might result in higher efficiency than the basic version. The results of the experiment are shown in Fig. 1. In the figure the mean efficiencies for the other best methods are also shown for making comparisons.

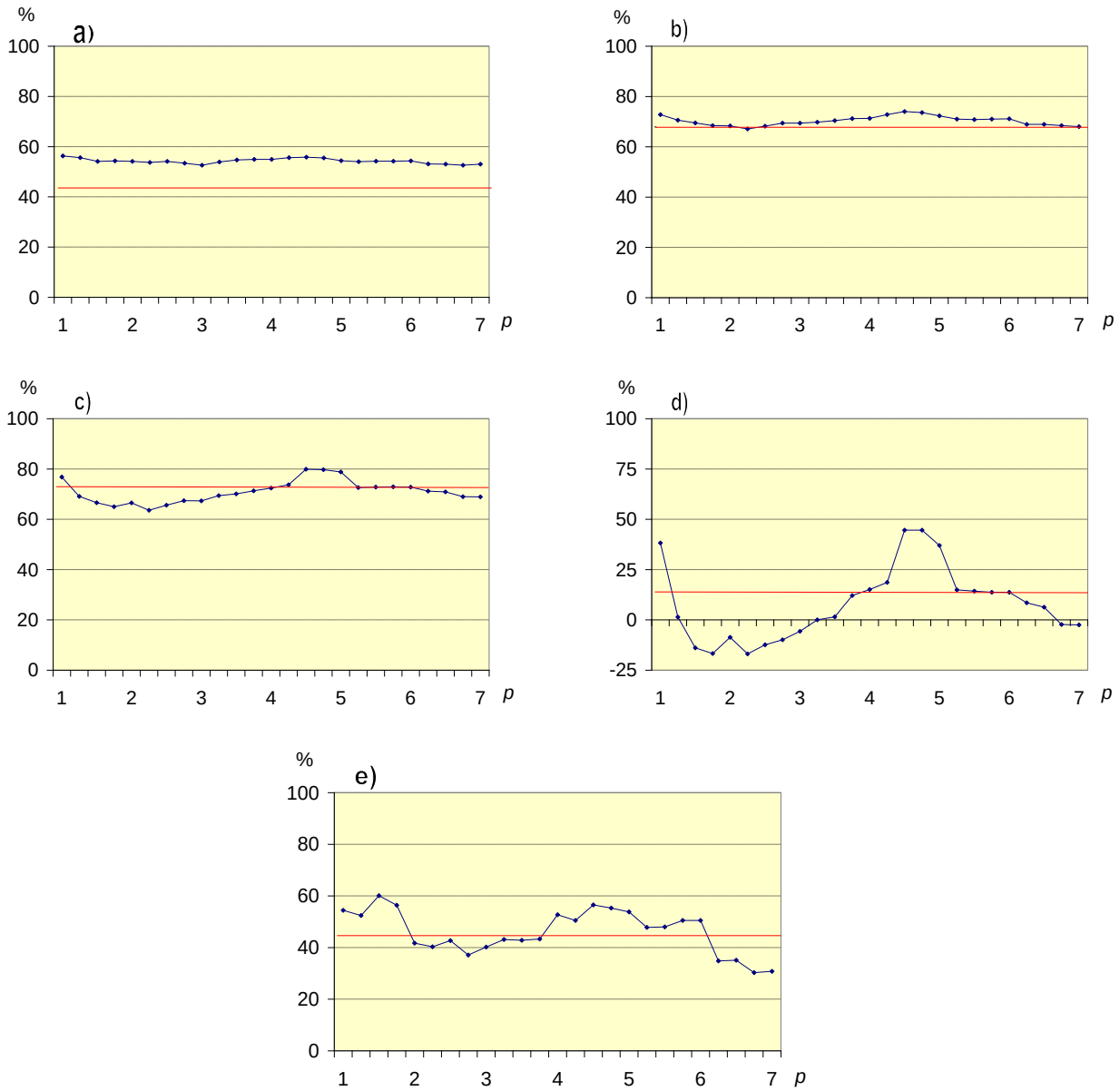


Figure 1. Efficiency in reducing the RMSE / CRMSE error of raw data in function of *p*. a) CRMSE monthly, b) CRMSE annual, c) RMSE of individual trends, d) RMSE of network-mean trends for 1900-1999, e) RMSE of network-mean trends for 1950-1999. Red horizontal lines represent the mean efficiency of full experiments with PRODIGE and MASH during the Benchmark homogenisation

Two maximums appear in function of *p*, the first is at *p* of 1 – 1.5, while the other is at *p* of 4.5 – 5.0. I cannot explain this double optimum, even do not know whether it would be similar for other datasets, or it is a peculiarity of Benchmark, more precisely the examined 15 networks of the surrogated temperature dataset of Benchmark. The absolute maximum of observed efficiency is $p = 1$ for the monthly CRMSE, $p = 1.5$ for the network-mean bias for 1950-1999, while for the other efficiency measures the absolute maximums are with $p = 4.5$. The variation of efficiency is small in Fig. 1a and 1b, moderately large in Fig. 1c, while quite large for network-mean biases (Fig. 1d and 1e). The efficiency of ACMANT is slightly higher than that of PRODIGE and MASH in annual CRMSE, substantially higher than that of PRODIGE and MASH in monthly CRMSE, while for efficiencies in trend-bias reduction this relation is parameter-dependent.

Finally I note that I made another experiment varying another parameter of ACMANT, i.e. the exponent of the denominator in eq. 42 of Domonkos (2011a), but in that case the observed variation of efficiency was much less than in the presented case.

5. DISCUSSION AND CONCLUSIONS

The study of Venema et al (2012) contains several statements about the performances of the methods participated in the Benchmark experiment. Now two of them about the late ACMANT are quoted here, for discussing if we have better understanding after the examinations presented. The two statements are (i) "ACMANT late contribution suggests that ACMANT is currently the most accurate method available", and (ii) "ACMANT late is optimized based on the benchmark data itself. It is thus not clear how much of this performance would be realised in an application to a real dataset."

First it has to be made clear that the projection of the observed efficiency results to the application to real data has limitation due to the differences between the surrogated data and real data that obviously exist in spite of the effort has been made to have the surrogate data similar to the real data. However, this limitation is not specific for ACMANT. An exception could be the harmonic annual cycle of station effects, which is exploited more intensively by ACMANT than by any other homogenisation method. However, the analyses of Sect. 3 proved that its effect is minor if any, in raising artificially the performance of ACMANT.

A more serious problem is that 15 networks of data is not very much either to find an optimum parameterisation or to achieve an accurate validation. It is because the within network errors are often interdependent and the distribution of the degree of errors is non-normal, but rather exponential. For illustrating the latter, a brief statistic of the network-mean biases for 1900-1999 in the Big Benchmark experiment is presented here. The absolute value of the bias was below 0.5°C in 94 cases (from the examined 100), in four cases the bias was between 0.51 and 0.63 (°C), but the highest two biases were 0.83°C and 1.20°C. The likely explanation is that rare unfavourable coincidence of change-points in different time series, missing data, as well as unfavourable interference with unusually high noise may result in large homogenisation errors even with such a sophisticated method as ACMANT. This fact limits the opportunity to draw final conclusions from the experiments with the 15 surrogated temperature networks, and Big Benchmark was examined only with ACMANT.

Both the Big Benchmark experiment and the parameterised examination showed that the performance of ACMANT is more stable with respect to the reduction of monthly and annual

CRMSE errors than in the reduction of trend-biases. The decrease of monthly CRMSE is spectacularly greater with ACMANT than with any other homogenisation method, and this difference is not sensitive to the chosen set of parameters of ACMANT. The performance of ACMANT in this characteristic may fall onto the level of the other best homogenisation methods only when the annual cycle of biases is entirely removed from the test datasets, which is, however, an unrealistic condition for the observed temperature data of mid- or high geographical latitudes. This good result with ACMANT is a consequence of the sophisticated treatment of different time-scales (i.e. multi-annual, annual and monthly) from which only one piece is the bivariate detection with two annual variables in the Main Detection segment.

The main conclusions are as follows:

- All the examinations confirm that ACMANT belongs to the family of the best homogenisation methods (i. e. PRODIGE, MASH, Craddock-test, HOMER and USHCN).
- ACMANT is particularly effective in reducing the RMSE of monthly temperature data. When dense networks with high spatial correlations are treated, this favourable characteristic of ACMANT is very stable. This characteristic of ACMANT might have importance in the future in providing input data for daily data homogenisation.
- The performance in trend-bias reduction is more parameter-dependent than in the reduction of RMSE errors. The examinations presented do not seem to be sufficient to find the optimal parameterisation of ACMANT, thus further analyses are needed.

Acknowledgements

The research was supported by the European projects COST ES0601 and EURO4M FP7-SPACE-2009-1/242093 and by the Spanish project “Tourism, Politics and Environment” ECO 2010-18158.

References

Alexandersson, H. and Moberg, A. 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. Int. J. Climatol. 17, 25-34.

Auer, I., Böhm, R. and coauthors, 2005: A new instrumental precipitation dataset for the greater Alpine region for the period 1800-2002. Int. J. Climatol. 25, 139-166. DOI: 10.1002/joc.1135.

Caussinus, H. and Lyazrhi, F. 1997: Choosing a linear model with a random number of change-points and outliers. Ann. Inst. Statist. Math. 49/4, 761-775.

Caussinus, H. and Mestre, O. 2004: Detection and correction of artificial shifts in climate series. J. Roy. Stat. Soc. Series C 53, 405-425.

Domonkos, P. 2006: Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation. Időjárás, 110, 63-87.

Domonkos, P. 2008: Testing of homogenisation methods: purposes, tools and problems of implementation. Proceedings of the 5th Seminar and Quality Control in Climatological Databases. (Ed. Lakatos, M., Szentimrey, T., Bihari, Z. and Szalai, S.),WCDMP-No. 71, WMO/TD-NO. 1493, 126-145.

Domonkos, P. 2011a: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). Int. J. Geosci., 2, 293-309, DOI: 10.4236/ijg.2011.23032.

Domonkos, P. 2011b: Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods. Theor. Appl. Climatol., 105, 455-467, DOI: 10.1007/s00704-011-0399-7.

Domonkos, P. and Štěpánek, P. 2009: Statistical characteristics of detectable inhomogeneities in observed meteorological time series. Studia Geophys. and Geod. 53, 239-260, doi: 10.1007/s11200-009-0015-9.

Domonkos, P., Venema, V. and Mestre, O. 2012: Efficiencies of homogenisation methods: Our present knowledge
and its limitation. In: Proceedings of the 7th Seminar and Quality Control in Climatological Databases.

Drogue, G., Mestre, O., Hoffmann, L., Iffly, J-F. and Pfister, L. 2005: Recent warming in a small region with semi-oceanic climate, 1949-1998: what is the ground truth? Theor. Appl. Climatol. 81, 1-10, doi: 10.1007/s00704 004-0088-x.

Menne, M.J. and Williams Jr, C.N. 2009: Homogenization of temperature series via pairwise comparisons. J. Climate, 22, 1700-1717. DOI: 10.1175/2008JCLI2263.1.

Moberg, A. and Alexandersson, H. 1997: Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. Int. J. Climatol. 17, 35-54.

Peterson, T.C. and Easterling, D.R. 1994: Creation of homogeneous composite climatological reference series. Int. J. Climatol. 14, 671-679.

Picard, F., Lebarbier, E., Budinska, E. and Robin, S. 2011: Joint segmentation of multivariate Gaussian processes using mixed linear models, Computational Statistics and Data Analysis, 55, 1160-1170.

Venema, V., Mestre, O. and the COST HOME team, 2011: Description of the COST-HOME monthly benchmark dataset and the submitted homogenized contributions. Report, Meteorological Institute, University of Bonn, Germany, available at: http://www2.meteo.uni-bonn.de/venema/articles/2011/report home.pdf, 2011.

Venema, V., Mestre, O. and the COST HOME team, 2012: Benchmarking monthly homogenization algorithms. Climate of the Past, 8, 89-115, doi:10.5194/cp-8-89-2012, 2012.

THEORETICAL QUESTIONS OF DAILY DATA HOMOGENIZATION AND DAILY BENCHMARK RESULTS FOR MASH

Tamás Szentimrey
Hungarian Meteorological Service, P.O. Box 38, H-1525 Budapest, Hungary
szentimrey.t@met.hu

1. INTRODUCTION

In the practice there are numerous methods and software for daily data homogenization but in general without any exact theoretical, mathematical basis. The essence of the arising theoretical, methodological questions can be formulated as follows.

- It is necessary to clarify that the homogenization or correction of data series is a distribution problem instead of a regression procedure.
- Such correction methods and validation characteristics should be applied which are in accordance with the above mentioned aim of homogenization.
- The methods applied in the practice also have to be examined theoretically in such a respect. For example there is the question what is the effect of the regression based methods for the higher order moments of distribution.

2. RELATION OF DAILY AND MONTHLY HOMOGENIZATION

2.1 The general structure of daily data homogenization

If we have daily data series the general way of homogenization is,

- calculation of monthly series,
- homogenization of monthly series taking advantage of the larger signal to noise ratio,
- homogenization of daily series using the detected monthly inhomogeneities.

So we have the question how can we use the valuable information of detected monthly inhomogeneities for the daily data homogenization?

2.2 A popular procedure e.g. the variable correction methods

The typical steps of the procedure are as follows.

1. Homogenization of monthly series:
   - Break points detection, correction in the first moment (mean).
   - Assumption: homogeneity in higher order moments (e.g. st. deviation).

2. Homogenization of daily series:
   There is a trial to homogenize also in higher order moments.
   The used monthly information are only the detected break points.

However the following questions are arising at this procedure:

- Is it adequate model that we have inhomogeneity in higher moments only at daily series but not at monthly ones? Can this model be accepted according to the probability theory?

- Why are not used the monthly correction factors for daily homogenization? It seems to lose some valuable information obtained during the monthly homogenization.

2.3 Problem of inhomogeneity in the higher order moments

There is a common assumption that the correction in mean is sufficient for monthly and annual series, and that the correction of higher order moments is necessary only in the case of daily data series. In general, it is tacitly assumed that the averaging is capable to filter out the inhomogeneities in the higher order moments. However, this assumption is false, for example, if there is an inhomogeneity in the standard deviation of daily data, we may have the same inhomogeneity in monthly data.

Proof.:

Daily data are $X(t)\ (t=1,2,..,30)$, monthly average is $\overline{X}=\frac{1}{30}\sum_{t=1}^{30}X(t)$.

Let us introduce an inhomogeneity in the standard deviation for the daily data:

$$X_{ih}(t)=\alpha\cdot\left(X(t)-\mathrm{E}\left(X(t)\right)\right)+\mathrm{E}\left(X(t)\right),\ \ (t=1,2,..,30).$$

The expected value is unchanged: $\mathrm{E}\left(X_{ih}(t)\right)=\mathrm{E}\left(X(t)\right)$, but the standard deviation has changed: $\mathrm{D}\left(X_{ih}(t)\right)=\alpha\cdot\mathrm{D}\left(X(t)\right)$.

Let us see the new monthly average: $\overline{X}_{ih}=\frac{1}{30}\sum_{t=1}^{30}X_{ih}(t)$.

The expected value is unchanged: $\mathrm{E}\left(\overline{X}_{ih}\right)=\mathrm{E}\left(\overline{X}\right)$, but the standard deviation changed with the same measure:

$$\mathrm{D}\left(\overline{X}_{ih}\right)=\mathrm{D}\left(\frac{1}{30}\sum_{t=1}^{30}X_{ih}(t)\right)=\mathrm{D}\left(\frac{1}{30}\sum_{t=1}^{30}\alpha\cdot X(t)\right)=\alpha\cdot\mathrm{D}\left(\frac{1}{30}\sum_{t=1}^{30}X(t)\right)=\alpha\cdot\mathrm{D}\left(\overline{X}\right).$$

For illustration an example is presented on Figure 1. The example series is from the daily benchmark dataset of COST Action HOME: network\000111\hotmd00011636d station.

On the figure the annual mean of the daily values can be seen. By the figure the generated higher order daily inhomogeneities appear also at the annual values! That means the averaging is not capable to filter out the inhomogeneities in the higher order moments.

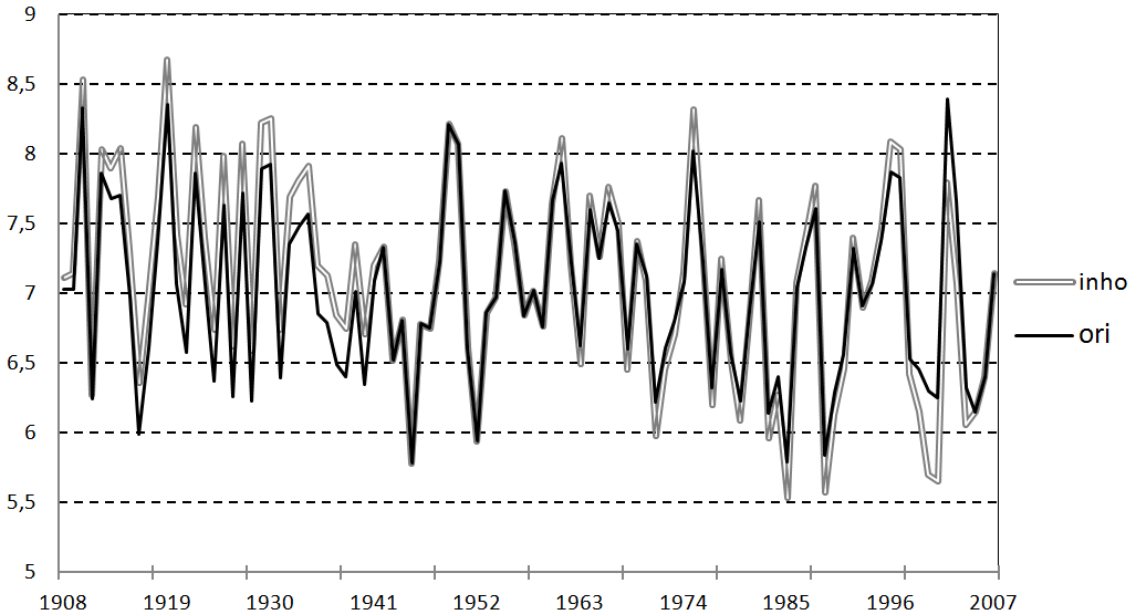


Figure 1. Annual mean series of the original and inhomogeneous daily series from the benchmark dataset of COST HOME: network\000111\hotmd00011636d station

3. DAILY BENCHMARK RESULTS FOR MASH

A daily "benchmark" (validation) dataset was created within COST Action ES0601 HOME in order to evaluate the homogenization methods used for daily data. Altogether 12 networks of daily data series with known breaks were generated where one network included 5 stations with surrogate daily temperature series for period 1908-2007. The inhomogeneities of the validation dataset were generated by switching distribution of stations (to simulate relocation - only for pairs), changing mean and changing higher moments, created for the length of 100 years with 5 breaks in each of the series, with no trends, no outliers, no missing data, no annual cycle (seasonality) of the breaks. For the purpose of the validation of methods 6 networks (30 stations) were used. The MASH method (Multiple Analysis of Series for Homogenization; Szentimrey, 1999, 2008, 2011) was tested for all these networks. Opposite to the fact that the daily series were generated with inhomogeneities also in higher moments and MASH was developed for homogenization in mean (first order) we thought instructive to perform the test procedure. For evaluation of the test results the RMSE and centered RMSE statistics were calculated for the difference series:

$Z_{INHO}(t) = X_{INHO}(t) - X_{ORI}(t)$, $Z_{HOM}(t) = X_{HOM}(t) - X_{ORI}(t)$, where

$X_{ORI}(t)$: original homogeneous series, $X_{INHO}(t)$: inhomogeneous series,

$X_{HOM}(t)$: homogenized series. The result statistics are in Table 1.

Table 1. The test statistics for evaluation of MASH

	RMSE(Z_{INHO})	RMSE(Z_{HOM})	CRMSE(Z_{INHO})	CRMSE(Z_{HOM})
1				
2	0.619	0.680	0.615	0.588
3	0.527	0.360	0.441	0.359
4	0.361	0.243	0.316	0.243
5	0.603	0.235	0.319	0.204
6	0.672	0.330	0.669	0.325
7	0.622	0.263	0.567	0.262
8	0.380	0.372	0.372	0.359
9	0.811	0.446	0.771	0.364
10	0.494	0.276	0.493	0.237
11	0.514	0.395	0.507	0.395
12	0.737	0.346	0.727	0.316
13	0.409	0.329	0.392	0.263
14	0.732	0.357	0.623	0.356
15	0.657	0.451	0.655	0.451
16	0.230	0.111	0.223	0.106
17	0.611	0.542	0.602	0.488
18	1.088	0.393	1.068	0.387
19	0.515	0.501	0.514	0.422
20	0.659	0.216	0.627	0.216
21	0.793	0.314	0.778	0.275
22	0.454	0.345	0.445	0.333
23	0.599	0.388	0.599	0.383
24	0.496	0.497	0.490	0.491
25	0.332	0.150	0.329	0.148
26	0.663	0.305	0.608	0.303
27	0.494	0.321	0.440	0.317
28	0.548	0.266	0.548	0.236
29	0.525	0.286	0.521	0.228
30	0.670	0.399	0.663	0.351
Mean	0.545	0.478	0.486	0.447
	0.579	**0.353**	**0.547**	**0.328**

4. EXAMINATION OF PARALLEL MEASUREMENTS

The basic assumption of the variable correction methods for daily series is that the correction in mean is sufficient for monthly and annual series, but in case of daily data series, the corrections should vary according to the meteorological situation of each day in order to represent the extremes. This idea was published in the paper by *Trewin* and *Trevitt* (1996), where parallel measurements were examined and compared to each other. Since then on the basis of the ideas formulated in the paper, a number of variable correction methods have been developed with the declared aim of being capable of correcting the daily data not only in mean (first moment) but also in the higher order moments. For example, we mention the following methods: higher order moments (HOM) method by *Della-Marta* and *Wanner* (2006) and spline daily homogenization (SPLIDHOM) method by *Mestre et al.* (2011), and there are numerous other similar methods applied in practice. However, in our humble

opinion, during the examinations only some physical experiences were considered without any exact theoretical, mathematical formulation of the problem. The empiric interpretation and formulation seem to be a misunderstanding. Moreover, there are some mathematical statements at the description of the methods – e. g., capability to correct the higher order moments – but without any proof, and this practice is of course contrary to the mathematical conventions.

4.1 Examinations by Trewin and Trevitt (1996)

First here is a quotation from the paper of *Della-Marta* and *Wanner* (2006): "One of the most robust methods capable of adjusting the higher-order moments of daily temperature data is that of *Trewin* and *Trevitt* (1996)." *Trewin* and *Trevitt* (1996) intended to homogenize daily data series in order to create composite temperature records. The following sentences are from their paper: "It is therefore necessary to make use of climatological records with inhomogeneities, and to develop a means of removing or minimizing the impact of inhomogeneities on these records. One way of doing this is by adjusting all parts of a record to be comparable with some 'reference period'. Standard procedures for such adjustments in mean temperatures have relied on the implicit assumption that, if two neighbouring stations both have homogeneous records over some period of time, the difference in daily maximum (or minimum) temperature between them will be a constant for any day in a given month of the year. This implies that the difference in monthly means will be a constant for that month from year to year." In general it is not true of course, but after some examination of real station data series they obtained the following result: "This is observed at Armidale (P. Burr, pers. comm.), ..,where the difference in minimum temperature between the town centre site used in this study and a second site approximately 2 km to the east, in the outer part of the town, has a mean value of 1.5 to 2 °C , but can increase to 4 °C on cold, clear nights. The assumption that the temperature difference between any two nearby sites is always constant must therefore be questioned."

The above conclusion was all right, but the next conclusion is a little bit surprising for us: "The relationship between the temperature characteristics of the two sites in each pair was examined, with the aim of determining an appropriate method for use in extrapolating records at one site to records at the other."

Probably here is the origin of the methods that apply varying corrections per days, and at this step a regression or interpolation problem was obtained for homogenization instead of the adequate distribution problem. Three interpolation techniques were considered by *Trewin* and *Trevitt* (1996) namely: the 'traditional' constant-difference approach, the 'regression' method, and the frequency distribution matching. The methods will be detailed in Section 6.1.

4.2 Mathematical examinations of parallel measurements

What was the reason of the development of the variable correction methods? Essentially, an observed phenomenon at the extremes, namely the differences of parallel measurements are larger in case of extremes. In our opinion, this observed phenomenon has a simple and logical reason, and it is superfluous to look for some complicated physical explanation for the inhomogeneity. The simple reason is that the extremes may be expected at different moments in case of parallel measurements, or in other words, there may be systematic biases in rank

order! It is a natural phenomenon, and for illustration a trivial example is presented according to the probability theory.

Example 4.2

Let $Y_1(t) \in N(0,1))$, $Y_2(t) \in N(0,1)$ $(t = 1,2,..,n)$ be standard normally distributed series with expected values $\mathrm{E}(Y_1(t)) = \mathrm{E}(Y_2(t)) = 0$, with standard deviations $\mathrm{D}(Y_1(t)) = \mathrm{D}(Y_2(t)) = 1$, and with correlation between the series $\mathrm{corr}(Y_1(t), Y_2(t)) = \rho$ $(t = 1,2,..,n)$.

Then the mean difference $\mathrm{E}(Y_1(t) - Y_2(t)) = 0$ of course, however, the difference $Y_1(t) - Y_2(t)$ is not independent from the elements $Y_1(t)$, $Y_2(t)$ if $\rho \neq 1$, and, e.g., the conditional expectation of difference $Y_1(t) - Y_2(t)$ given $Y_1(t)$, or equivalently the regression of difference $Y_1(t) - Y_2(t)$ on $Y_1(t)$ is $\mathrm{E}(Y_1(t) - Y_2(t) \,|\, Y_1(t)) = (1 - \rho) \cdot Y_1(t)$.

Consequently, the difference $Y_1(t) - Y_2(t)$ is an expectedly monotonous increasing function of $Y_1(t)$ if $\rho \neq 1$. This is the theory, but it can be demonstrated in practice too. We generated such standard normal series by the Monte Carlo method with parameters $\rho = 0.9$, $n = 1000$. In this case, $\mathrm{E}(Y_1(t) - Y_2(t) \,|\, Y_1(t)) = 0.1 \cdot Y_1(t)$ and the difference series $Y_1(t) - Y_2(t)$ as a function of series $Y_1(t)$ is plotted in Fig. 2.

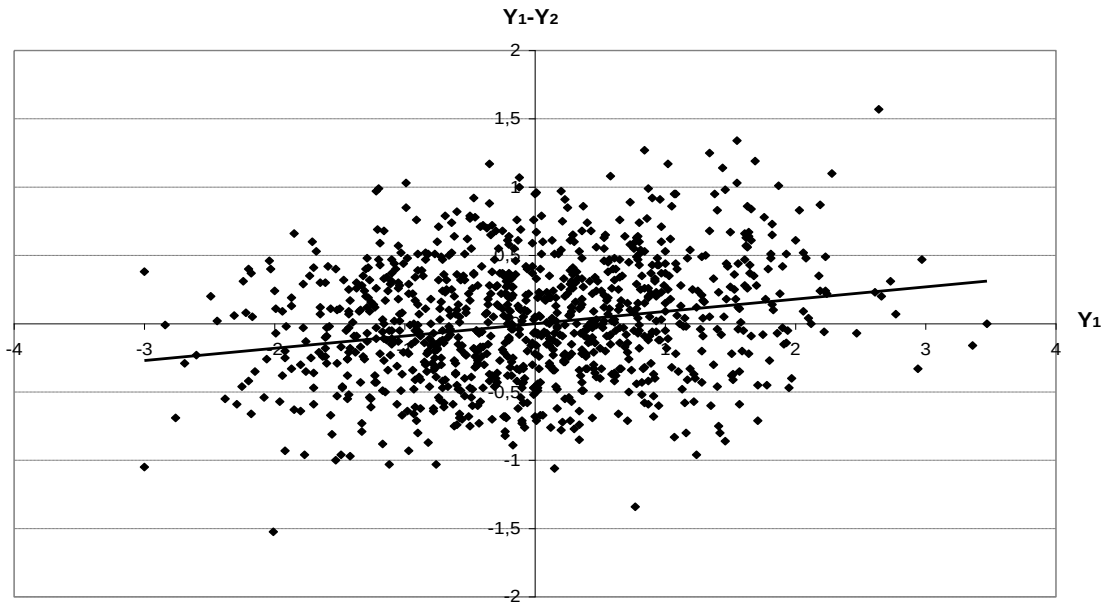


Figure 2. Difference series $Y_1(t) - Y_2(t)$ as a function of series $Y_1(t)$

It is evident that the conditional expectation of difference $Y_1(t) - Y_2(t)$ is monotonous increasing function of $Y_1(t)$, consequently the difference may be larger mainly in the case of extreme values. It is a general phenomenon not only observed for meteorological measurements. Presumably this experience is the reason for the idea that the correction of daily data should vary according to the meteorological situation of each day, in particular on the basis of some regression models. But it is a misunderstanding of the homogenization problem.

5. MATHEMATICAL FORMULATION OF THE DAILY DATA HOMOGENIZATION

Unfortunately, the exact theoretical, mathematical formulation of the problem of homogenization is generally neglected in meteorological studies. Therefore, we try to formulate this problem in accordance with mathematical conventions. First of all it is necessary to emphasize that homogenization is a distribution problem and not a regression one.

Notation
Let us assume we have daily data series:
$Y_1(t)\ (t=1,2,..,n)$: candidate time series of the new observing system.
$Y_2(t)\ (t=1,2,..,n)$: candidate time series of the old observing system.
$1 \le T < n$: change-point, series $Y_2(t)\ (t=1,2,..,T)$ can be used before and series $Y_1(t)\ (t=T+1,..,n)$ can be used after the change-point.

Definition
The aim of homogenization is the adjustment or correction of values $Y_2(t)\ (t=1,2,..,T)$ in order to have the corrected values $Y_{1,2h}(t)\ (t=1,2,..,T)$ with the same distribution as the elements of series $Y_1(t)\ (t=1,2,..,T)$, i.e.:

$$\mathrm{P}(Y_{1,2h}(t) < y) = \mathrm{P}(Y_1(t) < y), \quad y \in (-\infty, \infty)\ ,\ t=1,2,..,T\,. \tag{1}$$

Eq. (1) means the equality in distribution: $Y_{1,2h}(t) \overset{d}{=} Y_1(t)\ \ (t=1,2,..,T)$.

Consequence
Within the same climate area, if the variables $Y_1(t), Y_2(t)\ (t=1,2,..,T)$ have identical distribution, i.e., $Y_2(t) \overset{d}{=} Y_1(t)\ \ (t=1,2,..,T)$, then the merged series $Y_2(t)\ (t=1,2,..,T)$, $Y_1(t)$ $(t=T+1,..,n)$ is homogeneous.

Example
Let us assume we have parallel measurements $Y_1(t)$, $Y_2(t)\ (t=1,2,..,n)$ within the same climate area with distance 50 m between the locations. Then, as a consequence of micrometeorological processes, the series are probably different, $Y_2(t) \ne Y_1(t)\ (t=1,2,..,n)$, but they may be equal in distribution, $Y_2(t) \overset{d}{=} Y_1(t)\ \ (t=1,2,..,n)$. In this case, the mixed series $Y_2(t)\ (t=1,2,..,T)$, $Y_1(t)\ (t=T+1,..,n)$ can be taken as a homogeneous series. This mixed series is equivalent with the homogeneous series $Y_1(t)\ (t=1,2,..,n)$ also in respect of the distribution of extremes.

Returning to the general question, we have to see clearly that the aim of homogenization is to correct the distribution of $Y_2(t)$ according to $Y_1(t)$, instead of the estimation or regression of $Y_1(t)$ on $Y_2(t)$! Moreover, the correction of distribution is equivalent in essence with the correction or adjustment of the moments. The aim of the homogenization expressed in k^{th} moments:

$$m_k = \mathrm{E}\left((Y_{1,2h}(t))^k\right) = \mathrm{E}\left((Y_1(t))^k\right) \quad k = 1,2,\ldots \; ; \; t = 1,2,..,T \,, \tag{2}$$

where E is the usual notation of the expected value or mean equivalently.
Some remarkable formulas for the moments:

$$E = m_1 \,, \quad D^2 = m_2 - m_1^2 \tag{3}$$

where *E* denotes the expected value or mean, and *D* denotes the standard deviation.

In practice, numerous methods indicate the capability to correct the higher order moments but without any exact proof.

6. THE VARIABLE CORRECTION METHODS

We return to the methods suggested by *Trewin* and *Trevitt* (1996) which was mentioned in Section 4.1. Essentially, the underlying principles of the variable correction procedures developed later were formulated based on these methods. We do not agree with these principles as explained by our argument in Sections 4.1 and 5, but let us see some details and properties of the mathematical consequences.

6.1 The Trewin and Trevitt (1996) methods for parallel measurements

The short description is cited word for word again from the paper of *Della-Marta* and *Wanner* (2006):

"Trewin and Trevitt (1996) present three different methods to build a composite daily temperature series. Essential to the methods is the existence of simultaneous (in time) observations from the *new* and *old* observing system. These parallel measurements had been taken based on the recommendations of Karl et al. (1995), who suggest that a minimum of a 2-yr overlap between the new and old observing systems be made. In Australia, for example, this practice has only become routine since around 1993 and so many inhomogeneities needed to be adjusted using the traditional constant difference techniques with neighboring reference stations. In this way, Trewin (2001) created a homogenized daily temperature dataset that has subsequently been used by Collins et al. (2000) to assess trends in the frequency of extreme temperature events in Australia.

The three methods they intercompared were constant difference, linear regression, and frequency distribution matching.

The constant difference approach simply adjusted the older data with the newer data using the mean of the daily differences in the simultaneous (parallel) measurements.

The linear regression method fitted a linear model to the difference in daily simultaneous measurements between the two observing systems and the temperature at the older station. This model could then be used to adjust daily temperatures at the older station differentially depending on the temperature, thereby adjusting the higher-order moments.

Their third method determines the frequency distribution of each site during the simultaneous measurement period. The adjustment for each desired percentile is calculated simply as the difference between the two percentiles. This method assumes that there is no systematic bias in the rank order of the temperatures at the two sites.

They show that both the regression method and the frequency distribution matching technique have certain advantages; however, if the homogenization of extreme events is most needed, then their frequency distribution matching technique is more accurate."

Our mathematical comments to the methods are as follows.

6.1.1 Constant difference approach

Yes, this approach is correct if the inhomogeneity is in mean or expected value or first moment, which are the same with different names.

6.1.2 Linear regression method

This procedure is absolutely wrong for homogenization. To demonstrate the problem, a trivial counter-example is presented.

Theorem
Let us assume that the different series $Y_1(t)$, $Y_2(t)$ $(t=1,2,..,n)$ have identical distribution, with expected values $\mathrm{E}(Y_1(t))=\mathrm{E}(Y_2(t))=0$, standard deviations $\mathrm{D}(Y_1(t))=\mathrm{D}(Y_2(t))=1$, and correlation between the series $\mathrm{corr}(Y_1(t),Y_2(t))=\rho$ $(t=1,2,..,n)$.

(i) Then the linear regression of difference $Y_1(t)-Y_2(t)$ on $Y_2(t)$ is $(\rho-1)\cdot Y_2(t)$, consequently, the homogenized series after the suggested adjustment, $Y_{1,2h}(t)=Y_2(t)+(\rho-1)\cdot Y_2(t)=\rho\cdot Y_2(t)$ and $\rho\cdot Y_2(t)$ is just the linear regression of $Y_1(t)$ on $Y_2(t)$.

(ii) Moreover, since the expected values $\mathrm{E}(Y_{1,2h}(t))=\mathrm{E}(Y_1(t))=\mathrm{E}(Y_2(t))=0$, therefore – using Eq. (3) – , the second moment of $Y_{1,2h}(t)$ is equal to the variance $\mathrm{D}^2(Y_{1,2h}(t))=\delta^2<1$, while the common second moment of $Y_1(t)$, $Y_2(t)$ is equal to the variances $\mathrm{D}^2(Y_1(t))=\mathrm{D}^2(Y_2(t))=1$. Therefore, the second moment was decreased from 1 to $\delta^2<1$ during the regression.

Summing up, according to (i) this procedure is equivalent with the simple linear regression of $Y_1(t)$ on $Y_2(t)$. Furthermore, according to (ii) the following statement about the method is absolutely false: "This model could then be used to adjust daily temperatures at the older station differentially depending on the temperature, thereby adjusting the higher-order moments." The truth is just the opposite, since the correct second moment was damaged at our counter-example.

6.1.3 Frequency distribution matching technique

The main problem is the following assumption which is the fundament of the method: "This method assumes that there is no systematic bias in the rank order of the temperatures at the two sites."

Unfortunately, the reality and the mathematics are much more complicated, and the above assumption cannot be accepted as it is demonstrated in Fig. 2. The bias in rank order depends on the stochastic connection, and there may be systematic bias, since $Y_1(t)$, $Y_2(t)$ are not monotonous increasing functions of each others. At this method, the adjusted $Y_{1,2h}(t)$ is obtained essentially by a simple exchange $Y_2(t)$ for $Y_1(t)$ according to the rank orders. Why? For example, if $Y_1(t)$, $Y_2(t)$ were equal in distribution then such an exchange would not be necessary.

6.2 The general type of variable correction methods applied in the practice

On the basis of the former principles described in Sections 6.1.2 and 6.1.3 (regression and frequency distribution matching), a number of variable correction methods have been developed during the last years. The new improvement of these methods is that they do not need overlap observations, instead of this they use information from nearby reference stations, for example higher order moments (HOM) method by *Della-Marta* and *Wanner* (2006) and spline daily homogenization (SPLIDHOM) method by *Mestre et al.* (2011). We do not want to criticize the details of these methods however, we express again our skepticism on their common fundamental principles which were based on a pseudo problem demonstrated in *Example 4.2*. Moreover, we repeat the following sources of errors for consideration.

- The assumption of the frequency distribution matching technique, i.e., there is no systematic bias in the rank, cannot be accepted.
- The regression methods are not adequate to correct the higher order moments.

Our last remark is connected also with the higher order moments. In general, the papers about these methods indicate the capability to correct the higher order moments, but this statement is always without any exact mathematical proof. We are skeptic, however if somebody could send us a nice proof, we would be grateful for it.

7. SOME REMARKS ABOUT THE HOMOGENIZATION IN THE HIGHER-ORDER MOMENTS

We suggest considering the following remarks when developing homogenization methods with the capability to correct also the higher order moments.

Remark 1
The correction in the first two moments or, equivalently, in mean and standard deviation can be formulated by using the notations defined in Section 5 as follows:

$$Y_{1,2h}(t) = E_1(t) + \frac{D_1(t)}{D_2(t)}\left(Y_2(t) - E_2(t)\right) \quad \left(t = 1,2,..,T\right), \tag{4}$$

where $E_1(t) = \mathrm{E}\left(Y_1(t)\right)$, $E_2(t) = \mathrm{E}\left(Y_2(t)\right)$ are the means, and $D_1(t) = \mathrm{D}\left(Y_1(t)\right)$, $D_2(t) = \mathrm{D}\left(Y_2(t)\right)$ are the standard deviations. Then $\mathrm{E}\left(Y_{1,2h}(t)\right) = E_1(t)$, $\mathrm{D}\left(Y_{1,2h}(t)\right) = D_1(t)$.

In general, the detection of the change points and the estimation of correction factors are suggested to be based on the examination of monthly data series because of the larger signal to noise ratio.

Remark 2

If the joint distribution of the series is normal, $Y_1(t) \in N\left(E_1(t), D_1(t)\right)$, $Y_2(t) \in N\left(E_2(t), D_2(t)\right)$ $\left(t = 1,2,..,n\right)$ and $Y_{1,2h}(t)$ $\left(t = 1,2,..,T\right)$ is calculated according to Eq. (4), then $Y_{1,2h}(t), Y_1(t)$ $\left(t = 1,2,..,T\right)$ have identical distribution: $Y_{1,2h}(t) \overset{d}{=} Y_1(t)$ $\left(t = 1,2,..,T\right)$. Consequently, the mixed series $Y_{1,2h}(t)$ $\left(t = 1,2,..,T\right)$, $Y_1(t)$ $\left(t = T+1,..,n\right)$ is homogeneous, that means it is sufficient to correct only the first two moments in case of joint normal distribution.

Proof

Owing to Remark 2 and the joint normal distribution, $Y_{1,2h}(t) \in N\left(E_1(t), D_1(t)\right)$ $\left(t = 1,2,..,T\right)$.

8. CONCLUSION

It is necessary to define the exact mathematical theory for homogenization of climate data series. Homogenization is a probability distribution problem, and the methods applied in practice should be theoretically evaluated in this respect.

References

Della-Marta, P.M. and Wanner, H., 2006: A Method for homogenizing the extremes and mean of daily temperature measurements. J. Climate 19, 4179–4197.

Mestre, O., Gruber, C., Prieur, C., Caussinus, H., and Jourdain, S., 2011: Splidhom: a method for homogenization of daily temperature observations. J. Appl. Meteor. Climatol. 50, 2343–2358.

Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH), Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2008: Development of MASH homogenization procedure for daily data, Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary, 2006; WCDMP-No. 68, WMO-TD NO. 1434, 2008, pp. 116-125.

Szentimrey, T., 2011: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, pp. 64.

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

APPLICATION OF THE REDISTRIBUTION METHOD ON CUSTOM DATA SUBSETS

Predrag Petrović
Republic Hydrometeorological Service of Serbia, Kneza Višeslava 66, 11030 Belgrade, Serbia, tel. +381/11/3050-803
predrag.petrovic@hidmet.gov.rs

Abstract

The ReDistribution Method (Petrović, 2003) is independent of the temporal consistency of data series. This feature allows dealing with any data series subset, regardless of the gaps that might be featured in the series.

The ReDistribution Method might be successfully applied to diagnose causes of inhomogeneities more efficiently. Custom subsets might be extracted from the series by any criteria. For the observed wind direction and speed series, these subsets might be made according to wind speed categories.

The given example with three wind speed categories indicates different causes of the inhomogeneities of wind direction series, depending on the analyzed subset category. Detailed analysis of the yielded results might be helpful in estimating causes of the detected inhomogeneities. Homogenization method that should be applied afterwards must take these results into account, which opens new discussions about the most adequate procedure.

1. A BRIEF OVERVIEW OF THE METHOD

The ReDistribution Method is distribution-oriented method for detection of inhomogeneities in data series. The basic principle of the method is a comparison of two consecutive distributions over fixed moving window span. The output value is a ReDistribution Index (RDI) as a measure of variations of distribution in time. This index is calculated as

$$RDI = \frac{N_r}{N}$$

where N_r is the number of redistributed values (as a half sum of redistributed frequencies from all value categories) and N is the total number of values in the distribution (Petrović, 2004). Hence, the values of RDI mathematically range from zero (the "perfect match" of distributions) to unity (change of complete range of featured values without overlaps). Critical values of RDI, however, are the peak values (maximum change of distribution) of a magnitude significantly higher than the noise level (Fig. 1).

1.1 Special features of the method

The ReDistribution Method uses a fixed, user-defined number of observations for calculating distributions inside a moving window span. It is usually a whole number of years, such as 2, 4, or 6 years. It is important to emphasize that only observed values are used for calculations, which preserves possibility to use any desired number of observations. As a consequence, the ReDistribution Method is independent of the temporal consistency of data series.

Also, a mass of data builds distributions, while small gaps, outliers and/or erroneous values are minimized in distributions due to their very low frequency. Such values are diminished into a noise level of the RDI series. This practically means that no strict quality control or filling in missing data are necessary.

Finally, method's ability to run more than one series at a time is very suitable for wind data, since the complete wind data information consists of at least two independent series (direction and speed).

Bearing all these special features in mind, the ReDistribution Method allows dealing with any data series subset, regardless of its temporal inconsistency. Such subsets might be user-defined in order to inspect some special cases and thus detect and diagnose inhomogeneities more efficiently.

2. CUSTOM DATA SUBSETS

Custom data subset is any selection of data that meets the user-defined criteria. For the observed wind direction and speed series, these subsets might be made according to wind speed categories or angular sections (azimuth intervals) of wind direction. An example of three subsets by wind speed criteria is given as follows:

- weak winds, up to 2 m/s; this subset features the impact of instrument sensitivity and introduced instrument friction;
- moderate winds, between 2 m/s and 4 m/s; such subset is under influence of surrounding objects near the instrument;
- strong winds, over 4 m/s; this subset reveals serious wind redirections, since it is influenced by instrument orientation rather than change of general synoptic circulation that produces winds of this category.

Inhomogeneities that are featured in such subsets give an additional information about their probable cause, as described in the given definitions.

Naturally, such subsets might be used for detection of inhomogeneities in wind direction only, since the wind speed range is significantly reduced and thus inconvenient for wind speed homogeneity assessment.

3. EXAMPLE OF WIND SPEED CATEGORY SUBSETS

3.1 Initial settings and standard run of the method

As an example, hourly series of wind direction and speed of Niš for the period 1981-2010 is examined.

Standard run of the ReDistribution Method is made with the complete dataset, with moving window length of three years. The returned RDI series of wind direction and speed shows significant break points (Fig. 1). However, causes of the breaks appears one moving window before the RDI peak (in this example it is three years earlier).

Simultaneous peaks of both RDI series for direction and speed are recognized in 1990, and 2007 which, at present, indicate problems with instrument or station surroundings in 1987 and 2004, since metadata indicate preserved station location. Wind speed breaks are noted in 1984, 1993 and 1999, indicating problems with anemometer in 1981, 1990 and 1996. On the other hand, wind direction breaks might indicate problems with instrument orientation or change of surroundings, which is a seldom case without wind speed breaks.

3.2 Subset runs of the method

The runs of the ReDistribution Method over data subsets defined by given wind speed criteria reveal some more information about the breaks. Initial settings of moving window length are reset to use an average number of records that cover approximately three-year period. This number varies from one subset to another depending on the size of a subset. Thus, weak wind subset had the window length of 5000 records, moderate wind had 3000 records, while strong winds had only 400 records in the moving window data span. These figures approximately correspond to relative frequencies of wind by the defined subsets.

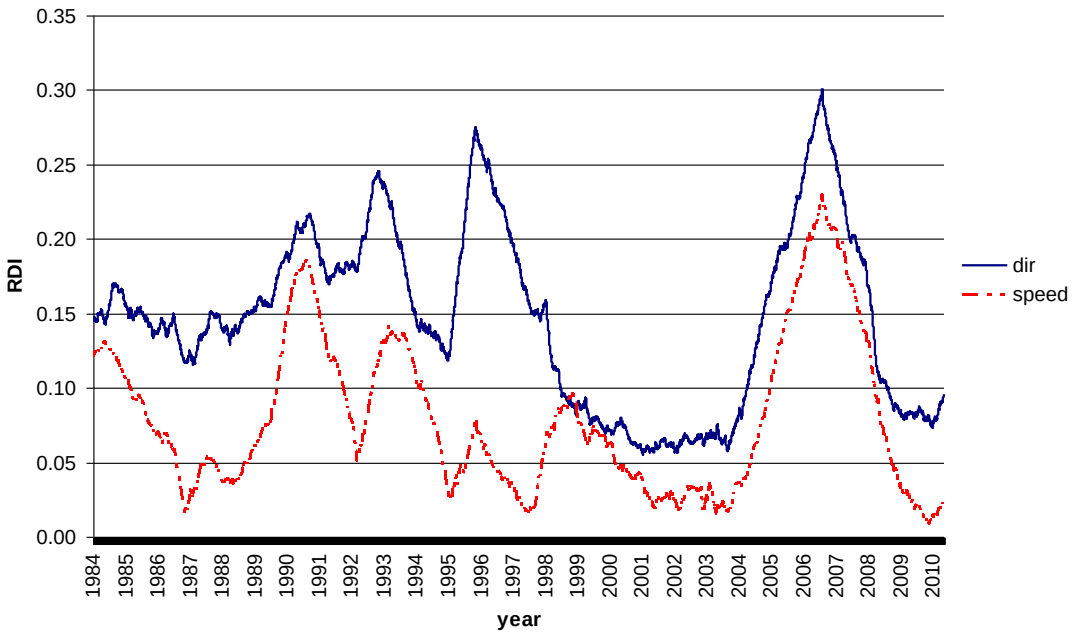


Figure 1. RDI series of wind direction and speed (all data), Niš, 1984-2010

The original graphs are created in number of records' scale, which reveals the exact information about inhomogeneities. However, better comprehension for users might be achieved using the time scale. Although it distorts the original information (due to uneven frequency of wind speeds over a category), it is important to have a proper information about the time of the detected break point, while the magnitude of the change remains intact.

Uneven response of corresponding RDI series is a basic information for break diagnostics.

4. RESULTS

The case of the break point in 2004, indicated by the RDI peak in 2007 is used as an example of the diagnostic part of the method, since it is very clear both in direction and in speed series (Fig. 1). Therefore, the explanations of the results will be given through the time period 2001-2010.

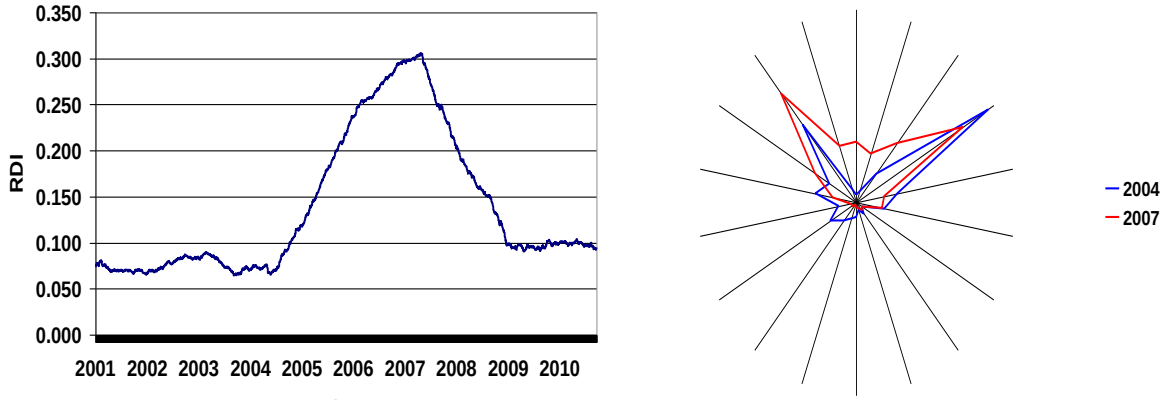


Figure 2. a) RDI series of wind direction for data subset of wind speed below 2 m/s, Niš, 2001-2010; b) wind directions before (in 2004) and after the break point (in 2007)

The run of the ReDistribution Method with weak wind subset indicates very regular and very significant peak of RDI series in 2007, which confirms the originally detected break in 2004 (Fig. 2a). Different noise level before and after the break indicate a change of the instrument. Higher noise level after the break point indicate an instrument with more friction and less sensitivity to the winds. The redirection of the weak winds is quite obvious, introducing more winds throughout angular section between NNW and NE. On the other hand, the most frequent ENE wind has a bit reduced frequency. Finally, winds from the section ESE-W nearly disappeared (Fig. 2b). This disappearance of a significant angular section winds indicate an obstacle which is quite near the instrument.

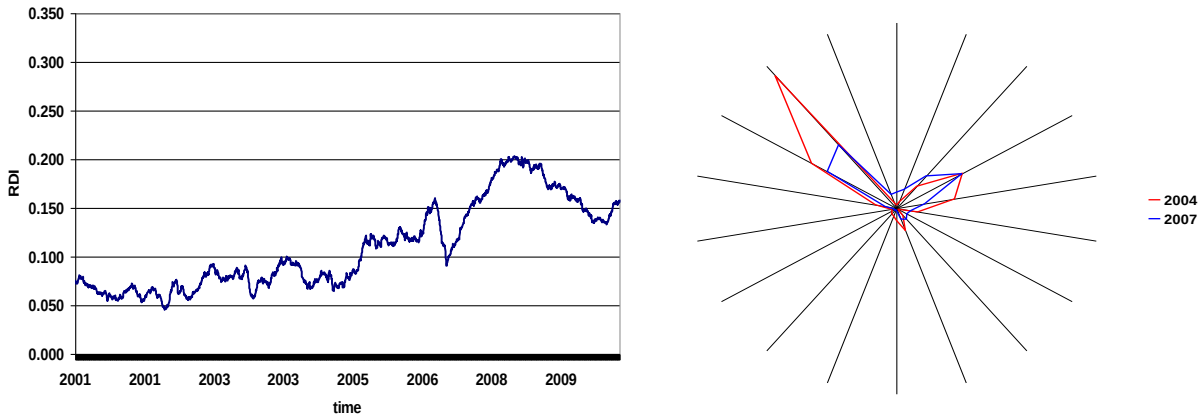


Figure 3. a) RDI series of wind direction for data subset of wind speed between 2 m/s and 4 m/s, Niš, 2001-2010; b) wind directions before (in 2004) and after the break point (in 2007)

Moderate winds had he same detected break point, but with much smaller magnitude (Fig. 3a). The wind redirection mostly affects NW winds, which frequency is reduced to nearly half of the value before the break. Also, some winds around ENE (NE and E) increased their frequencies, which explains the redirection process (Fig. 3b). This redirection is caused by an obstacle that appeared at some ten meters from the instrument, covering the angular section NW-NNE. This finally describes changes of instrument surroundings.

The run of the ReDistribution Method with strong wind subset returned the same break point, but quite diminished and sunk into the noise level (Fig. 4a). Thus, it is hard to detect any homogenity break at this point with certainty, so such RDI series result in no inhomogeneities detected. The directions of these winds keep their distribution, which indicates the preserved station location (Fig. 4b).

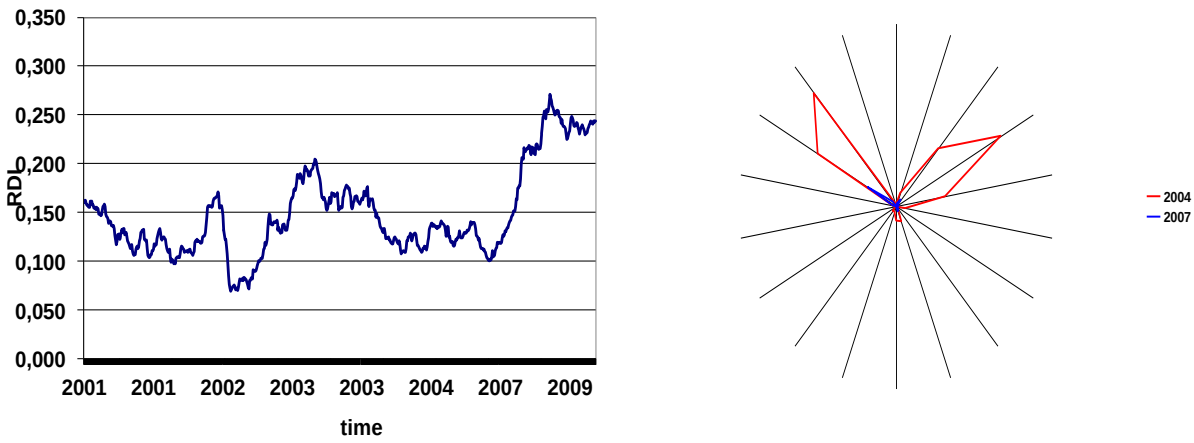


Figure 4. a) RDI series of wind direction for data subset of wind speed over 4 m/s, Niš, 2001-2010; b) wind directions before (in 2004) and after the break point (in 2007)

5. CONCLUSIONS

This variation of the ReDistribution Method introduces new possibilities for break point diagnostics. While standard use of the method returns information about the time and roughly type of cause of detected breaks, assessment of custom data subsets' distributions might give more accurate description of the break points. This is particularly important when a good decision about the possibilities for homogenisation must be made. Besides that, high precision of break detection in temporal scale is preserved.

One of the crucial disadvantages of the method variation is its time consumption. Since the calculation of the RDI series is time consuming itself, even more runs for every subset multiplies required time for work. Also, an user must have a lot of experience with wind data, from the measurement phase through digitalization to the final use of the data series for different assessments.

Nevertheless, this variation of the ReDistribution Method will enhance the knowledge about the wind data series, especially distributions of wind direction as very little used element. Since different causes of the inhomogeneities might require different methods for homogenisation, the final questioning of the right choice for homogenisation method is still active task for the future.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., Wieringa, J., 2003: Guidelines on climate data and homogenization, WMO, WCDMP-No. 53

Petrović, P. 2004: Detecting of inhomogeneities in time series using Real Precision Method, Fourth Seminar For Homogenization And Quality Control In Climatological Databases (Budapest, Hungary, 6-10. October 2003);

Petrović, P. 2007: Detection of Inhomogeneities in Wind Direction and Speed Data, Proceedings Of The Fifth Seminar For Homogenization And Quality Control In Climatological Databases, Budapest, Hungary, 29 May – 2 June 2006

Petrović, P. and M. Curley. 2009: "Detected Inhomogeneities In Wind Direction And Speed Data From Ireland", Proceedings Of The Sixth Seminar For Homogenization And Quality Control In Climatological Databases, WMO, WCDMP-No. 76, pp. 41-52

DETECTION AND CORRECTION OF BREAKPOINTS IN LONG-TERM GERMAN PRECIPITATION SERIES

E. Rustemeier, A. Kapala, V. Venema and C. Simmer
Meteorological Institute, University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany,
elke.rustemeier@uni-bonn.de

Abstract

Based on "original" Prodige homogenization software from Caussinus and Mestre (2004), full automatic detection and correction of artificial break points in climate records is developed, which is indispensable in handling large data sets. For the identification of breakpoints, scaled time series are compared to the other stations within the network. The detected breaks are corrected monthly by means of a multiple linear regression considering the annual cycle. The presented procedure is tested on benchmark data set of the COST-Action ES0601. In comparison to the power (performance) of other homogenization procedures, it is one of the most effective methods.

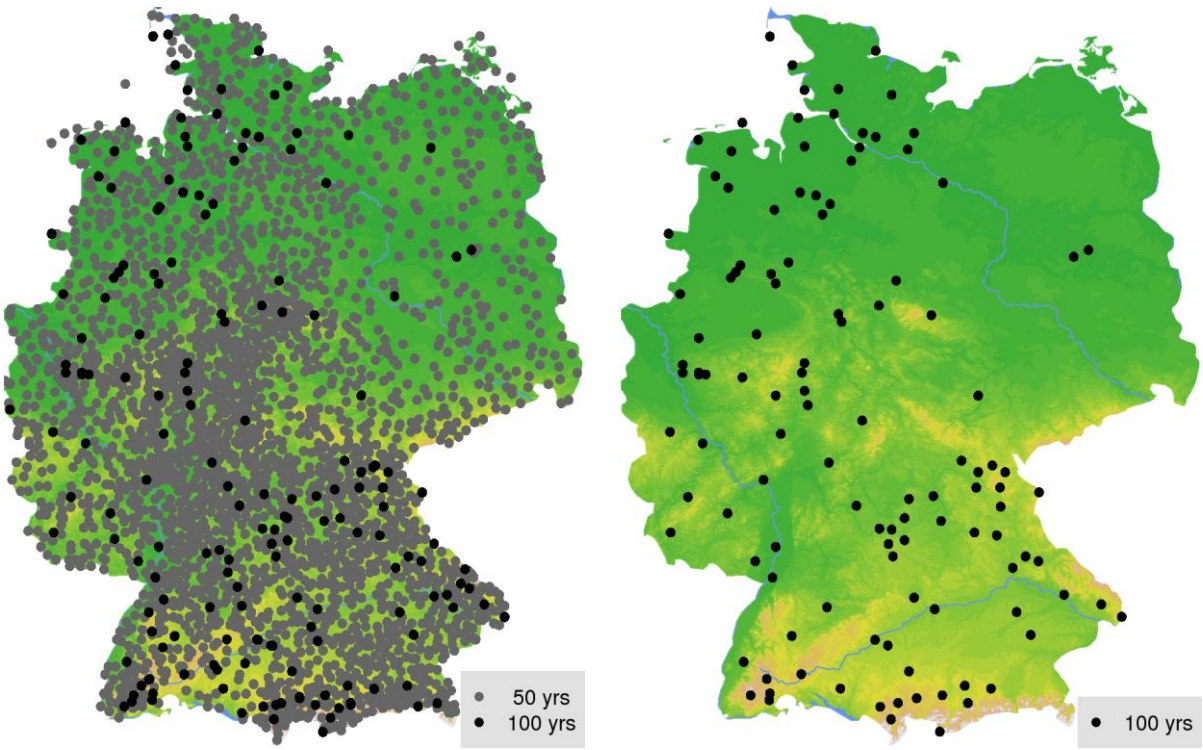
1. INTRODUCTION

It is a well-known fact that meteorological observations, in particular long-term records are inhomogeneous. In the climate change discussion, the need of analyzing precipitation series becomes obvious. Since for a robust trend analysis reliable data is a prerequisite, the data series investigated have to be homogeneous and of high quality. Inhomogeneities in time series may overlie real trends and may lead to wrong conclusions.

Precipitation has a high spatial variability, varying with distance and being strongly effected by topography. For this reason investigation requires a dense network of stations; not only to identify the regional climate, but also to provide enough data allowing the correction of errors in the data series and increasing the overall certainty of the results. Large data sets pose an additional challenge for the homogenization.

The examined data set contains about 2000 long-term precipitation series. In a first step, this study is focused on 118 stations, spanning 100 years of daily data and only few missing values (Fig. 1). The data has been digitized within the cooperative project KLIDADIGI by the Deutscher Wetterdienst and the Meteorological Institute of the University of Bonn.

This paper is dealing with the challenge and the successful development of a fully automatic method. In section 2 the benchmark data set is short described. The development of the method is described in section 3. The validation results are presented in section 4 followed by conclusions and outlook in section 5.



a) 2833 monthly stations (50 yrs) **b) 118 daily stations**

Figure 1. Location of long-term precipitation stations

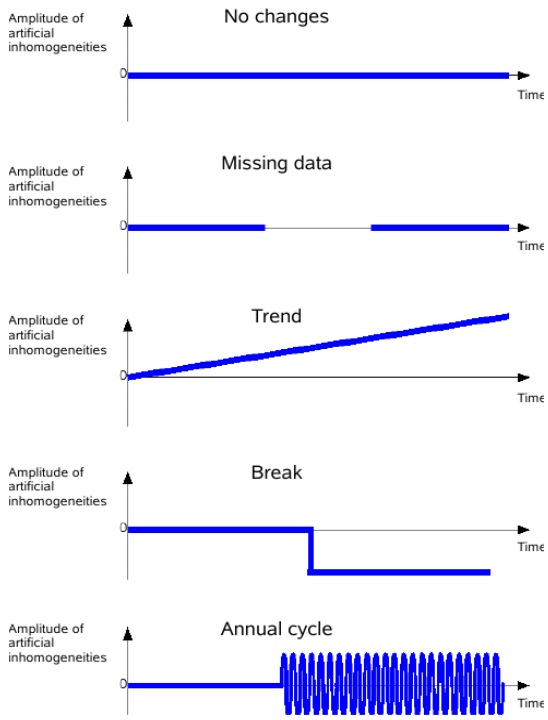
To date, for precipitation only partly manually methods are available, which, although being well known and tested, cannot be used, because of the required time and effort. This paper deals with this challenge and the successful development of an automatic method.

A common and successful approach is the use of meta-data to identify potential inhomogeneities. This approach cannot be used in the present case, since for old time series, station history is almost not available. Many records were lost during the wars. The existing documents are incomplete and only partly digitized. Consequently, meta-data cannot be included in automatic methods, which are necessary for the more than 2000 stations mentioned above. However, it is still essential to use the documentation when available, to verify the detected break points and to validate the applied homogenization method (Fig. 7 Zugspitze).

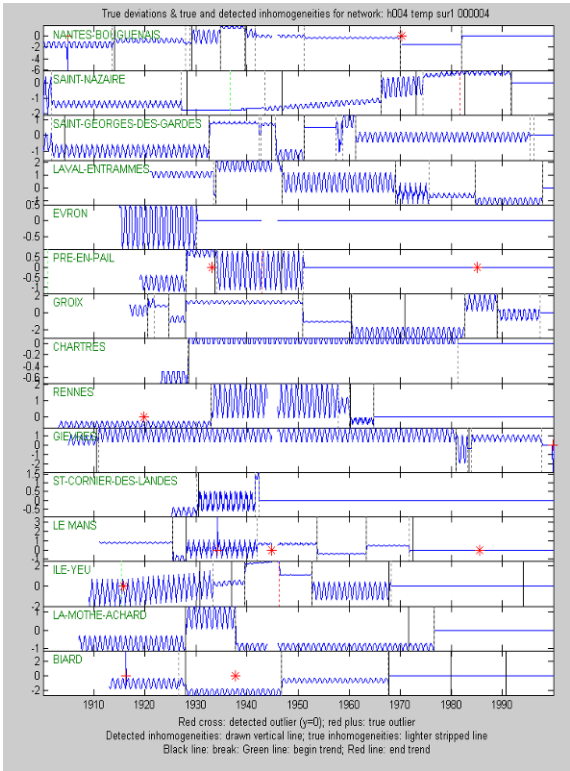
2. BENCHMARK DATA BASE

The new automatic method is tested on the artificial data set created for the COST (European Cooperation in Science and Technology) Action ES0601 (http://www.homogenisation.org). This data set was created for the comparison of existing homogenization software and is meanwhile openly available. It was created as described by Venema et al. (2012) using homogenized monthly observations as base. The surrogates imply artificial seasonal cycle and trends. Additionally, inhomogeneities such as outliers, trends and breaks are included to be detected and corrected in the COST Action blind study.

The data consist of 15 networks with 5, 9, or 15 stations spanning 100 years of monthly data.



a) Added inhomogeneities



b) Example of a surrogate temperature network (Victor Venema)

Figure 3. Inhomogeneities added into the surrogate networks

3. HOMOGENIZATION METHOD

The homogenization method consists of three steps. 1) Classifying the stations into networks with comparable precipitation behavior - this allows a relative homogenization and reduces the false detection indicated by a true change in the precipitation trend 2) Detection and 3) Correction of breaks. The last two points base on a modification of the software Prodige (Caussinus and Mestre, 2004).

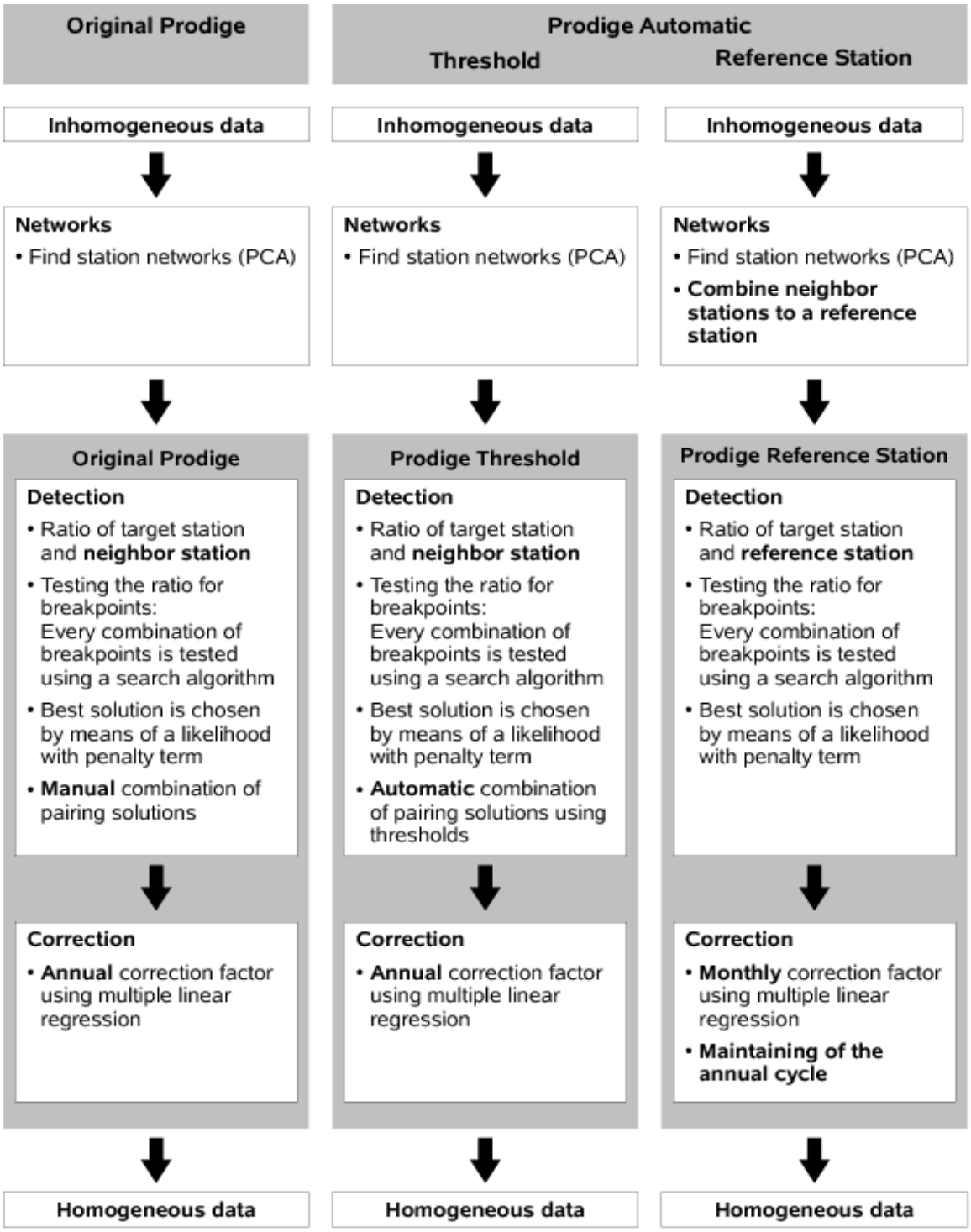


Figure 2. Overview over original and automatic Prodige

The original Prodige method has some limitations, which are also inherited in the methods developed from the original and described in this paper. The limitations are: 1) A homogenous time series is homogeneous between detected breaks. 2) Artificial trends can only be corrected by setting additional breaks. 3) Regarding precipitation, the logarithmic ratio of two stations has to be normal distributed and the variance during the entire time series has to be constant. 4) The disadvantage of Prodige being only partly automatic has triggered the development of a modified method as described in this paper.

3.1 Network selection

For homogenization, measuring stations with similar climate conditions are needed. The selection of stations is based on daily precipitation over 1 mm. This is due to the fact that small precipitation amounts are strongly inhomogeneous.

The data series are classified into regional networks of related precipitation characteristics using a principal component analysis followed by a *Varimax Rotation*. Input data are the previously introduced indices. Assuming the presence of inhomogeneities, precipitation values below 1 mm are deleted and the 30 years moving trend is subtracted before the networks are selected. This leads to much better-defined areas clearly corresponding to geographical sub regions.

3.2 Detection

The detection part as it is described by Mestre (2003a) depends on investigating logarithmic time series ratios $y_{i,j}$ of two stations i and j within the same climatic region (Sec.3.1)

$$y_{i,j} = \ln(station_i) - \ln(station_j)$$

An example for this difference time series $y_{i,j}$ can be seen in Fig. 3.

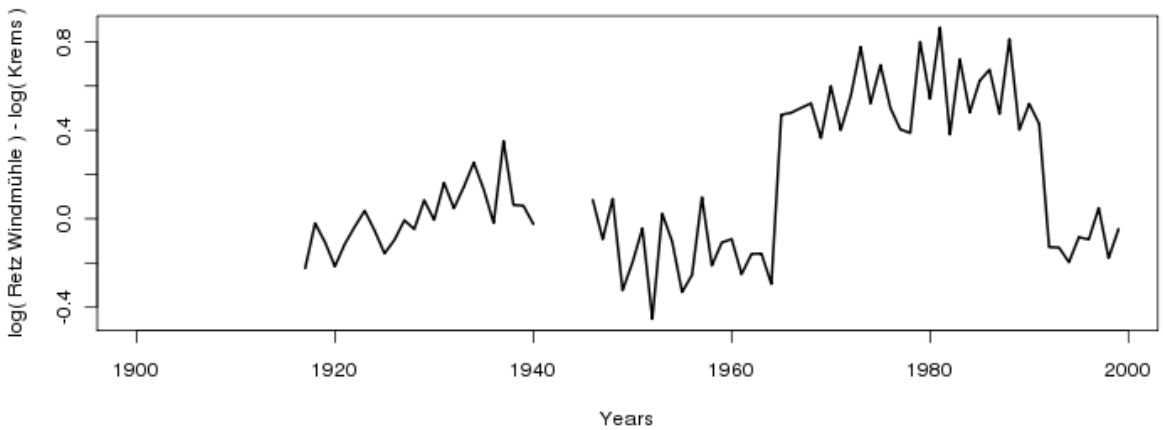


Figure 3. Log ratio of two surrogate time series in Network 4 based on the stations Retz and Krems

The results of this pairwise comparison are possible breakpoints for each pairing within any network. The next task is to assign the breaks to the individual stations.

Two approaches have been tested to automate this procedure. The first applies a threshold method and the second combines the neighbor stations to one reference station.

For the first method the detected breaks per combination are expressed as vectors for each pairs of stations *i* and *j*. The break vector consists of the entries 0 (no break) and 1 (break) for every year, for example $break_{i,j} = [0\,0\,0\,1\,0\,0\,0\,1\,0...]$.

To estimate the break points for a single station *i*, only those stations *j* are used with a correlation $cor_{i,j} > 0.8$. All break vectors are combined to a single one, depending on the correlation.

$$break_{i,t} = \frac{\sum break_{i,j,t} \cdot cor_{i,j}}{\sum cor_{i,j}}$$

The vector $break_i$ has an entry for every year summarizing the detected breaks in the pairs. The selection of the threshold for a break is based on many empirical experiments. Several runs are preformed with decreasing thresholds. In this way first strong inhomogeneities are detected and corrected, before the small breaks are searched.

The second method combines neighbor stations in the network to a single reference station. Only one comparison has to be preformed per station resulting in the direct detection of the breaks. The smaller the difference in climate signal from target and reference station the smaller are the breaks that can be found. Without breaks kriging would be the first choice, but in this case strong breaks in the nearest station cause a break in the reference series. For this reason the reference series is created from the mean of standardized time series of neighbor stations.

$$Refernce\ station_i = \frac{\sum standardized\left(\ln\left(station_k\right)\right)}{number\ of\ neighbor\ stations}$$

The difference series can be calculated as

$$y_i = standardized\left(\ln\left(station_i\right)\right) - reference\ station\ .$$

Then the detection algorithm is applied to all difference series y_i.

3.3 Correction

The detected breaks are corrected by means of multiple linear regression with binary coding based on Mestre (2003b). The result is an offset between the means at breakpoint positions that can be corrected. This information is only relative, therefore the usual approach is to keep the last segment of annual values constant and correct backwards.

In this paper a monthly correction factor was calculated by performing the regression for every month. But instead of taking the last segment as base the annual cycle is maintained on the condition that the range of standardized annual cycle within a network is smaller than two standard deviations (Fig. 4). This is fulfilled for networks in the Benchmark data set.

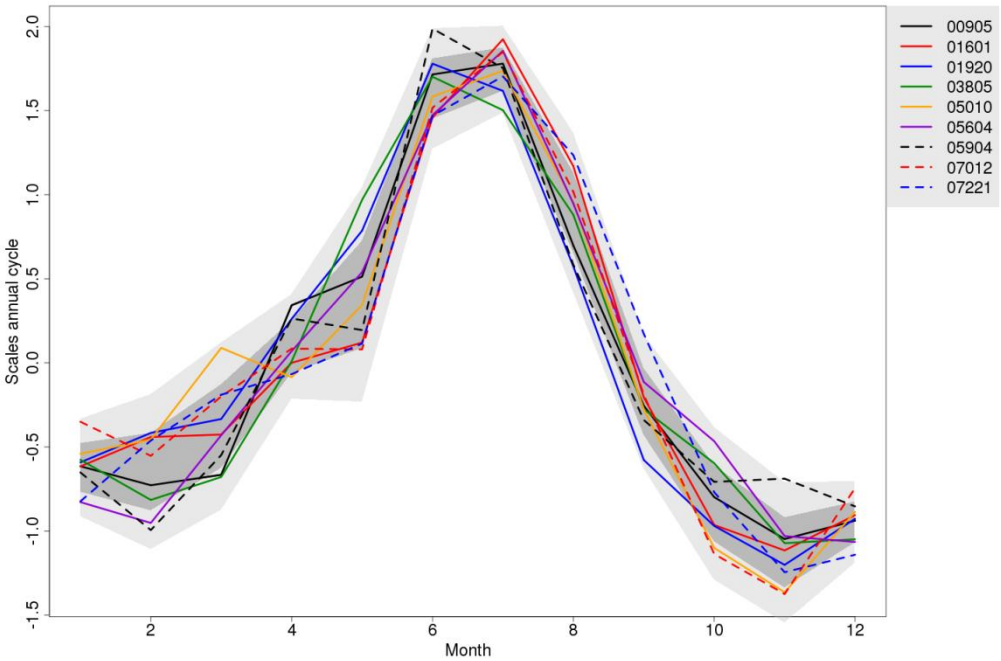


Figure 4. Scaled annual cycle of network 4

4. VALIDATION USING BENCHMARK DATA SET

The performance of including thresholds into regular Prodige can be seen in Figure 4. On the Benchmark data set, the automated algorithm has the same efficiency as the original software.

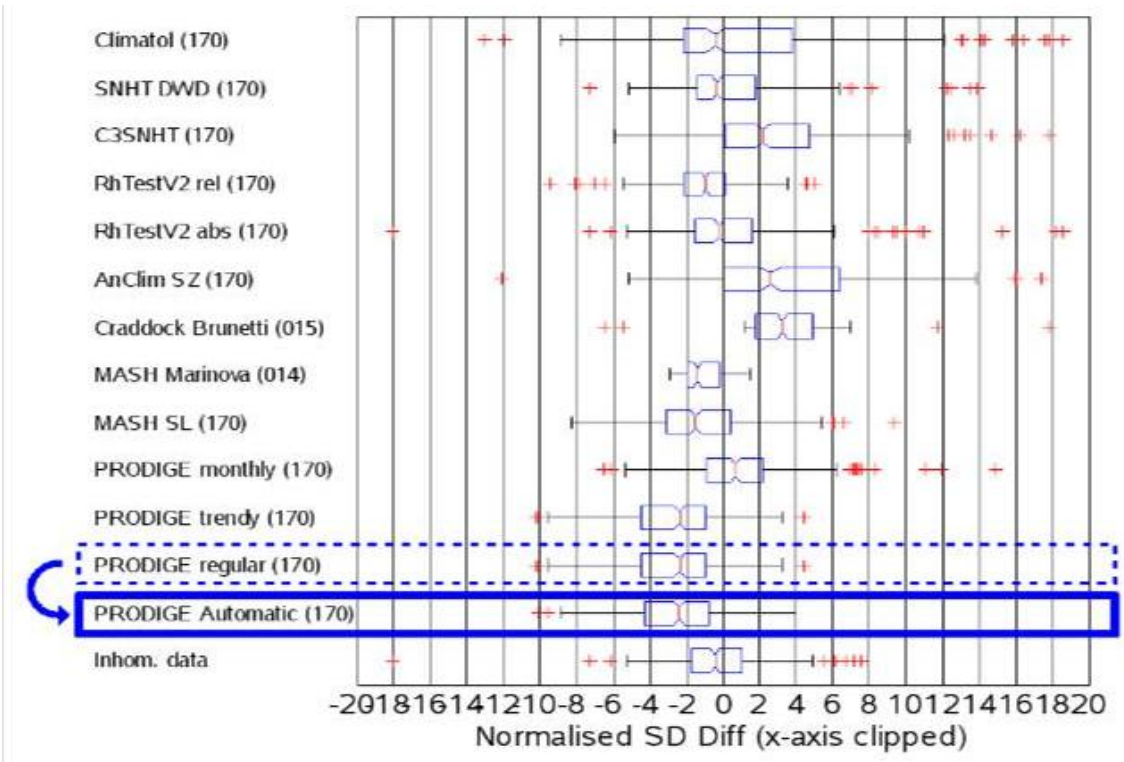


Figure 5. Standard deviation between original homogenous data and data homogenized by various methods (y-axis). The x-axis is normalized setting the mean performance of the contributions to 0. (Victor Venema)

The other software versions are compared with the threshold Prodige in Table 1. The detection skill is estimated by means of the Heidke Skill Score comparing the break point positions detected by the software with the known positions in the Benchmark data set. The correction is estimated by calculating the root mean square error between the homogenous and the homogenized data set and dividing by the inhomogenous data. This means a value below 1 shows an improvement of the data and a value above 1 indicates a change for the worse.

Table 1. Comparison of detection and correction skills of the algorithms described in this paper

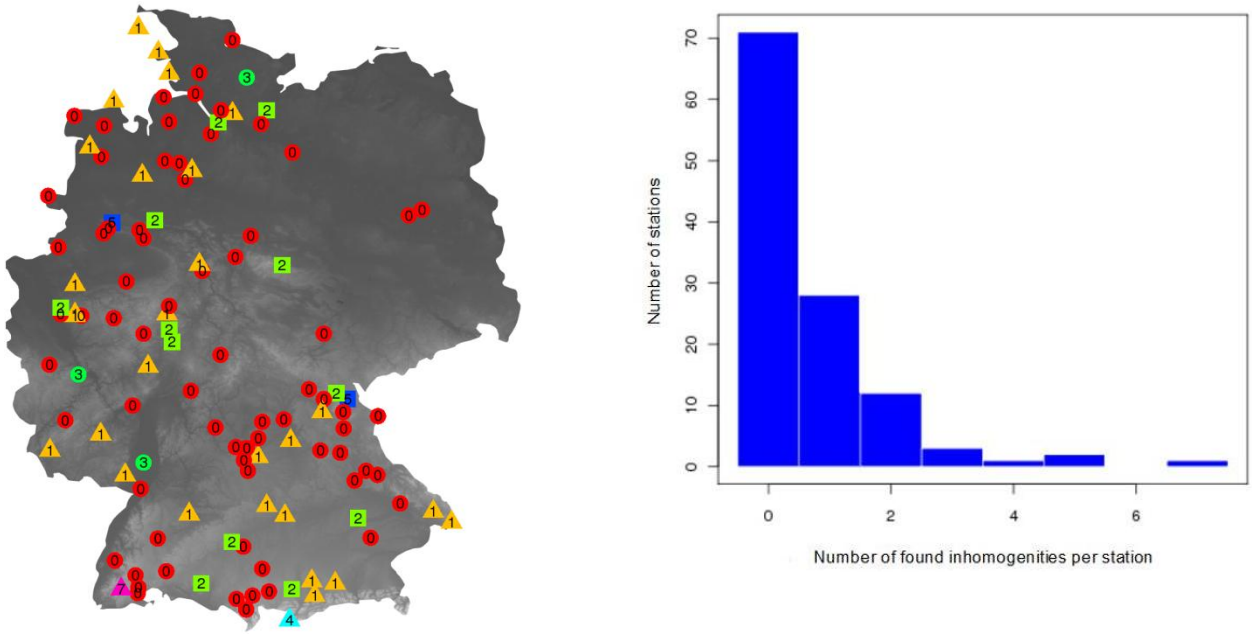
Algorithms	HSS[1]	CRMSE[2] yearly	CRMSE monthly
Prodige automatic: Regular Prodige with threshold (Fig 4)	0.28	0.60	1.03
Reference station all stations	0.32	0.59	0.89
Reference station least cor. excluded	0.29	0.64	0.92

[1] Heidke Skill Score to verify the break point position
[2] Centered Root Mean Square Error to verify the correction

Table 1. shows that both detection methods preform equally well. The main difference can be found in correction results. Calculating a monthly correction factor improves the monthly values, but leaves the quality of the annual data unchanged.

5. FIRST RESULTS

The algorithm *Prodige automatic* was applied to the daily data set described in section 1. Figure 6 shows a summary of the stations with data series, which have been corrected. In most of the stations (70 out of 118) no breaks were found, but up to 7 breaks per station have been detected, as can be seen in Figure 6a).



a) Number of breaks per station (map) **b) Number of breaks per station (histogram)**

Figure 6. Number of breaks per station shown as a) map and b) histogram

Figure 7 shows an example for a time series strongly contaminated by inhomogeneities, for which station history is available. In such case the detected break points can be easily validated by meta-data.

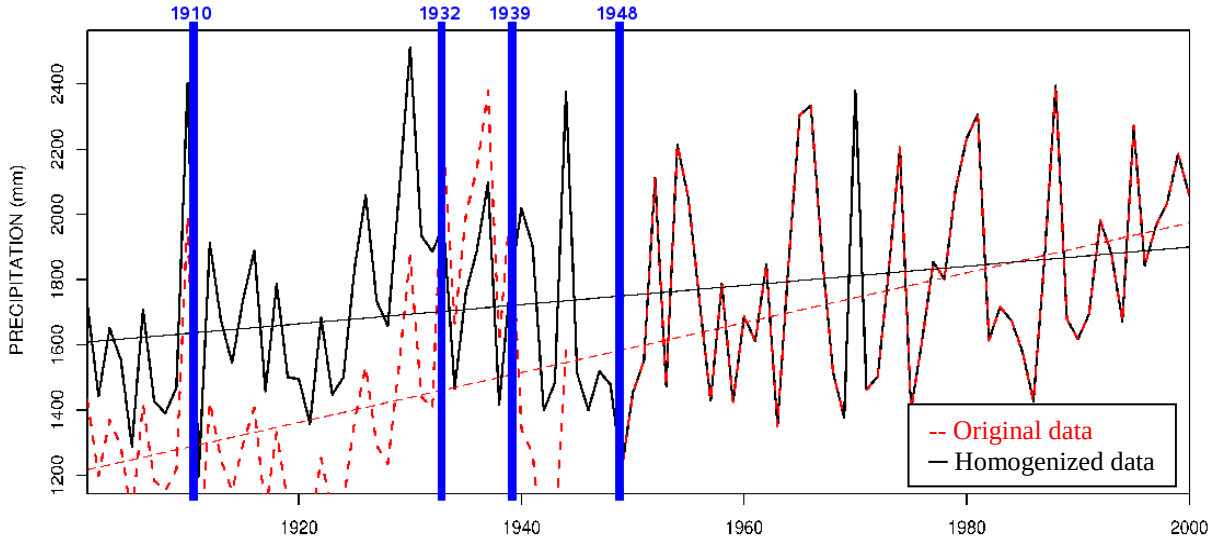


Figure 7. Annual precipitation of Zugspitze: Original (red) and corrected (black) with detected breaks (blue)

Table 2. Possible breaks in Zugspitze meta-data

Year	Reason
1910	Building constructions
1930	Cutting middle mountain top
1938	Cutting western mountain top
1945+	Instrumental change

6. CONCLUSIONS AND OUTLOOK

Especially for data sets with many stations it is important to have automatic algorithms. Both approaches, using either thresholds or constructing reference series, show good results on the applied Benchmark data set.

Changing from annual to monthly correction does not affect the annual trends if the annual cycle is included in the calculation. On the other hand, the monthly trend correction is improved.

A validation using real data is still necessary before relying on the automatic methods. This includes sensitivity studies on networks to discover the influence of single stations and the performance in less correlated networks. The influence on number of breaks and break point positions are also of interest, just like the certainty in trend correction.

Additionally the results should be compared with those of other independent algorithms.

References

Caussinus, H., and O. Mestre, Detection and correction of artificial shifts in climate series, Journal of the Royal Statistical Society. Series C (Applied Statistics), 53(3), 405–425, 2004.

Mestre, O., Detecting multiple change-points in a gaussian sample using dynamic programming, Proceedings of the fourth seminar for homogenization and quality control in climatological databases, Budapast, Hungary, WMO, WCDMP-No. 56, 89-92, 2003a

Mestre, O., Correcting climate series using ANOVA technique, Proceedings of the fourth seminar for homogenization and quality control in climatological databases, Budapast, Hungary, WMO, WCDMP-No. 56, 89-92, 93-96, 2003b

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89-115, doi:10.5194/cp-8-89-2012, 2012.

PORTUGUESE TEMPERATURE DATASET HOMOGENEITY WITH HOME R

L. Freitas[1], M. G. Pereira[1,2], L. Amorim[3,4], L. Caramelo[1], M. Mendes[3], Luís F. Nunes[4]
[1] Centre for Research and Technology of Agro-Environment and Biological Sciences (CITAB), University of Trás-os-Montes and Alto Douro, Vila Real, Portugal,
[2] Instituto Dom Luiz – Universidade de Lisboa, Lisboa, Portugal, gpereira@utad.pt
[3] Instituto de Meteorologia, IM, I.P., Rua C, Aeroporto de Lisboa, 1749-077 Lisboa, Portugal
[4] Laboratory for Systems, Instrumentation and Modeling in Science and Technology for Space and the Environment, Faculty of Sciences of University of Lisbon, Portugal
gpereira@utad.pt

Abstract

The maximum and minimum temperature Portuguese dataset is managed by the Institute of Meteorology (IM) which is national entity responsible to manage the observation network and to compile and archive the observed values. This network consists of a set of stations distributed throughout the territory with greater density in the urban areas and in sites of higher altitude within the major mountainous regions of the country. From the first records (mid 1800s), this network consisted of classical weather stations (CWS), equipped with conventional instrumentation such mercury thermometers and/or thermographs. By the end of the 1990s some of these were then complemented/replaced by automatic weather stations (AWS), with electronic sensors. The network also includes new stations installed at neighboring sites (e.g., airports). This dataset includes time series with more than one hundred years of data (monthly data), but the great majority (daily data) starts in 1940.

Besides the replacement of instrumentation (due to failure or damage), some of these stations have been relocated (usually from in the meantime urbanized regions to the periphery of the cities/localities) and/or have been replaced by automatic weather stations (located or not in the same places). Consequently, the series present heterogeneities that have to be identified and amended. On the other hand, the new software for detecting and correcting climatological series, developed under the COST Action ES0601, "Advances in homogenization methods of climate series: an integrated approach" (HOME), was just released. The HOME's main objective is to achieve a general method for homogenizing climate and environmental datasets and the software HOME R (HOMER) was developed based on the most adapted statistical procedures for detection and correction of varying parameters at different space and time scales. Thus, this work aims to use the HOMER software for detection and correction of monthly temperature series in the north region of mainland Portugal. After a preliminary analysis for quality control of the dataset, the time series for the same weather station (CWT and AWS) were merged. Then, the homogeneity analysis was performed on subsets of the dataset containing sufficiently long time series from the same homogeneous climatic regions. Homogenization procedure and results of the comparison analysis between the raw corrected datasets will be presented.

1. INTRODUCTION

This work results from the cooperation between a team from the Institute of Meteorology, IP Portugal (IM) and a group of researchers from the Centre for Research and Technology for Agro-Environmental and Biological Sciences (CITAB) at the University of Trás-os-Montes and Alto Douro (UTAD), which has been working in quality control and correction of environmental databases (e.g., Pereira et al., 2011). Both teams are interested in the study of the homogeneity of meteorological databases and have been working together in the homogenization of time series of minimum and maximum temperature in the Northern region of Portugal. This region is characterized by an irregular topography, a simple river network but diverse land use/occupation and a Csb (temperate with dry or temperate summer) and Csa (temperate with dry or hot summer) type of climate.

The Portuguese observational network started recording the first set of values in the mid-1800s, with classical weather stations (CWS), equipped with conventional instruments. Naturally, during the entire observing period some changes affecting climate data series have to be considered, such as the replacement of damaged instruments, the dislocation of the weather stations, usually from the urban to rural areas in the periphery of the cities/localities and those related to local conditions, as station surroundings significantly change over the years. Besides these typical changes, it is also of particular interest the replacement of manually operated instruments by electronic systems (automatic weather stations, AWS), with electronic sensors, located (or not) in the same places, at most of the sites of IM's network in the end of the 90's. At many of these sites, parallel observations with both methods were maintained during some years.

The general objective of this work is to contribute to assess the homogeneity of the Portuguese maximum and minimum temperature datasets making use of different methods and of the COST ACTION ES0601 (Advances in homogenization methods of climate series: an integrated approach HOME) main findings specifically the homogenization methods comparison analysis. Specific objective of this work are: (i) to use the new HOMER (Home, 2011) homogenization software which combines different methods for detection and correction of monthly temperature series in mainland Portugal; and, (ii) to test HOMER in what respects to several aspects, such as stations density, spatial domain, quality of the time series.

2. DATA

The database of minimum and maximum temperature used in this work corresponds to the values of the monthly arithmetic averages calculated from daily values observed at stations managed by the IM. This network of stations is well distributed throughout the national territory but has a slightly higher in densely populated regions or higher altitudes.

Stations that have been displaced or replaced by automatic stations have different identification numbers (ID). Information from these stations was merged to produce just one time series, representative of the local stations that originated it. The ID assigned to this time series is the same of the most recent one (from the automatic station, for example) since, in the future, the data will be recorded in this time series. On the other hand, the information from the fusion process is the metadata used in the homogenization analysis. The general characteristics of the original and merged series, such as ID, name, elevation, first and last date of registration and duration, are presented in Table 1.

Table 1 – Information on the minimum and maximum temperature observation station network used in this work: identification code (ID1) of each original time series and of the time series obtined after the fusion process (ID); stations name; station type; geographic coordinates (latitude and longitude); altitude (m); year of the first and last records; and length of the original (DUR1) and merged (DUR2) time series (in years)

ID	Initial ID	Station name	Type	Lat (ºN)	Lon(ºW)	Altitude (m)	Start	End	Dur1	Dur2
705	705	Anadia	AWS	40.44	08.44	45	1941	2010	69	69
622	622	Braga / Merelim	AWS	41.58	08.45	65	1998	2010	12	79
	023	Braga / Posto Agrário	CWS	41.55	08.40	190	1931	2007	76	
575	575	Bragança	CWS	41.80	06.74	690	1932	1998	66	66
575	575	Bragança	AWS	41.80	06.74	690	1999	2010	11	11
079	079	Caramulo	CWS	40.57	08.17	810	1937	1994	57	57
107	107	Coimbra / Bencanta	CWS	40.21	08.46	35	1941	2010	69	69
549	549	Coimbra / Geofísico	CWS	40.21	08.41	141	1864	2010	146	146
104	104	Dunas de Mira	CWS	40.45	08.76	14	1935	2005	70	70
683	082	Guarda	CWS	40.54	07.27	1019	1901	2000	99	109
	683	Guarda	AWS	40.53	07.28	1020	2001	2010	9	
635	035	Miranda do Douro	CWS	41.50	06.27	693	1942	2001	59	68
	635	Miranda do Douro	AWS	41.50	06.27	693	2000	2010	10	
632	032	Mirandela	CWS	41.51	07.19	250	1926	2000	74	84
	632	Mirandela	AWS	41.51	07.19	250	2001	2010	9	
605	005	Monção / Valinha	CWS	42.07	08.38	80	1967	2001	34	43
	605	Monção / Valinha	AWS	42.07	08.38	80	1999	2010	11	
611	011	Montalegre	CWS	41.82	07.79	1005	1880	2010	130	130
	611	Montalegre	AWS	41.82	07.79	1005	2000	2010	10	
685	085	Nelas	CWS	40.53	07.86	440	1961	1998	37	49
	685	Nelas	AWS	40.52	07.86	425	1999	2010	11	
568	568	Penhas Douradas	CWS	40.41	07.56	1380	1932	2003	71	78
	568	Penhas Douradas	AWS	40.41	07.56	1380	1998	2010	12	
055	055	Pinhão	CWS	41.17	07.55	130	1941	2010	69	69
	655	Pinhão	AWS	41.17	07.55	130	2010	2010	0	
545	545	Porto / Pedras Rubras	CWS	41.23	08.68	63	1967	1998	31	43
	545	Porto / Pedras Rubras	AWS	41.23	08.68	63	1999	2010	11	
546	546	Porto / Serra do Pilar	CWS	41.14	08.60	93	1863	2005	142	142
052	052	Régua	CWS	41.16	07.80	56	1933	2010	77	77
567	566	Vila Real	CWS	41.31	07.74	481	1928	1991	63	82
	567	Vila Real / C. C.	CWS	41.27	07.72	561	1992	2002	10	
	567	Vila Real / C. C.	AWS	41.27	07.72	561	1998	2010	12	
560	560	Viseu / C. C.	AWS	40.71	07.90	636	1991	2010	19	85
	075	Viseu / Escola Agrária	CWS	40.66	07.90	443	1925	2003	78	

This database contains time series, original or after the merger, with more than one hundred years in length such as the case of Coimbra-Geofísico (146 years), Guarda (109 years), Montalegre (130 years), Porto - Serra do Pilar (142 years). The network includes many other stations with much shorter time series that, for this reason, were not used.

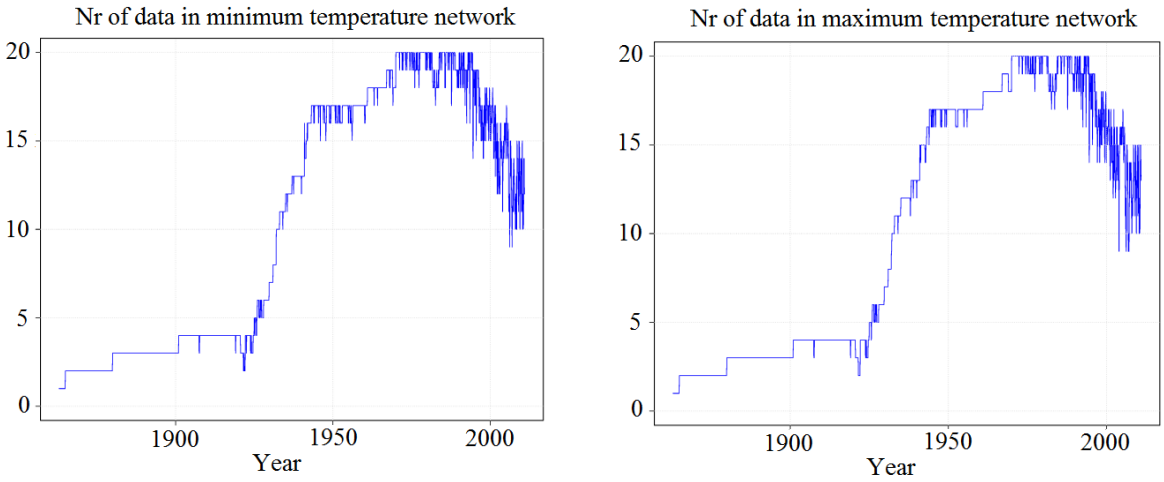


Figure 1. Temporal evolution of the number stations without missing data on the basis of minimum temperature (left) and maximum (right), located in the northern region of Portugal, for the entire period 1863-1941

The algorithms to detect and correct the inhomogeneity problems require the existence of a sufficient number of time series, as long as possible. The preliminary analysis of the database allowed finding a good tradeoff between the density of weather stations in the study area and the length of the time series if we consider study period beginning in 1941. Although the number of stations with missing values begin to increase from 2000 (Figure 1), it was decided to analyze the data from 1941 to 2010 with the aim of analyzing the longest period possible.

3. METHODOLOGY

In this work different methods were used: (i) to perform preliminary statistical analysis and quality control; and, (ii) to detect and correct problems of homogeneity. In the first case, specific software to detect outliers was developed and used along with the Climatol R package (Guijarro, 2011) for quality control of the data. The developed software for outlier detection in time series is based on the quartile values of the series itself. The outliers were identified as the values outside of the range [IQR * k-P25, P75 + k * IQR], where P25 (P75) represents the 25th (75th) percentile or the first (third) quartile, IQR is interquartile distance, defined as P75-P25, and k is a coefficient which usually assume the value 1.5 to identify outliers and the value 3.0 to identify extreme values.

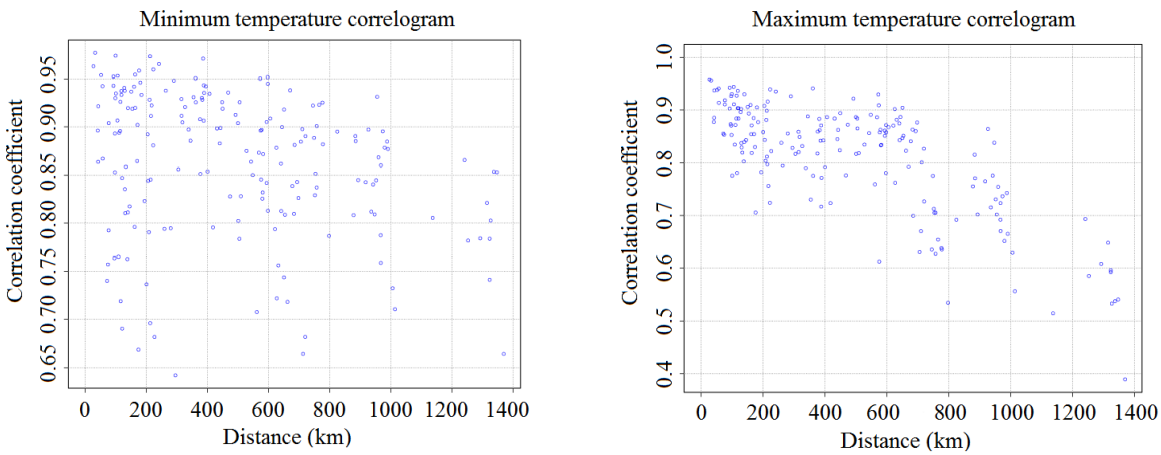


Figure 2. Correlogram of the series of differences in minimum temperature (left) and maximum (right) located in Northern Portugal for the period 1941-2010

The outliers and extreme values identified taking into account the temporal evolution of the time series are called single station outliers (SSO) and single station extreme values (SSEV). This procedure was performed for each of the series and for each month independently. The identified SSO and SSEV were then validated with a comparison analysis based on the time series of temperature differences with the neighboring stations. The criterion to select the neighboring time series can be based on the distance or on the correlation coefficient between the candidate and all other stations. In this work, six of the best correlated nearby stations were selected as long as they present a correlation coefficient greater than or equal to 0.6. Taking into account the correlogram between the temperature difference time series for both the minimum and maximum temperature (Figure 2), these conditions are not too restrictive to achieve. The outliers identified by this procedure, will be called neighbor station outliers (NSO). The SSO and SSEV were only removed from the databases if were also identified as NSO.

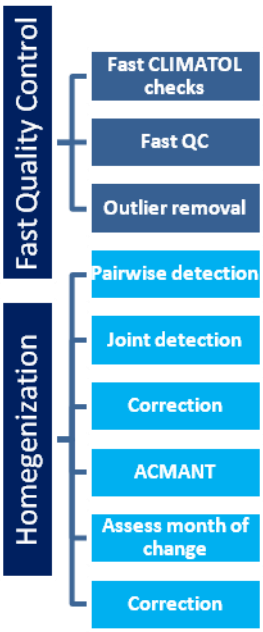


Figure 3. Scheme features HOMER, developed and made available within the COST Action ES0601

The analysis of homogeneity of the datasets was performed using the HOMER software, which was developed and made available within the COST Action ES0601. This application was written to be executed in R, which can be defined as an environment or a free language for statistical computing and graphics facilities. The HOMER (Figure 3) has two distinct groups of features. One dedicated quality control, which mainly uses the Climatol R package (Guijarro, 2011), and allows: preliminary statistical analysis of data including correlograms, histogram and boxplots monthly; the performance of cluster analysis, and estimate the density of the network. A file with the outliers identified in each station can be created and those outliers be removed from the original database. The second feature set is dedicated to the detection and correction of heterogeneities. HOMER provides three techniques for break detection: pairwise, joint and ACMANT. The pairwise technique uses predefined series (based on the correlation and distance to the candidate series) but for the sake of visualization simplicity, only the detection results with six (better correlated) other series are shown. In this technique, three types of plots are produced for each time series: pairwise series; position and amplitude of detected breaks; and position and neighbor of detected breaks. Unlike the other detection methods, the results of pairwise are not automatically considered by HOMER. Instead, the user can visually identify potential breaks and, if desired, add these discontinuities to the list of detections made with the other techniques and correct them. The joint segmentation model is a detection of break points procedure difficult to implement due to high computational execution time. With the development of new algorithms that allow their timely implementation, can now be applied to the detection of changes throughout the series (Picard et al., 2011). The ACMANT technique was also developed under the COST Action HOME (Domonkos et al., 2011) and consists of the bivariate detection of changes to the annual and seasonal scale which includes a penalty term.

Several aspects of the functioning of HOMER should be highlighted, such as:

- Running in automatic mode or with user intervention;
- Operate in an iterative mode, in the sense that it can perform the detection and correction features more than once, that is, while detections were made;
- Methods can be applied to original (raw) data or to data that have undergone quality control (qc);
- The detection methods can be applied to input data (raw or qc) or to corrected data by HOMER or other methods;
- The detection method ACMANT can only be used after the first correction procedure, which is only possible after the first application of the other detection methods (pairwise+joint);
- The correction is always performed on the input data (raw or qc);
- Run in interactive mode, which means that the user can intervene in the identification and elimination of outliers from the database, in the validation of detected breaks with joint and ACMANT detection techniques as well as in the addition of breakpoints not detected by the joint or ACMANT but suggested by pairwise.

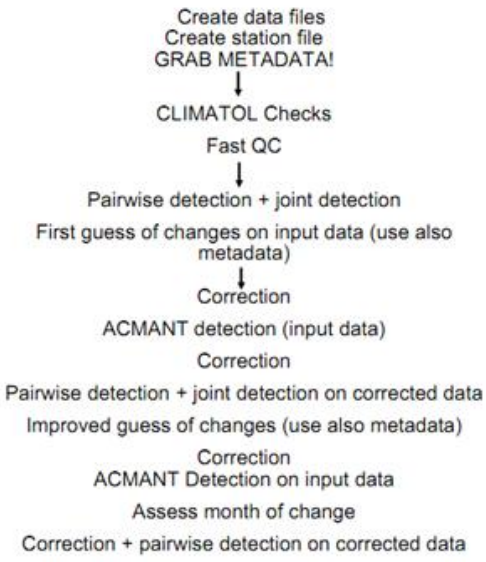


Figure 4. Proposed aplication scheme of HOMER. Extracted from HOMER manual, HOME_R.pdf

The HOMER also offers the user the ability to produce graphs of the input (raw or qc) and homogenized time series. After identifying and correcting the break points, the homogenized series are also corrected with regard to missing values.

Table 2. Number of outliers detected on minimum and maximum temperature time series (Table 1) analyzing each singular station (SSO), comparing with neighbor stations (NSO), and total number of outliers (NO) removed from the datasets. The number of extreme values detected (SSEV) in each time series is also shown

Weather station	Number of maximum temperature outliers				Number of minimum temperature outliers			
	SSO	SSEV	NSO	NO	SSO	SSEV	NSO	NO
Anadia	7			0	7			0
Braga	14			0	5		1	1
Braganca	2			0	6			0
Caramulo	5			0	11			0
Coimbra_Bencanta	7		1	1	8			0
Coimbra_Geofisico	9			0	9			0
DunasdeMira	8		1	1	15		1	1
Guarda	6			0	10			0
Miranda do Douro	10			0	6			0
Mirandela	5		1	1	2		1	1
Moncao_Valinha	2			0	4			0
Montalegre	4			0	9			0
Nelas	4			0	3		3	3
Penhas Douradas	8			0	10		3	3
Pinhao_SantaBarbara	16	1	8	8	9	2	1	1
Porto_PedrasRubras	0			0	7			0
Porto_SerradoPilar	5			0	8			0
Regua	4			0	4			0
VilaReal_CC	10			0	5			0
Viseu_CC	5			0	6		3	3

As referred previously, the HOMER can be applied in fully automatically mode or with user intervention in the validation or addition of breaks. Operator intervention can produce different solutions, while the automatic application of HOMER can also be characterized by difficulties associated with the data itself, particularly its quality (e.g., missing values, network density) and the temporal location of the detections.

The authors of HOMER propose an application scheme that includes user intervention at various stages of the process (Figure 4). In this paper, HOMER was applied following this scheme, but human intervention was limited to a second phase of implementation in order to minimize the subjectivity of the solutions found. The HOMER was applied fully automatically from the first application of the methods of detection pairwise + joint to the second application of the method ACMANT on input data. Only at this stage, the hypothesis to include detections suggested / made with the initial pairwise methods and pairwise + joint performed on corrected data was considered.

4. RESULTS

4.1 Fast quality control

The quality control of data was based on the results of the identification of SSO, SSEV and NSO. The results of this analysis for the minimum and maximum temperature are presented in Table 2. It should be noted that were only identified 13 (11) outliers in the minimum (maximum) temperature dataset. Only the values that simultaneously fulfilled the criterion to be classified as SSO and NSO were considered outliers and excluded from the datasets and from the analysis of homogeneity.

All the functionalities of quality control available on HOMER were applied and special attention was given to the boxplot charts and to the cluster analysis.

In fact, there is a clear and statistically significant distinction between the stations located at high altitude in comparison with those located at a lower altitude, in the case of minimum temperature, while the major difference in spatial distribution of the maximum temperature occurs between stations located in the coast and inland. These differences are more significant in winter than in summer (Figure 5).

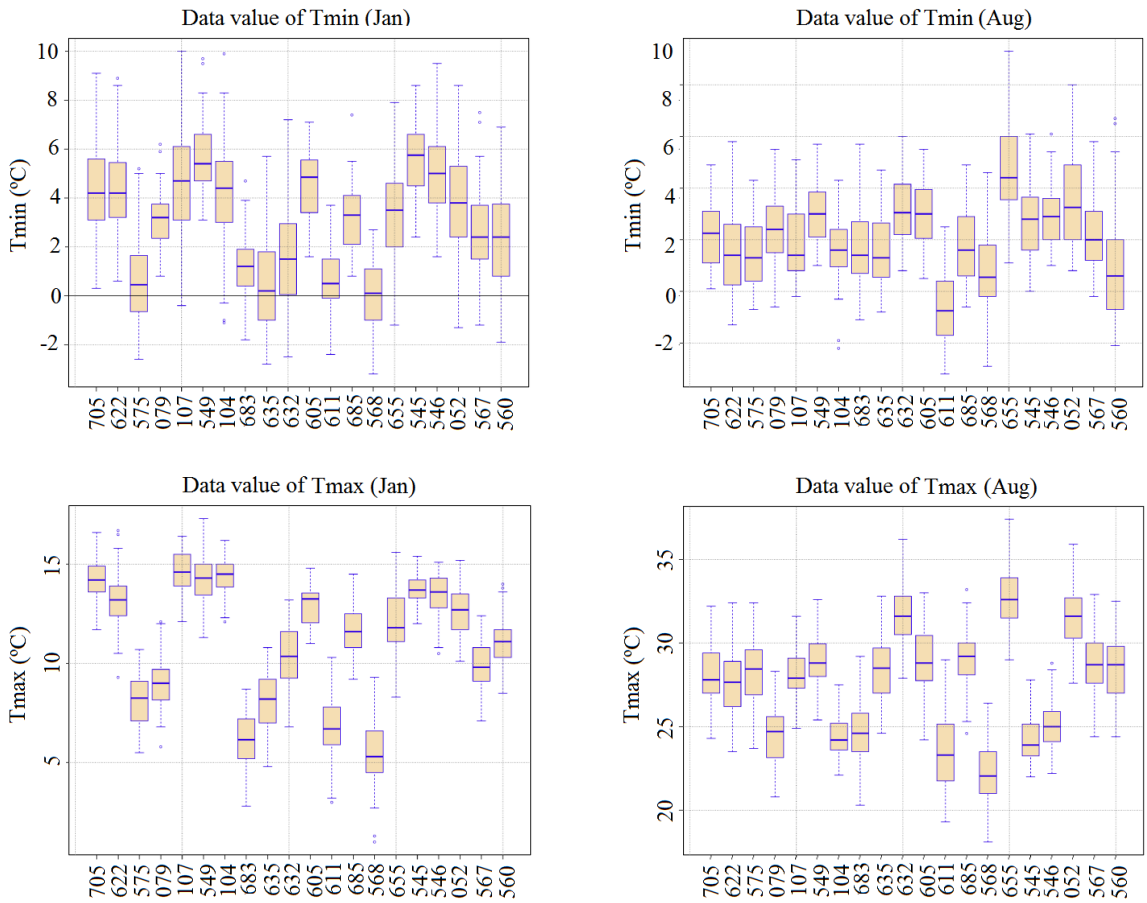


Figure 5. Boxplot of the temperature minimum (top) and the maximum temperature (bottom), recorded at stations in the Northern region of Portugal, in January and August

4.2 Homogeneity analysis with HOMER

Table 3 and Table 4 summarize the results obtained with the application of HOMER for the analysis of homogeneity of maximum and minimum temperature after having conducted the quality control of databases and following the method of application described in the previous section. These tables present, essentially, the first estimates of the location of the breaks detected by the various methods in the input data while the columns C represent the phases of data correction. It should be noted that the detections are identified by year of occurrence. The "Assess month of change" (AMC) procedure allows the identification of the exact month of the break which could be in one of the previous or the following years.

Table 3. Location of the breaks detected in the time series of maximum temperature recorded in each of the stations in the Northern Portugal, with the pairwise, joint and ACMANT detection methods available in the HOMER software, performed on input and corrected data. Columns C corresponds to correction procedure while ACM (removed) corresponds to the breaks that have to be removed due to difficulties to correct the data after the Assessing the Month of Change. Breaks with a star (*) were not validated by the user

Station	pairwise input data	joint input data	C	ACMANT input data	C	pairwise+joint corrected data	C	ACMANT input data	ACMANT input data (added)	AMC (removed)
Anadia	49, 79	49		49, 77		2002				2002
Braga	50, 65, 89	48, 81		48, 59, 67		49				50
Braganca	43, 47	71, 93		71, 93						
Caramulo	62, 77					78				
Coimbra_Bencanta	45, 52	48		53, 90						
Coimbra_Geofisico	71, 69	49, 71		49, 71						
DunasdeMira	90, 68, 71					68, 72, 90			90	
Guarda	73			73		2001				2001
Miranda do Douro	79					79		78		
Mirandela	50, 72	50, 93								
Moncao_Valinha		92		94						
Montalegre	73	95		53, 72, 95		97, 99		52*		
Nelas		90		69, 90		93, 98		89		93
Penhas Douradas		89		93				72*		
Pinhao	52, 55, 74, 96, 2002	74, 96		55, 74, 96		97, 2003		98	52	97, 98, 2003
Porto_PedrasRubras										
Porto_SerradoPilar	73	92							73	
Regua	53, 79	93				53				
VilaReal_CC	53, 68	53		58, 91						
Viseu_CC	80, 52, 93	80, 93		53, 93						

Table 4. As in table 3, but for the minimum temperature

Station	pairwise input data	joint input data	C	ACMANT input data	C	pairwise+joint corrected data	C	ACMANT input data	ACMANT input data (added)	AMC (removed)
Anadia	71			71		70		63, 70, 85		70
Braga	52, 73, 79	92		63, 92		79, 2007			79	
Braganca	62, 79, 92	48, 79		46, 62, 79		69				46
Caramulo										
Coimbra_Bencanta	69, 95	79		79		50, 94			50, 69	
Coimbra_Geofisico				49, 67						
DunasdeMira	42, 87, 79	48, 72, 80, 87		72, 80, 84, 87		91, 93		69, 78	93	
Guarda	46, 70					46, 70			46, 70	
Miranda do Douro		73		73		72				
Mirandela	47, 61, 76	76, 86, 97		76						
Moncao_Valinha	74			71				99		
Montalegre	59, 70, 73	50		50		70, 73			60, 73	
Nelas				2008		94, 98				
Penhas Douradas	80	50				86			86	
Pinhao		01		58, 2003, 2008		59, 2007				
Porto_PedrasRubras	68, 74, 90	90		90		95				
Porto_SerradoPilar				55, 58, 92						
Regua	64, 78, 87	78, 87		78, 87, 2000		64			64	
VilaReal_CC	46, 74	44, 74		44, 46, 74						
Viseu_CC	60	57, 76, 93		93		79, 89		55	74, 83	

It is also important to note that not all years reported in the tables correspond to breaks detections. In fact, the values presented in the column "ACMANT input data" with a star (*) and information contained in "AMC (removed)" column correspond to detections identified but not included in the list of breakpoints. The values of the first case correspond to additional breaks detected with the ACMANT on input data, but not validate by the user because it is localized in a year when there are no significant differences in relation with neighboring series, and for this reason it is considered a false detection and, therefore removed the detection list. The breakpoints listed in the “ACM (removed)" column correspond to cases where the user has validated the results of the detection methods and assessed the exact month of change, but was not possible to correct it and the break was also removed from the detection list. Tables 3 and 4 do not include the residual number of breaks detected by the pairwise+joint and ACMANT methods on corrected data (6 in maximum temperature and 3 in minimum temperature time series) because the vast majority of them coincide with those that have been removed from the list of detections as explained previously.

The reasons for not being able to correct the break points identified with the HOMER are mainly the existence of missing values before and / or after its location, the position of the breakpoints to be very near the beginning or the end of series, and the existence of very close points of detection, which is also not recommended. It should also be noted that the detections performed on the corrected data results from the fact that they are performed on series with less noise and detections were performed on data without missing values. None of these conditions corresponds to the characteristics of the original data. In summary, we detected 61 break points in the minimum temperature time series and 55 in the database of the maximum temperature.

4.3 HOMER sensitivity tests

Results from the cluster analysis were used to test the influence of the station density and spatial domain on the results obtained with HOMER software. Maximum and minimum

temperature network was divided between stations located in the coastal and interior areas and between low and high altitude stations, respectively. In order to avoid the subjectivity of the user intervention, HOMER detection and correction functionalities were applied in a full automatic mode. Table 4 shows the results obtained with the homogenization analysis performed on the entire network (except for the Dunas de Mira weather station) and on subsets of maximum temperature time series.

Several aspects should be highlighted, namely: (i) the existence breaks in the exactly and similar temporal location but also the detection of breaks on different locations; (ii) the difference between the total number of detected breaks (35 analyzed when the entire dataset is, 40 when the network is divided).

5. DISCUSSION AND CONCLUSIONS

The importance of quality control for the analysis of homogeneity has been reported in the vast literature on the subject, is regarded as one of the "ten commandments" for homogenization (Olivier and Aguilar, 2011) and is listed in the manual use of the HOMER. The existing information on this issue also point towards the analysis of outliers in the monthly data to be very limited by the need for a careful analysis of daily data, to identify its nature and provide their eventual correction. In fact, the more correct procedure would be to develop and implement a system of quality control of daily data before computing the monthly averages. It should also be noted that in this study only a few number of outliers were identified in the time series of both minimum and maximum temperature. These outliers were removed from the database by replacing the original monthly average by the value attributed to the lack of data. This is another important aspect, since the original time series already have some missing values and the outlier removal could limit the ability to identify and correct the database.

Table 5. Maximum temperature breaks detected with automatic use of HOMER over the entire dataset or for subsets of coastal and inland weather stations. Similar and coincident and breaks are marked in light grey and in grey, respectively, while dissimalr breaks are marked in black. Total number of breaks (SUM) and difference between the numbers of breaks in each time series (DIF) are also shown

Station Name	Breaks detected (entire network)								SUM	Breaks detected (coastal and inland stations)								SUM	DIF
Braganca	03/72								1	12/48	06/67							2	**1**
Caramulo									0									0	**0**
Guarda	03/48	12/73							2	12/73								1	**-1**
MirandadoDouro									0									0	**0**
Mirandela	12/50	12/71							2	12/50	12/57	11/72						3	**1**
Montalegre	05/96								1									0	**-1**
Nelas	12/69	12/86							2									0	**-2**
PenhasDouradas									0									0	**0**
Pinhao_SantaBarbara	04/53	12/58	11/74	08/78	11/91	12/96	12/98	12/00	8	04/53	12/58	02/78	11/91	12/96	12/98	12/00		7	**-1**
Regua	12/53	12/57	12/71	12/00					4	12/57	12/72	12/99						3	**-1**
VilaReal_CC	12/53	10/59	11/82	03/92					4	12/53	10/59	08/78	02/92					4	**0**
Viseu_CC	04/52	05/80	09/94	03/00					4	12/57	04/81	01/93	09/94					4	**0**
Anadia	12/49	06/78							2	12/49								1	**-1**
Braga_Merelim	01/50	12/57							2	10/50	08/56	09/59	10/72					4	**2**
Coimbra_Bencanta	04/53								1	12/51	04/52	12/75	04/76					4	**3**
Coimbra_Geofisico	10/49	12/71							2	12/43	06/49	10/52	12/69	12/71				5	**3**
Moncao_Valinha									0	12/88	04/89							2	**2**
Porto_PedrasRubras									0									0	**0**
Porto_SerradoPilar									0									0	**0**
								SUM	35								SUM	40	**5**

Other quality control functionalities available in HOMER were used and it is important to underline the consistency between the cluster analysis results and the monthly boxplots regarding the characterization of the spatial distribution of minimum and maximum temperature. It appears that the stations with similar median values are grouped by cluster analysis. The statistical significant differences detected in the main clusters of weather stations suggest that the analysis of homogeneity of the databases could have been performed independently, by separating the stations located high altitude and low altitude, in the case of minimum temperature, and the stations on the coast and inland, in the case of maximum temperature. However, splitting the database into two subsets of stations could lead to an insufficient number of stations to complete the homogenization analysis, especially in the case of stations located at high altitude.

The process of homogenization of the maximum temperature proved to be much simpler and faster than the minimum temperature. In fact, the number of outliers and break points of the minimum temperature is lower than those detected for the maximum temperature. In addition, the number of break points in the minimum temperature that could not be fixed with the HOMER is higher than for the maximum temperature. Obtained results elucidate some of the characteristics of the analysis of homogeneity of databases with HOMER, including:

- Different detection methods produce different results; in fact, the analysis of the results presented in Table 3 and Table 4, show, for example, that not all detections made with pairwise, joint and ACMANT are coincident;
- False detections may occur and, in this work, this deficiency is more evident with ACMANT than with the joint method;
- User intervention proved to be important to validate the breaks detected with joint and ACMANT methods but also to add the breakpoints identified with the pairwise method; in fact, it appears that potential breakpoints (identified with pairwise but not inserted automatically by HOMER in the list of detections) were only identified by the joint or by ACMANT detection procedures when analyzing the corrected database.

Notwithstanding these limitations, the homogenization analysis of the time series of minimum temperature and maximum temperature at stations located in the northern region of Portugal with HOMER was successful. Three different methods of detection, recognized by the community recognizes as being particularly suitable for the analysis of temperature homogeneity were used. These methods were applied successively, first automatically and then with user intervention, resulting in corrected databases in the sense that any additional break was detected or cannot be corrected with the HOMER due to characteristics of the data.

References

Domonkos P., R. Poza, and D. Efthymiadis (2011) Newest development of ACMANT. dv.Sci. Res., 6, 7-11, 2011. doi:10.5194/asr-6-7-2011

Guijarro, J.A. (2011) Climatol version 2.0, an R contributed package for homogenization of climatological series. State Meteorological Agency, Balearic Islands Office, Spain, http://webs.ono.com/climatol/climatol.html

HOME (2011). Homepage of the COST Action ES0601 - Advances in Homogenisation Methods of Climate Series: an Integrated Approach (HOME), http://www.homogenisation.org, 2011.

Olivier Mestre, Enric Aguilar (2011) Homogenization (personal communication), COST ES0601 Training School, 5-7th of october, 2011 Tarragona

Pereira M.G., Malamude B.D., Trigo R.M., Alves P.I. (2011) The History and Characteristics of the 1980−2005 Portuguese Rural Fire Database. Natural Hazards and Earth System Sciences, 11, 1-16, doi:10.5194/nhess-11-1-2011

Picard F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam B. and Robin S. (2011) Joint segmentation, calling, and normalization of multiple CGH profiles. Biostatistics. doi:10.1093/biostatistics/kxq076

CONSIDERATIONS ON SNOW AND TEMPERATURES TIME SERIES IN WESTERN ITALIAN ALPS

Fratianni S., Terzago S., Acquaotta F.
Department of Earth Science, University of Torino, Via Valperga Caluso, 35-10125 Torino (Italy), tel. 00390116705102 Fax 00390116705148
simona.fratianni@unito.it

Abstract

In the frame of MEDARE (MEditerranean DAta Rescue) and STRADA (Strategies of Adaptation to climate change) projects, Western Italian Alps precipitation and snow depth time series recorded in stations located between 960 and 2316 m a.s.l. have been analysed in order to investigate variability and trends over the 1925-2010 period. The results outline a significant decrease of snow depth over seasonal time scale in the period 1951-2010. The snow depth decrease ranges from -0.19 cm/y up to -1.40 cm/y and the contribution to this negative trend comes mainly from spring months. The climatic conditions that caused this considerable snow depth reduction have been examined. Below 1000 m a.s.l. the snow depth decrease is due to a significant decrease in snow precipitation, as a consequence of significant increase of maximum and minimum temperatures. At higher elevation the negative trend is more marked for the combination of a stronger increase in minimum temperatures and stronger decrease in the number of days with Tmax<0 with respect to low elevation sites.

Keywords: snow depth, winter precipitation and temperatures trends, Alps, climate change.

1. INTRODUCTION

The Alps plays a key role in the hydrological cycle, being the source of the river systems that supply water to the Po, Rhine, Rhone and Danube basins. A shift in climatic regimes, particularly winter precipitation and snow cover duration, would impact heavily on the river systems originating in that area, leading to disruptions of the existing socio-economic structures of populations living within the mountains and downstream (Beniston, 1997).

Moreover snow represents a resource for hydro-power generation, irrigation, domestic and industrial water supply and for tourism based on winter sports (Terzago et al., 2012 in press).
The investigation on snow precipitation and snow cover variability is fundamental in the frame of climate change studies and in developing strategies for the mitigation of climate change effects. At local scale the need of information about snow precipitation amount, duration and variability is required for a correct socio-economic planning which takes into account the changing environment. Literature on snow climatology is rich for the Swiss Alps (Laternser et Schneebeli, 2003; Beniston, 1997; Scherrer et al., 2004) and French Alps (Durand et al., 2009; Martin et al., 1994).

Concerning Austrian Alps Schöner et al. (2009) used the snow measurement network of the Sonnblick region to describe temporal trends of snow-depth and Hantel et al. (2000, 2007) studied the sensitivity of Alpine snow cover duration.

Concerning Italian Alps, the information on winter precipitation variability is still scarce compared to the Swiss, French and Austrian sides. A general study considering the entire Alpine region over the 1971–1992 winters has been conducted by Quadrelli et al. (2001). Even though some regional studies have already been produced (Biancotti et al., 1998;

Fratianni et al., 2002; Fazzini et Gaddo, 2003), this field can be still considered as underexplored due to the difficulty of finding long term and continuous time series.

This work has been conducted in the frame of the MEditerranean climate DAta REscue (MEDARE) initiative, born under the auspice of the World Meteorological Organization, with the main goal of developing, consolidating and progressing climate data and metadata rescue activities across the Greater Mediterranean Region (GMR). Moreover, this study is part of the interregional Italy-Switzerland project STR.ADA. (STRategies of ADAptation to climate change), draw to evaluate the climatic variability in the North of Piedmont (NW Italy) in relation to avalanches risk.

The effort is then addressed to recover historical climatic time series and enhance a unique and unexplored daily climatic dataset. The data originally reported over bulletins were digitized, quality controlled, checked for homogeneity and then analysed to investigate trends and interannual variability of temperatures and snow.

2. THE DATA

The historical time series used in this study have been recovered from the paper archives of the Ufficio Idrografico del Bacino del Po (Po Basin Hydrographic Office) operational since 1920's up to 1990 and then merged to the Arpa (Regional Agency for Environmental Protection) of Piemonte and Lombardia (Fig. 1).

All the measurements are performed manually by the observers with thermometers, graduated snow stakes and snow tablets (Fig. 2).

Figure 2. Meteorological station of Lake Serrù

The main parameters used in this study are temperatures maximum and minimum (TM and Tn) and the snow depth (HS). In most cases fresh snow precipitation (HN) observations are not systematically registered, so, in order to have continuous and comparable data, HN is calculated by subtraction of two consecutive HS values. Other derived quantities are the number of snowy days SD, defined as the number of days when solid precipitation HN > 0 cm and the number of frost and ice days (Tn ≤ 0°C and TM ≤ 0°C respectively).

The data finally analysed refer to the ten stations covering the whole Piedmontese Alps and in particular the North of Piedmont (Ossola Valley in the frame of STR.ADA. project) and ranging between 960 and 2316 m a.s.l. (Tab. 1 and Fig. 3). The longest time series supply 86-years records (1925-2010) and the shortest ones 50-years records (1961-2010) and they all almost continuous.

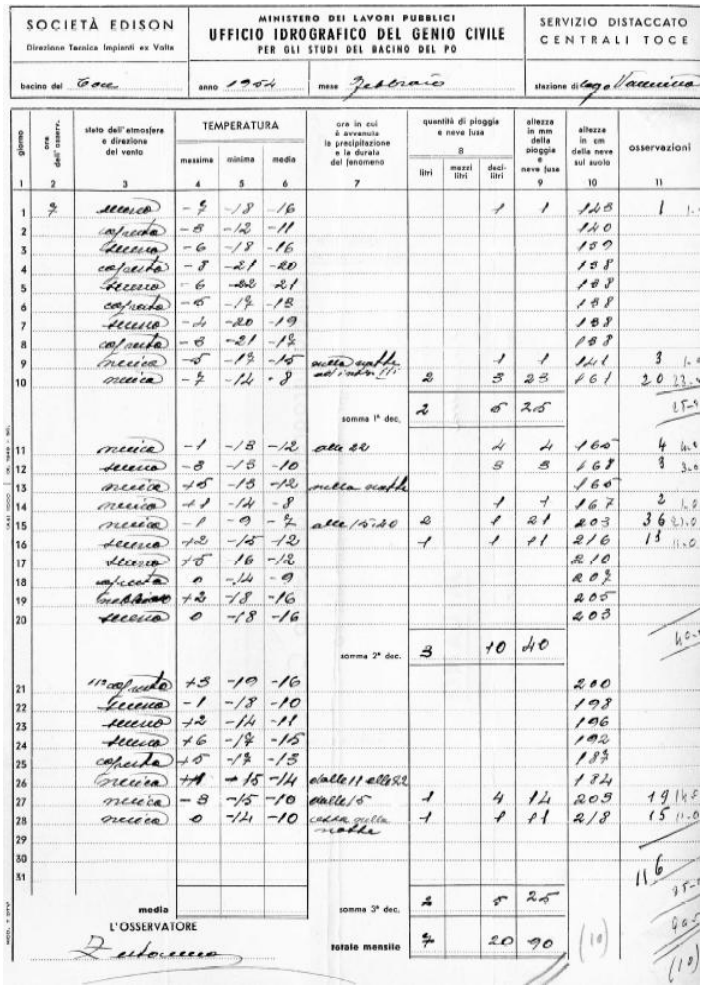


Figure 1. Extract of the paper which were digitalised data and metadata: an example of the Lake Vannino station in February 1954

Table 1. Denomination, elevation, position and operational periods of the measurement sites considered in this study

Station	Elevation [M A.S.L.]	UTM X [M]	UTM Y [M]	Period
Camposecco	2316	426645	5101275	1951-2010
Lago Vannino	2177	451230	5137189	1951-2010
Lago Toggia	2165	456227	5142763	1932-2010
Alpi Devero	1634	443114	5129624	1951-2010
Agaro	1600	422567	5094533	1961-2010
Lago Castello	1589	345381	4942026	1943-2010
Ceresole Reale	1573	362763	5032442	1926-2010
Acceglio Saretto	1540	335855	4927442	1925-2010
Alpe Cavalli	1500	431707	5104302	1932-2010
Lago Piastra	960	371372	4898574	1926-2010

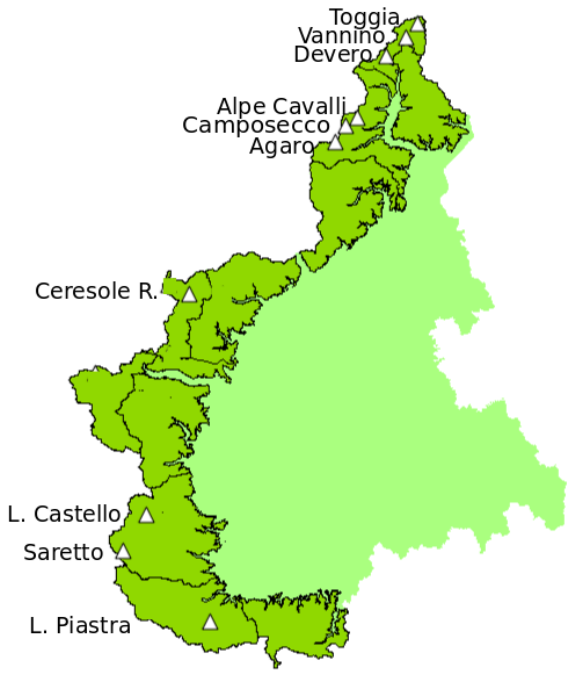


Figure 3. Geographical position of the stations selected for this study, located in the Piedmont, NW Italy

3. METHODOLOGY

The temperature maximum and minimum, snow depth and fresh snow data performed and registered by observers on the bulletins have been digitized together with all the notes regarding the measurements or the instruments anomalies.

A parallel in-depth historical research has been carried out in order to acquire stations metadata. Particular attention has been addressed to eventual relocations or changes undergone during the stations lifetime, which could reflect in dishomogeneities in the time series and relevant changes in the data not related to climatic factors (Aguilar et al., 2003; Acquaotta et al., 2009).

All the time series have been quality controlled in order to identify and eventually correct anomalous values and errors due to the observers or to the process of digitization. Temperatures time series used in this study have been quality checked using the RClimdex package (Zhang and Yang, 2004) which highlights unreasonable values. For daily snow depth data a procedure for the identification of the data outlying pre-fixed thresholds (the 99th percentile calculated on non-zero values) is applied to two derived from HS time series, the snow accumulation at and snow depletion dt in order to check the reliability of abrupt changes in snow thickness.

Formula for snow accumulation (at) and depletion (dt):

$$
\begin{array}{lclcl}
a(t) & = & HS(t) - HS(t\text{-}1) & \text{if} & HS(t) - HS(t\text{-}1) > 0 \\
 & & 0 & \text{if} & HS(t) - HS(t\text{-}1) \leq 0 \\
d(t) & = & HS(t) - HS(t{+}1) & \text{if} & HS(t) - HS(t{+}1) > 0 \\
 & & & \text{if} & HS(t) - HS(t{+}1) \leq 0
\end{array}
$$

The identification of the inhomogeneities in daily temperatures time series provided essential information for the homogenization which has been performed using the SPLIDHOM method (Mestre et al., 2011), implemented in the EU COST Action ES0601 “HOME”(Venema et al., 2012).

Daily values have been aggregated over monthly and seasonal time scales. These data have been retained only if at least 80% of the daily values were available (Klein Tank et al., 2002) and then a trend analysis has been performed by least square linear fitting (Brunetti et al., 2001). The significance of the trends has been evaluated with the non parametric Mann-Kendall test, at 95% confidence level (Sneyers, 1990).

4. TRENDS IN SNOW DEPTH AND SNOW PRECIPITATION

The analysis of trends has been conducted at different time scales: over the complete snow season November-May (NM) and over winter December-February (DJF) and spring March-May (MAM) period (Tab. 2).

In Table 2 the snow depth trends over the full period (left) and over the last 60 years (right) are compared. If we consider the full records, over the November-May snow season all the stations register a negative trend that is statistically significant in two cases: at Acceglio Saretto (1540 m), Southern Piedmont, snow depth decreased by -0.25 cm/y since 1925 and at Toggia (2165 m), Northern Piedmont, snow depth decreased by -0.96 cm/y since 1932. The other stations, located mainly at middle and low elevation, present a non-significant snow depth reduction.

Table 2: Seasonal (NM), winter (DJF) and spring (MAM) trends in snow depth (HS) over the maximum period available and over the 1951-2010 common period. The statistically significant trends (95% confidence level) are highlighted in bold

	Elev	Max Period	Trend HS Max Period [cm/seas]			Trend HS 1951-2010 [cm/seas]		
Station	**[m a.s.l.]**		**NM**	**DJF**	**MAM**	**NM**	**DJF**	**MAM**
Camposecco	2316	1951-2010	-	-	-	**-1.05**	**-1.12**	**-1.74**
Lago Vannino	2177	1951-2010	-	-	-	**-1.39**	**-1.82**	**-1.39**
Toggia	2165	1932-2010	**-0.96**	**-1.22**	-0.85	**-1.40**	**-1.74**	**-1.20**
Alpi Devero	1634	1951-2010	-	-	-	**-0.75**	**-0.78**	**-1.04**
Agaro	1600	1961-2010	-0.33	-0.17	-0.47	**-**	**-**	**-**
Lago Castello	1589	1943-2010	-0.15	-0.20	-0.14	**-0.27**	-0.33	**-0.28**
Ceresole Reale	1573	1926-2010	-0.15	-0.14	-0.17	**-0.32**	-0.34	**-0.33**
Acceglio Saretto	1540	1925-2010	**-0.25**	-0.26	**-0.30**	**-0.41**	-0.40	**-0.49**
Alpe Cavalli	1500	1932-2010	-0.30	-0.31	-0.29	**-0.58**	-0.62	**-0.63**
Lago Piastra	960	1926-2010	-0.06	-0.11	-0.01	**-0.19**	-0.23	**-0.18**

Focusing the analysis over the period 1951-2010, a general decrease of the average seasonal snow depth over Piedmontese Alps occurred. This negative trend is statistically significant in all the stations. The overall decrease ranges between -0.19 cm/y in the most Southern station (Lago Piastra, 960 m a.s.l.) and -1.40 cm/y in the most Northern station (Lago Toggia, 2165 m a.s.l.). The stations located in Northern Piedmont above 2000 m a.s.l. experienced the strongest snow depth reduction, contrary to expectations for high altitude sites. Also comparing middle elevation stations at about 1500-1600 m a.s.l. the decrease is strongest in Northern Piedmont, so it can be concluded that this area has been the most subject to the snow

thickness reduction in the last 60 years. A similar behavior is found over the French side, where negative snow depth trends are strongest in the Northern Alps (Durand et al., 2009). The analysis over the complete period and over the 1951-2010 subperiod puts in evidence that the main contribution to the negative snow depth trend is due to the last decades of the record. A strong snow depth decrease since mid 1980s has been documented also over the other slopes of Alps, in Switzerland (Laternser and Schneebeli, 2003), France (Durand et al., 2009) and Austria (Schoner et al., 2009).

A finer analysis at sub-seasonal scale has been performed to find out in which part of the snow season the snow depth reduction is strongest, whether in winter (DJF) or in spring (MAM). Below 1600 m winter and spring snow depth reduction are comparable, but the trends are significant only in the last part of the snow season. Above 1600 m winter snow depth decreased significantly both in winter and in spring, confirming that these stations, located mainly in Northern Piedmont are the most subjected to climatic change.

Considering the seasonal accumulated snow precipitation and the seasonal number of days with snow precipitation both over the maximum period and over the 1951-2010 sub-period (Table 3), in the majority of stations these parameters show negative trends in general not statistically significant. However it can be outlined that during the last decades the decrease in snowfall amount is largest below 2000 m a.s.l, so the mid and low elevation mountains receive less snow precipitation than in the past. According to these results, both over the full records and over the 1951-2010 period there is no clear evidence of a significant change in the number of days with snow precipitation, so it is unlikely that the observed decrease in snow depth is due to a change in the frequency of meteorological perturbations favourable to snow precipitation.

Table 3: Trends in seasonal (November-May) accumulated snowfall (HN) and number of snowy days (SD) over the maximum period available and over the 1951-2010 common period. The statistically significant trends (95% confidence level) are highlighted in bold

		Max Period	Trend Max Period		Trend 1951-2010	
Station	**Elev**	**years**	**HN [cm/y]**	**SD [days/y]**	**HN [cm/y]**	**SD [days/y]**
Camposecco	2316	1951-2010	-	-	-0.56	-0.04
Lago Vannino	2177	1951-2010	-	-	-1.54	-0.10
Toggia	2165	1932-2010	0.74	0.08	0.59	-0.02
Alpi Devero	1634	1951-2010	-	-	-0.37	0.02
Agaro	1600	1961-2010	**-0,51**	**-0.28**	-	-
Lago Castello	1589	1943-2010	-0.51	0.07	-0.85	0.03
Ceresole Reale	1573	1926-2010	-0.08	-0.08	-0.94	-0.08
Acceglio Saretto	1540	1925-2010	-0.73	-0.03	-1.31	-0.09
Alpe Cavalli	1500	1932-2010	-0.52	**-0.08**	-1.41	**-0.16**
Lago Piastra	960	1926-2010	-0.61	-0.02	**-1.92**	-0.05

For each station, the seasonal snow depth standardized anomalies have also been considered. All the stations time series have been averaged to get the regional Standardized Anomaly Index, which express the anomaly of the studied parameter respect to the mean value over a reference 30-year period, in this case 1971-2000.

During the last 25 years almost all seasons were characterized by negative anomalies. The SAI calculated on snow precipitation shows the same phases of positive snow depth anomalies. The snow precipitation has been below average along the 1940 till the mid 50's, during the 1960's and from the mid 80's up to 2008. In the last 25 years only 6 seasons

present positive anomalies, all the others had below average snow precipitation. In particular it has been registered the longest sequence of seasons with negative anomalies, from 1995 to 2003 (Fig. 4).

Snow precipitation shows fluctuations of irregular period of about one decade, with relative maxima around 1940, 1950, 1960. The most prominent peak is registered in the '70s, when average DJF HN was around 300 cm in all the tree stations. The maximum has been followed by an absolute minimum around 1990, with snowfall amount reduced by 1/3.

It is outlined a significant decrease of snow depth over seasonal time scale in the period 1951-2010. In all the sites a strong contribution to this negative trend comes from spring and in the Northern Alps the snow depth reduction is significant also in winter months. In the following section these findings are related to temperatures in order to explore the connections with the snow reduction.

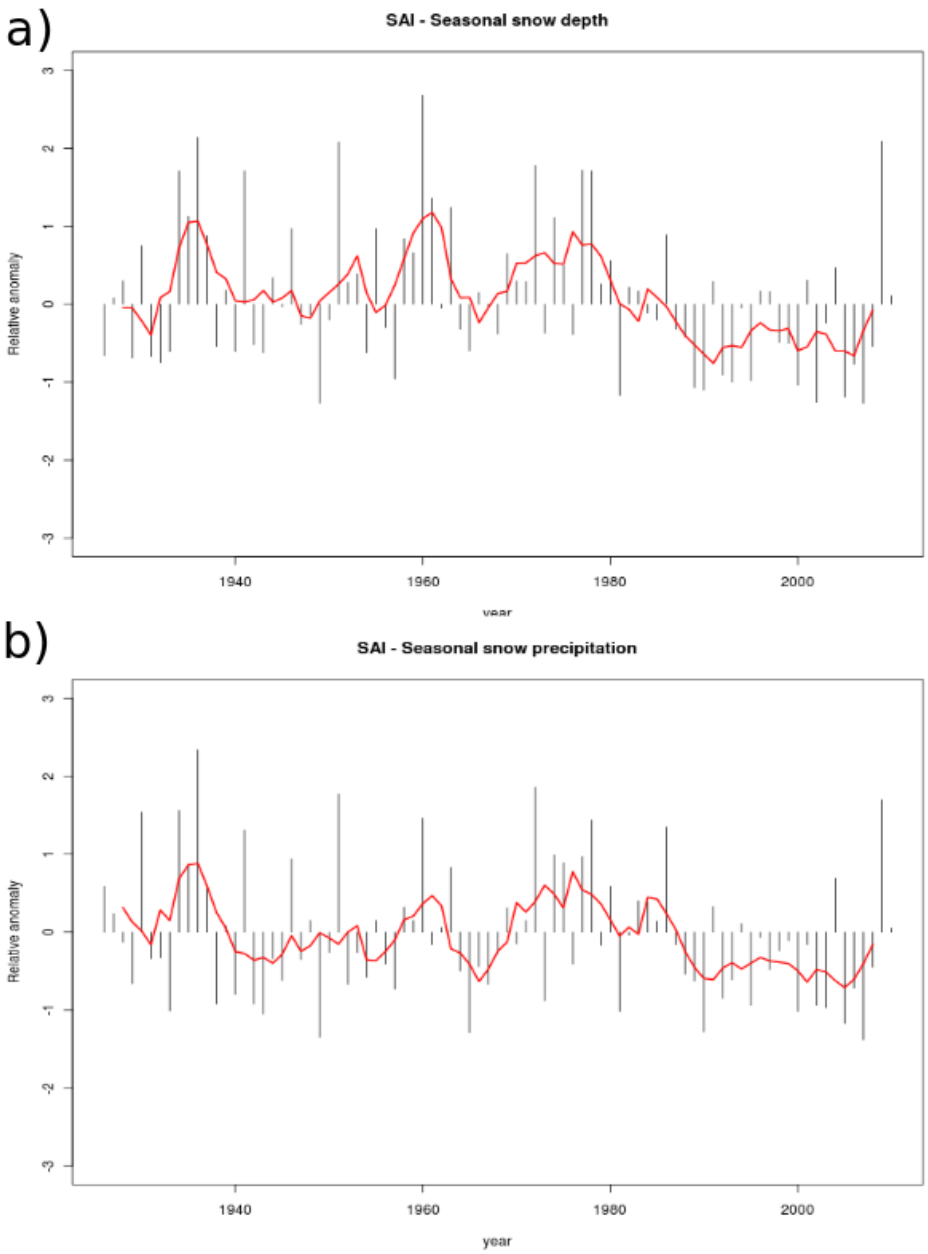


Figure 3: Temporal variability of seasonal average snow depth (a) and cumulated snowfall (b)

5. TRENDS IN MAXIMUM AND MINIMUM TEMPERATURES

The investigation on the changes in minimum and maximum temperatures time series shows increasing and statistically significant trends both in winter and in spring over the maximum period (Tab. 4). In winter the strongest increase in maximum temperatures (0.09 °C/year) is found in Acceglio Saretto while for minimum temperatures (0.07°C/year) in Camposecco. In spring the highest trend is detected in Lago Toggia: the maximum temperature increased by 0.05°C/year while the minimum temperature increased by 0.03°C/year. Concerning the number of ice and frost days, they decreased even though in the last case the trend is not statistically significant. In winter the highest ice day decrease (-0.48 day/year) is registered in Acceglio Saretto.

These results can be compared to those referring to the common period 1951-2010 reported as well in Table 4. In winter the maximum and minimum temperatures show increasing trends but not statistically significant while the trends in climate indices significantly decreased. The highest trend is calculated for the number of ice days in Acceglio Saretto (-0.65 days/year).

In spring the maximum and minimum temperature trends are positive and statistically significant while the climate indices trends are negative and statistically significant. The maximum temperature increase is registered in Acceglio Saretto (0.08°C/year) while the highest ice days decrease is identified in Lago Toggia (-0.3 days/year).

Table 4: Winter (DJF) and spring (MAM) trends in maximum (Tmax) and minimum (Tmin) temperature, frost day (FD)and ice day (ID) over the maximum period available and over the 1951-2010 common period. The statistically significant trends (95% confidence level) are highlighted in bold

Station	Elevation [m a.s.l.]	Period	Seasonal	Tmax	Tmin	FD	ID
Camposecco	2316	1951-2010	DJF	0.01	**0.07**	-0.03	-0.15
			MAM	**-0.06**	**0.07**	**-0.27**	**0.2**
Lago Vannino	2177	1951-2010	DJF	**0.02**	**0.04**	-0.01	-0.11
			MAM	**0.04**	**0.06**	**-0.25**	**-0.21**
Lago Toggia	2165	1932-2010	DJF	**0.02**	**0.03**	-0.0004	**-0.22**
			MAM	**0.05**	**0.03**	**-0.13**	**-0.23**
		1951-2010	DJF	**0.02**	0.02	9.3^e-4	**-0.24**
			MAM	**0.06**	**0.03**	**-0.19**	**-0.3**
Alpi Devero	1634	1951-2010	DJF	0.003	**-0.01**	0.03	-0.12
			MAM	**-0.05**	-0.02	**0.28**	**0.09**
Acceglio Saretto	1549	1930-2010	DJF	**0.09**	**0.04**	**-0.1**	**-0.48**
			MAM	**0.03**	0.01	**-0.11**	**-0.02**
		1951-2010	DJF	**0.1**	**0.04**	**-0.1**	**-0.65**
			MAM	**0.08**	**0.06**	**-0.4**	**-0.07**
Alpe Cavalli	1500	1937-2010	DJF	-0.003	0.013	0.04	-0.03
			MAM	**0.03**	**0.01**	0.07	-0.02
		1951-2010	DJF	-0.01	**0.02**	-0.01	0.03
			MAM	**0.05**	**0.03**	-0.08	**-0.05**

6. CONCLUSIONS

The present study gives a contribution to the assessment of the temporal and spatial variability of the climatic conditions at high elevation sites in Western Italian Alps since 1925. In accordance to the MEDARE initiative several historical daily snow depth and temperatures

time series have been recovered from the paper archives of several Institutions. The time series are representative of all the Western Italian Alps and they allow to explore a range of altitude between 900 and 2300 m a.s.l., spanning a 85 year period, from 1925 up to 2010.

The investigation on temporal evolution of snow precipitation shows HN fluctuations with an irregular period of about one decade, with relative maxima around 1940, 1950, 1960, and the absolute maximum in the '70s. Snow was abundant till the early 80's, then a sequence of poor snow winter leads to the absolute minimum of the record in the 1990's. A similar investigation has been conducted over the Switzerland (Laternser et Schneebeli, 2003) considering individual winters and the results are comparable between the two sides of Alps.

This study outlines a significant decrease of snow depth in all the stations over seasonal (November to May) time scale in the period 1951-2010. The main contribution to this negative trend comes from spring when the snow depth decrease is comparable to the one registered over the complete season. In particular November-May trends range from -0.19 cm/y in the lowest (960 m a.s.l.) and most Southern station up to -1.40 cm/y at 2177 m a.s.l. in the most Northern station. At comparable elevation, around 1500 m a.s.l. the strength of the snow depth decrease grows moving Northward. In addition to the strongest snow depth reduction in spring, the Northern stations suffer also a significant decrease in winter months, so North Piedmontese Alps result the most sensitive to climate change. This important result will allow to better investigate the relationship between climatic change and avalanches, the main goal of the project interreg STRADA.

The shallower winter and seasonal snow depth can be explained by an increasing in temperatures values. The raise in daily temperature anomalies is particularly evident for winter months and it ranges from +0.02°C/year to 0.07°C/year for minimum temperatures and from 0.003 °C/year to 0.10°C/year for maximum temperatures.

The combination of the higher -especially minimum- temperatures and minor number of frost days make the high elevation sites in Northern Piedmont the most vulnerable to snow depth reduction to climatic change. Low elevation sites (below 1000 m a.s.l.) are mainly affected by scarcer snow precipitation due to higher temperatures and consequent predominance of liquid respect to solid precipitation.

In conclusion, a general increase in maximum and minimum temperatures has been found in high elevation sites in Piedmontese Alps during the snow season November-May. The increase of temperatures together with the decrease in the number of ice and frost days enhance the snow melt and may reduce the frequency and amount of snow precipitation at low elevation. At high elevations, the main consequence is the faster ablation of the snow pack, with consequent snow depth reduction and decrease in snow cover persistence.

Acknowledgements

We gratefully acknowledge the valuable help Marco Cordola (Civil Protection Piedmont Region) in recovering and collecting some meteorological data and the patient work of digitization carried out by Silvia Brunatti, Daniela Testa, Noemi Canevarolo, Silvia Cadorna and Matteo Collimedaglia. Finally thanks to Roberto Cremonini of Arpa (Regional Agency for Environmental Protection) Piedmont for the fruitful discussions risen during the work.

References

Acquaotta F, Fratianni S, Cassardo C, Cremonini R. 2009. On the continuity and climatic variability of the meteorological stations in Torino, Asti, Vercelli and Oropa. Meteorology and Atmospheric Physics 103: 279-287.

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidance on metadata and homogenization. WMO-TD No. 1186. Available online: http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/index_en.html.

Beniston M, Diaz HF, Bradley RS. 1997 (a). Climatic change at high elevation sites: an overview. Climatic Change 36: 233-251.

Beniston M. 1997 (b). Variations of snow depth and duration in the Swiss Alps over the last 50 years: links to changes in large-scale climatic forcings. Climatic Change 36: 281 – 300.

Biancotti A, Carotta M, Motta L, Turroni E. 1998. Le precipitazioni nevose sulle Alpi piemontesi. Trentennio 1966-1996. Collana Studi Climatologici in Piemonte 2: 11-80.

Brunetti M., Colacino M., Maugeri M. and Nanni T. 2001. Trends in the daily intensity of precipitation in Italy from 1951 to 1996 Int. J. Climatol. vol 21: pp 299-316.

Durand Y, Giraud G, Laternser M, Etchevers P, Mérindol L, Lesaffre B. 2009. Reanalysis of 47 Years of Climate in the French Alps (1958–2005): Climatology and Trends for Snow Cover. Journal of Applied Meteorology and Climatology 48: 2487–2512.

Fazzini M, Gaddo M. 2003. La neve in trentino. Analisi statistica del fenomeno nell'ultimo ventennio. Neve e Valanghe, 48: 28-35.

Hantel M, Ehrendorfer M, Haslinger AM. 2000. Climate sensitivity of snow cover duration in Austria. International Journal of Climatology 20, 615-640.

Hantel M, Hirtl-Wielke LM. 2007. Sensitivity of Alpine snow cover to European temperature. International Journal of Climatology 27: 1265–1275.

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, Antonio López J, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P. 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology 22: 1441–1453. doi: 10.1002/joc.773.

Laternser M, Schneebeli M. 2003. Long-term snow climate trends of the Swiss Alps (1931-99). International Journal of Climatology 23, 733-750.

Martin E, Brun E, Durand Y. 1994. Sensitivity of the French Alps snow cover to the variation of climatic variables. Annales Geophysicae 12: 469-477.

Quadrelli R, Lazzari M, Cacciamani C, Tibaldi S. 2001. Observerved winter Alpine precipitation variabilità and links with large – scala circulation patterns. Climate Research, 17: 275 -285.

Scherrer SC, Appenzeller C, Laternser M. 2004. Trends in Swiss Alpine snow days: the role of local and large scale climate variability. Geophysical Research Letters 31, L13215.

Schöner W, Auer I, Böhm R. 2009. Long term trend of snow depth at Sonnblick (Austrian Alps) and its relation to climate change. Hydrological Processes, 23: 1052–1063.

Zhang X, Yang F. 2004. RClimDex User Manual. Climate Research Branch, Environment Canada: Downsview Canada Available online http://cccma.seos.uvic.ca/ETCCDMI/software.shtml.

Sneyers R. 1990. On the statistical analysis of series of observation. Technical Note NO. 143. WMO, GENEVA.

Terzago S, Cremonini R, Cassardo C, Fratianni S. 2012. Analysis of snow precipitation during the period 2000–2009 and evaluation of a snow cover algorithm in the South Western Italian Alps. Geografia Fisica e Dinamica Quaternaria, in press.

Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S. 2011. SPLIDHOM: A method for homogenization of daily temperature observations. American Meteorological Society, 50, 2343-2358.

Venema V K C, Mestre O, Aguilar E, Auer I, Guijarro J A, Domonkos P, Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Müller-Westermeier G, Lakatos M, Williams C N, Menne M J, Lindau R, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquaotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P, and Brandsma T. 2012. Benchmarking monthly homogenization algorithms. Clim. Past, 8, 89-115.

DATA HOMOGENIZATION IN THE CARPATCLIM PROJECT

Lakatos, M.[1], Szentimrey, T.[1], Bihari, Z.[1], Szalai, S.[2]
[1]Hungarian Meteorological Service
[2]Szent István University

Abstract

The main aim of the CARPACLIM project is to improve and extend the climate data basis in the Carpathian Region for applied regional climatological studies such as a Climate Atlas and drought monitoring. The project investigates the temporal and spatial structure of the climate in the Carpathian Mountains and the Carpathian basin with unified or at least directly comparable methods. The final outcome of the CARPATCLIM project will be high quality climate in-situ time series and 10x10 km grids per country and the whole region, including a metadata catalogue documenting the existing homogenised datasets. Quality control and homogenisation procedures, together with corresponding metadata of the improved datasets are required to be published without limitation. The commonly used homogenisation and DQ software is the MASH (Multiple Analysis of Series for Homogenization) has been developed by Tamás Szentimrey at the Hungarian Meteorological Service.

1. INTRODUCTION

Joint Research Centre (Ispra) published a tender call supported by the European Parliament in June 2010. A consortium of 10 organisations won and started its work by 22 December 2010. The project contains three modules with the following tasks: Module 1: Improve the availability and accessibility of a homogeneous and spatially representative time series of climate data for the Carpathian Region through data rescue, quality control, and data homogenisation. Module 2: Ensure Carpathian countries data harmonisation with special emphasis on a cross-country harmonisation and production of gridded climatologies per country. Module 3: Develop a Climate Atlas as a basis for climate assessment and further applied climatological studies as well as for drought monitoring in the Carpathian Region. The methodology of the data homogenization in Module 1 is introduced in this paper.

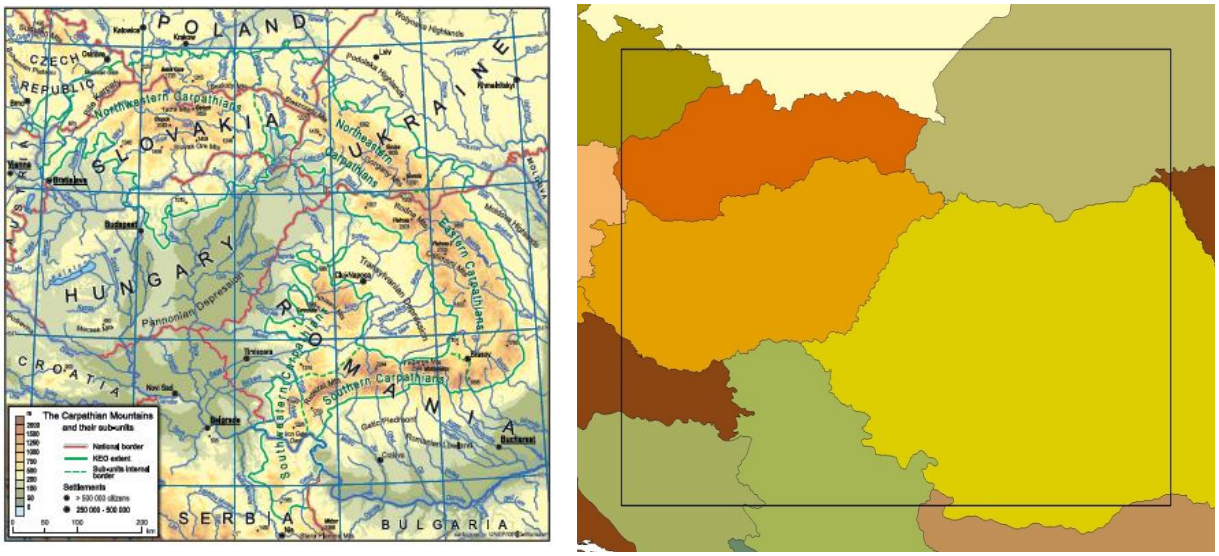


Figure 1. The area of interest area between latitudes 50°N and 44°N, and longitudes 17°E and 27°E, approximately partly including the territory of Czech Republic, Slovakia, Poland, Ukraine, Romania, Serbia, Croatia, Austria and Hungary

2. THE METHODOLOGY OF HOMOGENIZATION

The work in progress in Module 1 are the following: - Data rescue and digitisation of analogue datasets of climate observations - Quality checking including data gap elimination of existing climate time series - Homogenization of existing climate time series The MASH (Multiple Analysis of Series for Homogenization, Szentimrey) procedure was proposed for data quality control and data homogenization in the tender service. For ensuring the largest possible station use, each contributor executes the necessary work phases individually but by the same methods and software. It assures the national consistency. The international harmonization will be guaranteed by near border data exchange. The use of the near border data of the neighbouring countries will assure the smooth fitting of individual national maps in the region.

The MASHv3.02 software has been already used in several countries in the Carpathian Region. At the Hungarian Meteorological Service it was applied for numerous climate elements. The chosen homogenization model is depending on the distribution type of the element in question. Additive model: mean, maximum and minimum temperature, air pressure, means relative humidity, sunshine duration. Multiplicative, cumulative model: precipitation, minimum relative humidity, wind speed. The original MASH (Szentimrey, 1999) procedure was developed for homogenization of monthly series. The time resolution of the data specified in the Service Tasks is daily. The present version: MASHv3.02 (Szentimrey, 2007) has been developed for homogenization of daily series, besides the monthly data.

The homogenization, the data quality control and the data completion can be made on national level, implemented by the common software MASH. Between the neighbouring countries the near border station data series is exchanged in order to cross-border harmonization.

3. THE MAIN ADVANTAGES OF APPLYING MASH

- In general the daily data homogenization is based on the monthly data homogenization, according to the several methods applied in the practice. The benchmark results show that the MASH procedure is one of the best monthly homogenization methods examined in COST HOME action (Venema et al, 2012)
- MASH is an automatic procedure, and it is worth considering if the station network is dense, or numerous stations have to be examined
- MASH is able to use the metadata (namely the possible dates of break points) automatically. It allows the effective usage of the metadata
- During the execution of the quality control and homogenization the test results e.g. detected errors, degree of inhomogeneity, number of break points, estimated corrections and certain verification results are documented in automatically generated tables
- Information of quality control and homogenization processes performed can be tracked on these tables. The summary of them can be added to the homogenized series as the newly created metadata.

Table I. Set of meteorological variables in daily temporal resolution to be homogenized

Variable	Description	Units
Ta	2 m mean daily air temperature	°C
Tmin	Minimum air temperature from 18:00 to 06:00	°C
Tmax	Maximum air temperature from 06:00 to 18:00	°C
p	Accumulated total precipitation from 06:00 to 06:00	mm
DD	10 m wind direction	0°-360°
VV	10 m horizontal wind speed	m/s
Sunshine	Sunshine duration	hours
cc	Cloud cover	tenths
Rglobal	Global radiation	MJ/m2/day
RH	Relative humidity	%
pvapour	Surface vapour pressure	hPa
pair	Surface air pressure	hPa
Snow depth	Snow depth	cm

4. THE MAIN STEPS FOR CREATION OF THE STATION DATA SERIES

I. Compilation of the raw station data series.

1. Selection of the stations (with the spherical coordinates: φ, λ), determination of the time period.
2. Collecting the daily station data series (missing data are allowed) and the metadata per countries. Exchange of the near border station data series and the metadata between the neighbouring countries.

II. Homogenization, quality control, data completion for the station data series by MASH on national level, with using the near border data.

1. Derivation of monthly station data series from the daily station data series collected in I.2. Homogenization, quality control, data completion of the monthly station data series. Metadata (probable dates of break points) can be used automatically.
2. Daily station data series (I.2): homogenization, quality control, data completion. This procedure is based on the results of II.1.
3. Exchange of the near border homogenized data for cross-border harmonization and for gridding (Module 2: modelling, interpolation).
4. Evaluation of the results of the homogenization and quality control was carried out. Controlling of the cross-border harmonization of the data series. The cross-border harmonization will be continued after modelling procedure of gridding in Module 2.

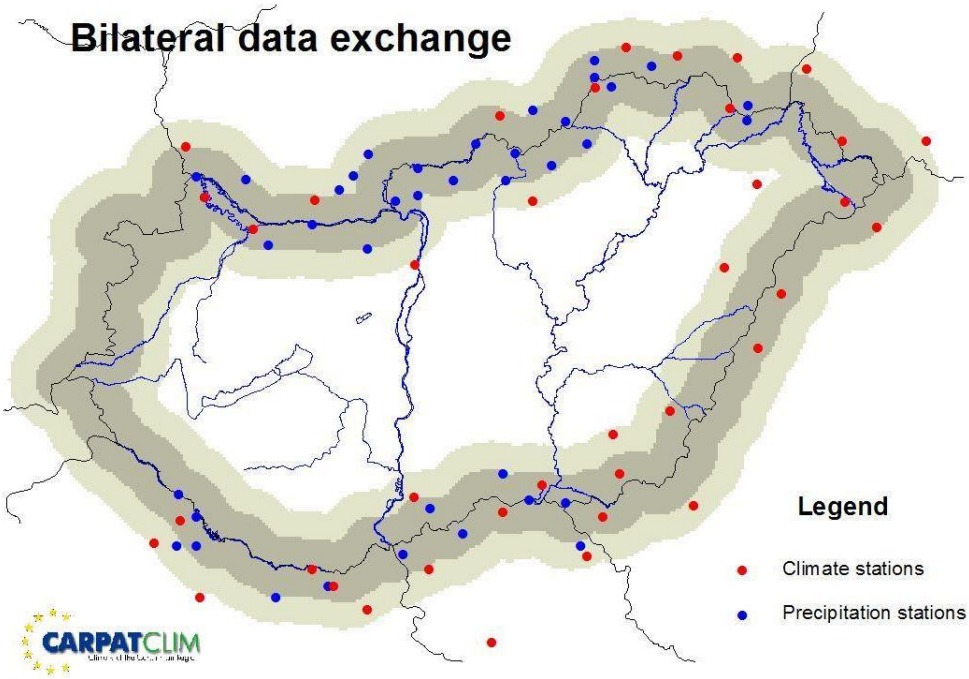


Figure 2. The scheme of the data exchange around the Hungarian border

5. THE PROCEDURE FOR HOMOGENIZATION OF DAILY SERIES BY MASHV3.03

The main steps of homogenization of daily series with quality control and missing data completion are as follows.

1. Monthly series from daily series.
2. MASH homogenization procedure for monthly series, estimation of monthly inhomogeneities. Meta data can be used automatically
3. On the basis of estimated monthly inhomogeneities, smooth estimation for daily inhomogeneities.
4. Automatic correction of daily series.
5. Automatic quality control (QC) for homogenized daily data.
6. Automatic missing daily data completion.
7. Monthly series from homogenized, controlled, completed daily data.
8. Test of homogeneity for the new monthly series by MASH.

During the procedure the results as detected break points and errors, estimated corrections and certain verification test results are documented in automatically generated tables.

Two types of the result files of the MASH procedure can be differentiated:.

A: The first type is output files of the homogenized, controlled and completed series, inhomogeneity series, detected breaks and detected errors.

B: The second type is output files of the test results, verification tables in order to evaluate the homogenization procedure:

- In the verification tables: test statistics before and after homogenization, statistics to characterize the modification of series, the spatial representativity of the station network, evaluation of metadata etc.
- Tables of quality control results for daily data.

6. CONCLUSION

Applying the MASH procedure allows to perform the data homogenization quality control and data completion for the Carpathian region for numerous stations and elements during the tender service as it is quasi automatic, mathematically well established procedure. The members of consortium execute the homogenization individually with the common software MASH Daily, it assures the national consistency. The international harmonization will be guaranteed by near border data exchange. The automatic test results organized in tables allow measuring the inhomogeneities of the meteorological observations in the region before and the data quality improvement after the homogenization. The participants of the project started to use the software with the manual (Szentimrey, 2011) furthermore they have already tested it on their data series and the preliminary results have been obtained and delivered to the JRC. The CARPATCLIM project assists the availability of a set of homogeneous and spatially representative time series of climate data for the Carpathian Region through data rescue, quality control, and data homogenisation.

References

JRC, 2010: Climate of the Carpathian Region. Technical Specifications (Contract Notice OJEU 2010/S 110-166082 dated 9 June 2010). http://www.jrc.ec.europa.eu/callsfortender

Hungarian Meteorological Service, 1997: Proceedings of the first seminar for Homogenisation of Surface Climatological Data, Budapest, 6-12 October 1996. Budapest.

Hungarian Meteorological Service. 2001: Proceedings of the third seminar for Homogenisation and Quality Control in Climatological Databases, Budapest, September 2000. Budapest.

Szalai, S., 2011: Climatological establishment of site studies and land use: the Carpathian region, Növénytermelés, Vol. 60, 2011, Suppl. in, DOI:0.1556/Novenyterm.60.2011.Suppl.1

Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH), Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2011: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, pp. 64.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89-115, doi:10.5194/cp-8-89-2012, 2012.

HOMOGENIZATION OF LONG MONTHLY AIR TEMPERATURE TIME SERIES COLLECTED IN UKRAINE

Osadchy V.[1], Nabyvanets Y.[1], Skrynyk O.[2], Krakovska S.[1] and Skrynyk O.[1]
[1] Ukrainian Hydrometeorological Institute, Kyiv, Ukraine
[2] National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
krava@uhmi.org.ua, skrynyk@univ.kiev.ua

Abstract

Climate study is not possible without well proved and qualitative empirical data - time series of meteorological parameters. Term "well proved" means that time series are homogeneous; they do not have a lot of missed values (gaps) and outliers. So, the first step of any climatological investigations should be checking of time series quality that can be done by using of MASH software (Dr. Tamás Szentimrey)

The goal of this study is to present first results of using of MASH software in Ukrainian Hydrometeorological Institute. The MASH procedure was used to perform quality control and to homogenize very long monthly air temperature time series collected in Ukraine during time period of 1881 – 2007. We used observations data of 38 Ukrainian stations that were uniformly distributed on the territory of Ukraine. An average distance between stations was about 100 km. There was some amount of gaps in practically all stations except Kyiv, where there were no missed values. Percentage of gaps was different for different stations with limit of 20 %.

Territory of Ukraine was divided into 6 regions with approximately similar climatological conditions within. Homogenization procedure was performed for each region separately.

We used an algorithm of MASH running according to new version of MASH Manual (2011) and performed two running of MASH for each region. The first run was performed without metadata while during the second one metadata have been used. Comparison of the obtained results shows that MASH software defines break points very well.
In general, revealed inhomogeneties in the monthly air temperature time series collected on the Ukrainian territory were less than 1^0C, which means that they are rather good and may be used for climatological studies.

1. INTRODUCTION

Well proved and qualitative empirical data, i.e. time series of meteorological parameters, is an essential background for climate study. Term "well proved" means that time series are homogeneous; without missed values (gaps) and outliers. Therefore, the first step of any climatological studies should be the checking of time series quality and homogenization. Such an assessment of time series can be performed by using MASH software (Szentimrey, 1999; Szentimrey, 2011). According to the results of benchmarking conducted in the framework of COST Action HOME (Venema et al, 2012) this homogenization procedure is one of the best.

The objective of this study is to discuss results of using MASH software in Ukrainian Hydrometeorological Institute (UHMI). The MASH procedure was used to perform quality control and to homogenize very long monthly air temperature time series collected in Ukraine during time period of 1900 – 2007. This period was chosen because some work regarding study of regional climate change during the last century is planned in UHMI for the nearest future.

2. DESCRIPTION OF THE TEMPERATURE DATASETS EMPLOYED

At the present Ukrainian meteorological network includes 188 meteorological (climatological) stations and about 370 hydrological posts where precipitations are measured (Vyshnevsky and Tokar, 1998). 114 Ukrainian stations are presented at the WMO catalog of observation stations.

There are 2 stages of meteorological data quality control (QC) in Ukraine. At the first stage information passes internal observation station quality control procedure. During the second one, meteorological information goes under QC performed in Meteorological Division of Central Geophysical Observatory of the Ministry of Emergencies of Ukraine. Finally, meteorological information is published in special tables which guaranties that it is rather qualitative and reliable. But we should mention that checking of homogeneity of temperature data using up-to-date homogenization procedures has not been performed.

Observations data of 33 climatological stations uniformly located on the territory of Ukraine were used for this study. Stations with minimal amount of missed values, maximal period of observations and providing more reliable data have been selected. Schematic map displaying spatial distribution of the employed climatological stations is presented in Fig. 1. An average minimal distance between stations was about 100 km. The altitude of the stations ranges from 22 to 327 m.

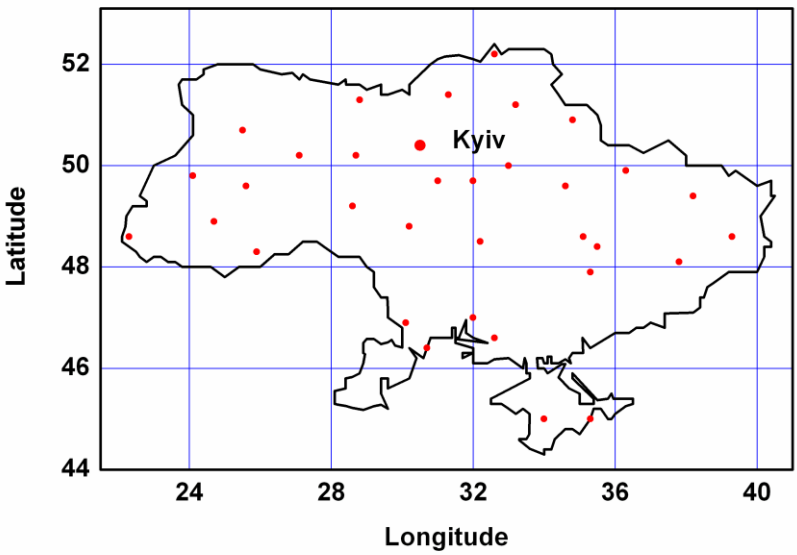


Figure 1. Ukrainian climatological stations considered in the study

Practically all stations except station Kyiv were characterized with certain amount of gaps in data series. Percentage of gaps was different for different stations with limit of 20 %. More detailed information regarding employed stations is presented in Table 1.

Historical description of Ukrainian stations is published in the special issues (Climatological handbook USSR, 1968; Climatological handbook, 2011) which served as sources of possible break points (metadata). Most of Ukrainian stations were relocated several times; some of them were relocated at very long distance (about 20 km). Besides, measuring methodology for temperature was changed at all Ukrainian climatological stations. In 1966 four time observations were replaced with eight times observations. We should note that some

correction for monthly air temperature for a period with four time observations was proposed and implemented in published tables.

3. USED MASH ALGORITHM

MASH software was applied according to the algorithm shared among participants of CARPATCLIM project by T. Szentimrey. The algorithm is quasi automatic. However, in our case using of the algorithm was not enough because test statistics (TS) still remained quite high. Therefore, according to the MASH Manual some additional steps were used.

Firstly, we used MASH procedure without metadata in order to assess the efficiency of this software in break points detection. After this we run the MASH software second time with collected metadata. According to the recommendations of Venema et al (2012) the results of last running could be considered as most accurate. Namely these results will be taken for further climatological studies.

Table 1. Detailed information concerning Ukrainian climatologocal stations involved

MASH index	Name	ID	Coordinates		Height, m	Beginning of observations	Gaps (1900-2007), %
			Lat	Long			
1	Semenivka	33049	52,2	32,6	160	1927	17
2	Chernigiv	33135	51,4	31,3	139	1883	13
3	Ovruch	33213	51,3	28,8	168	1894	12
4	Konotop	33261	51,2	33,2	144	1923?	11
5	Sumy	33275	50,9	34,8	180	1896	5
6	Lutsk	33187	50,7	25,5	233	not known exact.	20
7	Kyiv	33345	50,4	30,5	166	1812	0
8	Zhytomyr	33325	50,2	28,7	219	1886	2
9	Shepetivka	33317	50,2	27,1	277	not known exact.	15
10	Lubny	33377	50,0	33,0	156	1892	2
11	Kharkiv	34300	49,9	36,3	154	1892	2
12	Lviv	33393	49,8	24,1	319	not known exact.	5
13	Myronivka	33466	49,7	31,0	151	1913	20
14	Zolotonosha	33484	49,7	32,0	94	1895	2
15	Ternopil	33415	49,6	25,6	327	not known exact.	18
16	Poltava	33506	49,6	34,6	160	1886	2
17	Vinnytsya	33562	49,2	28,6	296	1908	14
18	Ivano-Frankivsk	33526	48,9	24,7	270	1887	19
19	Uman	33587	48,8	30,2	214	1885	4
20	Uzhhorod	33631	48,6	22,3	115	1872	18
21	Dnipropetrovsk	34504	48,6	35,1	141	1839	4
22	Lugansk	34523	48,6	39,3	59	1836	0
23	Kirovograd	33711	48,5	32,2	170	1874	2
24	Synelnykove	34505	48,4	35,5	147	1915	19
25	Chernivtsi	33658	48,3	25,9	242	not known exact.	10
26	Donetsk	34519	48,1	37,8	224	1926	16
27	Zaporizhzhya	34601	47,9	35,3	59	not known exact.	13
28	Mykolaiv	33846	47,0	32,0	49	1801?	7
29	Rozdilna	33834	46,9	30,1	146	1924	16
30	Kherson	33902	46,6	32,6	47	1825	8
31	Odesa	33835	46,4	30,8	42	1839	1
32	Simferopol	33946	45,0	34,0	180	1821	5
33	Feodosiya	33976	45,0	35,4	22	1870	1

4. RESULTS OF HOMOGENIZATION

Firstly, we would like to present homogenization results without taking into account metadata. We should mention that the degree of inhomogeneity of considered time series was rather high. The average value of TS before homogenization exceeds the critical value several times even in average (see second part of Table 2).

Table 2. The homogeneity test results for year series (MASHVERO.RES)

VERIFICATION OF HOMOGENIZATION (Ordered Statistics)
I. TEST STATISTICS FOR SERIES INHOMOGENEITY
Null hypothesis: the examined series are homogeneous.
Critical value (significance level 0.05): 21.73

1. Test Statistics After Homogenization

Series	Index	TSA	Series	Index	TSA	Series	Index	TSA
3352	18	60.31	3313	2	59.37	3304	1	51.93
3346	13	51.61	3332	8	50.16	3321	3	46.15
3384	28	42.82	3390	30	42.07	3331	9	41.70
3365	25	40.97	3451	26	40.40	3394	32	40.20
3460	27	38.47	3397	33	33.61	3450	21	32.59
3371	23	32.10	3450	24	31.83	3356	17	31.55
3452	22	30.79	3339	12	28.16	3383	29	28.02
3318	6	27.44	3430	11	26.52	3337	10	22.27
3383	31	20.67	3348	14	20.36	3327	5	20.20
3326	4	20.19	3358	19	19.13	3350	16	18.83
3334	7	17.76	3363	20	17.68	3341	15	15.61
AVERAGE:		**33.38**						

2. Test Statistics Before Homogenization

Series	Index	TSB	Series	Index	TSB	Series	Index	TSB
3450	21	950.98	3339	12	833.91	3390	30	568.78
3334	7	566.12	3331	9	547.36	3341	15	542.41
3356	17	436.55	3451	26	408.96	3352	18	368.82
3384	28	352.27	3304	1	344.21	3371	23	315.76
3318	6	307.45	3313	2	246.17	3326	4	239.66
3332	8	167.23	3450	24	157.60	3460	27	155.51
3346	13	139.68	3452	22	134.71	3363	20	125.57
3337	10	124.84	3327	5	118.17	3394	32	115.87
3430	11	109.51	3383	31	100.09	3358	19	96.64
3350	16	95.73	3397	33	94.70	3321	3	87.32
3365	25	66.97	3348	14	66.80	3383	29	61.24
AVERAGE:		**274.17**						

Table 3. Homogenization results for all series together (the Verisum file)

Index 1:	3451	26	87.74	3450	21	43.00	3460	27	35.54
Index 2:	3451	26	81.66	3460	27	70.59	3326	4	63.08
Index 3:	3352	18	51.67	3450	24	47.05	3313	2	46.60
Index 4:	3430	11	89.84	3332	8	79.71	3451	26	71.40
Index 5:	3390	30	77.56	3450	24	57.54	3384	28	56.41
Index 6:	3348	14	90.41	3341	15	90.35	3352	18	58.16
Index 7:	3451	26	84.97	3326	4	71.96	3394	32	63.87
Index 8:	3352	18	55.15	3358	19	46.47	3450	24	38.58
Index 9:	3341	15	72.06	3352	18	56.61	3371	23	48.79
Index 10:	3451	26	79.60	3346	13	68.12	3383	29	59.22
Index 11:	3450	24	52.91	3460	27	52.02	3326	4	48.88
Index 12:	3384	28	97.18	3390	30	89.75	3352	18	66.28

Index 13:	3451	26	68.75	3326	4	58.01	3337	10	56.46
Index 14:	3450	24	67.06	3450	21	52.66	3327	5	51.89
Index 15:	3326	4	79.79	3321	3	69.41	3358	19	60.90
Index 16:	3346	13	45.58	3358	19	44.70	3326	4	43.33
Index 17:	3326	4	70.59	3327	5	61.93	3397	33	50.13

The test statistics proximate the critical value after homogenization. This is also shown in the Table 2 where the results of verification procedure for year series are presented. Table 2 is a part of the file MASHVERO.RES.

The MASH procedure reduced substantially the test statistics also in the case of monthly series. Homogenization results for all series together are presented in Table 3. This is the file "Verisum".

In order to assess the efficiency of MASH software in break points detection we have compared MASH results with available metadata. However, it should be mentioned that our metadata is not complete. We have mainly taken into account only changing in location of stations. Approximately 30% of detected break points can be explained by our metadata (Fig. 2). We believe that this is very good result. We have to emphasize that MASH works very well in break detection especially when changes were considerable: all big changes were detected by MASH.

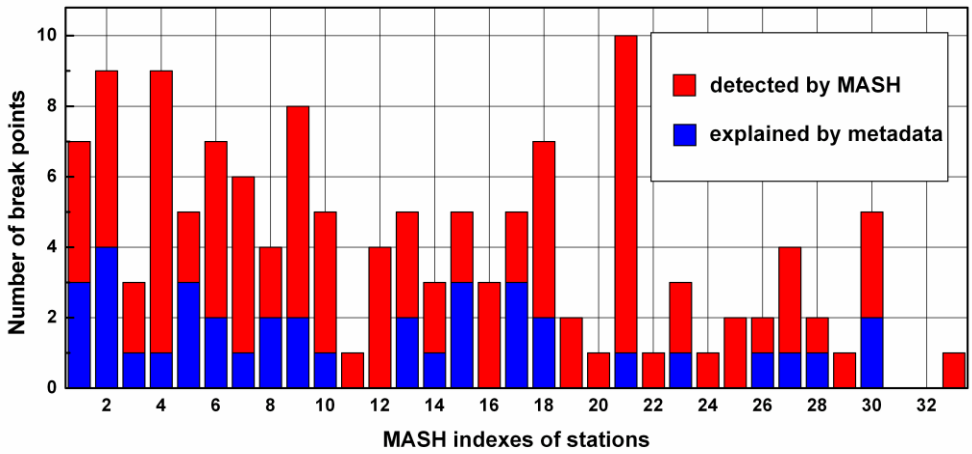


Figure 2. Number of break points detected by MASH and explained by metadata

Taking into account mentioned above we can conclude that executing of MASH with available metadata will not change the homogenization results considerably. Our calculations approve this conclusion.

5. EFECT OF REVEALED INHOMOGENEITY ON LINEAR TRENDS IN CONSIDERED TEMPERATURE TIME SERIES

Climate change and variability, especially on regional scale, is very important and interesting climatilogical issue. It is also important to know if revealed inhomogeneities influence conclusion regarding regional climate change. As it was mentioned above this will be an object of a special study in UHMI. Here we present some tentative results.

Calculation of time series linear trends coefficients is the simplest way to assess climate changes. According to previous findings of Ukrainian climatologist (Climate of Ukraine, 2003) the most considerable changes of regional climate were observed for the end of winter and beginning of spring. Based on those conclusions, the time series for March and calculated coefficients of linear trends (k) for original (with filled gaps) and homogenized series have been selected for our study. Results are presented in Table 4. Besides coefficients of linear trends the values of the determination coefficients are presented (r^2) in the table as well.

Table 4. Linear trends of original and homogenized time series for March

MASH index	Station	Original data		Homogenized data	
		k, °C/100 year	r^2, %	k, °C/100 year	r^2, %
1	Semenivka	2.17	5.4	2.22	5.7
2	Chernigiv	2.01	4.9	2.04	5.0
3	Ovruch	1.49	3.0	1.65	3.6
4	Konotop	2.19	5.5	2.05	4.9
5	Sumy	2.01	4.7	1.92	4.4
6	Lutsk	0.63	0.5	0.94	1.1
7	Kyiv	2.16	5.9	2.04	5.3
8	Zhytomyr	1.18	1.8	1.33	2.3
9	Shepetivka	0.94	1.2	0.95	1.2
10	Lubny	2.07	5.3	2.05	5.2
11	Kharkiv	2.12	5.4	2.21	5.9
12	Lviv	**-0.07**	0.0	**0.88**	1.1
13	Myronivka	1.73	3.6	1.68	3.4
14	Zolotonosha	1.81	3.9	1.79	3.8
15	Ternopil	0.37	0.2	0.80	0.8
16	Poltava	2.25	6.0	2.32	6.4
17	Vinnytsya	0.97	1.2	1.13	1.6
18	Ivano-Frankivsk	0.41	0.2	0.51	0.4
19	Uman	1.74	3.7	1.57	3.1
20	Uzhhorod	0.44	0.3	0.54	0.5
21	Dnipropetrovsk	1.01	1.4	1.82	4.6
22	Lugansk	2.13	5.7	1.48	2.8
23	Kirovograd	0.67	0.6	1.51	3.1
24	Synelnykove	1.78	4.1	1.69	3.8
25	Chernivtsi	0.59	0.5	0.63	0.5
26	Donetsk	1.89	5.0	1.54	3.4
27	Zaporizhzhya	1.53	3.5	1.48	3.3
28	Mykolaiv	0.79	1.1	1.26	2.7
29	Rozdilna	1.06	1.8	1.19	2.2
30	Kherson	0.46	0.4	1.07	2.1
31	Odesa	1.43	4.2	1.43	2.1
32	Simferopol	0.04	0.0	0.04	0.0
33	Feodosiya	0.40	0.4	0.71	1.1

Coefficient of linear trend changes its sign at only one station (Lviv). At other stations changes are not so considerable. There are two stations where linear trends are the same for both original and homogenized time series (Odesa and Simferopol). This is in agreement with homogenization results: exactly at these stations break points were not detected.

6. CONCLUSION

According to our results the absolute values of the revealed inhomogeneties in the employed monthly air temperature time series were not so large (less than 1^0C). In spite of this, the degree of inhomogeneity before implementing MASH procedure was rather high. Using MASH allowed decreasing values of TS considerably.

The homogenized time series can be used for further climate studies. Besides, homogenized time series can serve as reference time series for homogenization and quality controlling of another Ukrainian stations data.

Our results also confirm the importance of data homogenization in climate change studies. Inhomogeneity can cause to not only quantitative difference in estimated climate change parameters but also in qualitative conclusions (like in case of the station Lviv).

Acknowledgements

Authors would like to thank Dr. Tamas Szentimrey for his valuable comments, suggestions and help regarding all aspects of MASH homogenization software.

This work has been partly done within the framework of CARPATCLIM project.

References

Climate of Ukraine (Edited by Lipinsky V.M., Dyachuk V.A. and Babichenko V.M.), 2003, Kyiv, 343p. (in Ukrainian)

Climatological handbook USSR, issue 10, Ukrainian SSR "History and physiographic description of meteorological stations, Kiev, 1968, 456 p. (in Russian)

Climatological handbook "History and physiographic description of Ukrainian meteorological stations, Kiev, 2011, 462 p. (in Ukrainian)

Szentimrey,T., 1999: Multiple Analysis of Series for Homogenization (MASH), Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey,T., 2011: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p. 66.

Venema,V.K.C., Mestre,O., Aguilar,E., Auer,I., Guijarro,J.A., Domonkos,P., Vertacnik,G., Szentimrey,T., Stepanek,P., Zahradnicek,P., Viarre,J., Müller-Westermeier,G., Lakatos,M., Williams,C.N., Menne,M.J., Lindau,R., Rasol,D., Rustemeier,E., Kolokythas,K., Marinova,T., Andresen,L., Acquaotta,F., Fratianni,S., Cheval,S., Klancar,M., Brunetti,M., Gruber,C., Prohom Duran,M., Likso,T., Esteban,P., and Brandsma,T. 2012: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89-115.

Vyshnevsky,V.I., Tokar N.F., 1998: The meteorological network is a base of branch functionality, Proceeding of UHMI, 246, P. 5-20. (in Ukrainian)

PROGRAMME

BUDAPEST, HUNGARY

24 – 27 OCTOBER 2011

Venue:

The Headquarters of the Hungarian Meteorological Service (1 Kitaibel Pál street, Budapest)

Monday, 24 October

8:30-9:00 Registration

9:00-12:00
Openning addresses by
the President of HMS
the Organizers
Mestre, O. (FR): Main results of COST HOME
Szentimrey, T. (HU): Theoretical aspects of homogenization
Venema, V., Mestre, O. and COST-HOME group: A blind test of monthly homogenisation algorithms

Lunch break

13:30-17:00
Benchmark:

Willett, K. and the ISTI Benchmarking and Assessment Working Group: Creating a global benchmark cycle for the International Surface Temperature Initiative
Williams, C. N., Menne, M., Thorne, P. (US): Structural Uncertainty in the USHCN temperature records
Guijarro, J. A. (ES): Influence of network density on homogenization performance
Mestre, O. (FR): A focus on correction of the early part of the networks in the benchmark experiment
Domonkos, P., Venema, V. (ES, DE): Efficiencies of homogenisation methods: our present knowledge and its limitation

18:00 Welcome party

Tuesday, 25 October

9:00-12:00
Methods and software for monthly data:

Grimvall, A., Sirisack, S., and Sysoev, O. (SE): Segmentation of Multiple Time Series Using a Tree Algorithm and Dynamic Programming
Robin, S. (FR): Some statistical tools for change-points detection
Lindau, R., Venema, V. (DE): What is the correct number of breaks points hidden in a climate record?
Domonkos, P. (ES): ACMANT: Why is it efficient?
Mestre, O., Domonkos, P., Picard, F., Robin, S., Lebarbier, E., Prohom, M., Aguilar, E. (FR,ES): Combining pairwise/joint/ACMANT detections in HOMER software

Lunch break

13:30-17:00
Methods for daily data:

Stepanek, P. (CZ): Comparison of correction methods for inhomogeneities in daily time series
Szentimrey, T. (HU): Theoretical questions of daily data homogenization
Auer, I., Chimani, B. (AT): AT-HOM – a homogenized daily temperature and precipitation data set for climate impact studies in Austria
Stepanek, P., Zahradnicek, P., Mozny, M. (CZ): Experiences with quality control and homogenization of daily data of various meteorological elements in the Czech Republic
Petrović, P. (RS): Application Of The ReDistribution Method On Custom Data Subsets

Wednesday, 26 October

9:00-12:00
Homogenisation and quality control of climate data series:

Rustemeier, E., Kapala, K., Venema, V., Simmer, C. (DE): Detection and correction of breakpoints in longterm German precipitation series
Dubuisson, B., Gibelin, A. L., Jourdain, S., Colombon, N. (FR): Homogenization of monthly Temperature and Precipitation over France
Andresen, L. (NO): Homogenization of monthly long-term temperature series of mainland Norway
Vízi, Zs., Saunders, M. (UK): Quality control and detection of inhomogeneities in wind observation data in the Met office Integrated Data Archive System (MIDAS)
Bertrand, C. (BE): Data quality control within the RMI climatological network.

Lunch break

13:30-18:00
Management Committee of COST HOME

19:00 Seminar banquet

Thursday, 27 October

9:00-12:30
Homogenisation and quality control of climate data series:

Nordli, Ø. (NO): The homogenisation of the Oslo temperature series 1837-2011
Freitas, L., Pereira, M. G., Amorim, L., Caramelo, L., Mendes, M., Nunes, L. F. (PT): Portuguese temperature dataset homogeneity with HOME R
Guijarro, J. A. (ES): Experiences of homogenization of different climatological variables with the Climatol package
Fratianni, S. (IT): Considerations on snow data
Mikulová, K., Breja, S., Šťastný, P., Bochníček, O., Kajaba, P. (SK): Risks dealing with data preparation of sunshine duration for homogenization process
Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S. (HU): Data homogenization in CarpatClim project

Lunch break

14:00-17:00
Possibility for presentations of various software

LIST OF PARTICIPANTS

AUSTRIA

BARBARA CHIMANI
Zentralanstalt für Meteorologie und Geodynamik
barbara.chimani@zamg.ac.at

INGEBORG AUER
Zentralanstalt für Meteorologie und Geodynamik
ingeborg.auer@zamg.ac.at

BELGIUM

CEDRIC BERTRAN
Royal Meteorological Institute of Belgium
cedric.bertrand@meteo.be

BULGARIA

TANIA MARINOVA
National Institute of Meteorology and Hydrology
Tania.Marinova@meteo.bg

CROATIA

JANJA MILKOVIC
Meteorological and Hydrological Service of Croatia
milkovic@cirus.dhz.hr

CZECH REPUBLIC

MICHAL ZAK
Czech Hydrometeorological Institute
michal.zak@chmi.cz

PAVEL ZAHRADNICEK
Czech Hydrometeorological Institute
zahradnicek@chmi.cz

PETR ŠTĚPÁNEK
Czech Hydrometeorological Institute
petr.stepanek@chmi.cz

VIT KVETON
Czech Hydrometeorological Institute
vit.kveton@chmi.cz

FRANCE

BRIGITTE DUBUISSON
Meteo-France
brigitte.dubuisson@meteo.fr

STÉPHANE ROBIN
INRA
robin@agroparistech.fr

OLIVIER MESTRE
Météo-France
olivier.mestre@meteo.fr

GERMANY

ELKE RUSTEMEIER
Meteorological Institute, University of Bonn
elke.rustemeier@uni-bonn.de

RALF LINDAU
Meteorological Institute, University of Bonn
rlindau@uni-bonn.de

VICTOR VENEMA
University of Bonn
Victor.Venema@uni-bonn.de

JOERG RAPP
DEUTSCHER WETTERDIENST
Joerg.Rapp@dwd.de

HUNGARY

TAMÁS SZENTIMREY
Hungarian Meteorological Service
szentimrey.t@met.hu

ZITA BIHARI
Hungarian Meteorological Service
lakatos.m@met.hu

MÓNIKA LAKATOS
Hungarian Meteorological Service
bihari.z@met.hu

IRELAND

MARY CURLEY
Met Eireann
mary.curley@met.ie

ITALY

MICHELE BRUNETTI
ISAC-CNR
m.brunetti@isac.cnr.it

SIMONA FRATIANNI
UNIVERSITY OF TURIN
simona.fratianni@unito.it

TIBERIU-EUGEN ANTOFIE
JRC-IES
tiberiu.antofie@jrc.ec.europa.eu

NIGERIA

ANDREW ONIARAH
Nigerian Meteorological Agency
a_oniarah@yahoo.com

FRANCIS OLAJIDE ADEFUYE
Nigerian Meteorological Agency,
Climate Data Processing Centre
adefuye_jide@yahoo.com

NORWAY

LARS ANDRESEN
Norwegian Meteorological Institute
lars.andresen@met.no

NORDLI ØYVIND
Norwegian Meteorological Institute
oyvind.nordli@met.no

POLAND

AGNIESZKA WYPYCH
Jagiellonian University, Krakow
agnieszka.wypych@uj.edu.pl

DANUTA CZEKIERDA
Institute Of Meteorology And Water
Management – National Research Institute
danuta.czekierda@imgw.pl

PORTUGAL

MARIO PEREIRA
Centre for Research and Technology of
Agro-Environment
and Biological Sciences (CITAB),
Universidade de Lisboa
gpereira@utad.pt

ROMANIA

SORIN CHEVAL
Ministry for Environment National
Research and Development Institute for
Environmental Protection
sorin.cheval@incdpm.ro

SERBIA

PETROVIĆ PREDRAG
Republic Hydrometeorological Service of
Serbia
predrag.petrovic@hidmet.gov.rs

SLOVAKIA

PETER KAJABA
Slovak Hydrometeorological Institute
peter.kajaba@shmu.sk

KATARÍNA MIKULOVÁ
Slovak Hydrometeorological Institute
katarina.mikulova@shmu.sk

OLIVER BOCHNÍČEK
Slovak Hydrometeorological Institute
oliver.bochnicek@shmu.sk

SPAIN

JOSÉ ANTONIO GUIJARRO
State Meteorological Agency
jguijarrop@aemet.es

PÉTER DOMONKOS
Centre for Climate Change, Univ. Rovira i
Virgili
peter.domonkos@urv.cat

SWEDEN

ERIK ENGSTRÖM
Swedish Meteorological & Hydrological Institute
Erik.Engstrom@smhi.se

THOMAS CARLUND
Swedish Meteorological & Hydrological Institute
Thomas.Carlund@smhi.se

ANDERS GRIMWAL
Science, Linkoping University
anders.grimvall@liu.se

UK

KATHARINE WILLETT
Met Office Hadley Centre
kate.willett@metoffice.gov.uk

ZSUZSANNA VÍZI
University College London
sv@mssl.ucl.ac.uk

USA

CLAUDE.N.WILLIAMS
NOAA
Claude.N.Williams@noaa.gov

ROBERT LUND
Clemson University
lund@clemson.e