Climate Data and Monitoring
WCDMP-No. 85

NINTH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES

AND

FOURTH CONFERENCE ON SPATIAL INTERPOLATION TECHNIQUES IN CLIMATOLOGY AND METEOROLOGY

(Budapest, Hungary, 03 – 07 April 2017)

WEATHER CLIMATE WATER

WORLD
METEOROLOGICAL
ORGANIZATION

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 81 17
Email: Publications@wmo.int

NINTH SEMINAR

FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES

AND

FOURTH CONFERENCE

ON SPATIAL INTERPOLATION TECHNIQUES IN CLIMATOLOGY AND METEOROLOGY

Budapest, Hungary, 03 – 07 April 2017

Organized by the Hungarian Meteorological Service (OMSZ)

Supported by Climate monitoring products for Europe based on Surface in-situ Observations high-resolution gridded information products is generated in Copernicus Climate Change Service (C3S) Climate Data Store (CDS) project WMO and OMSZ

Edited by Tamás Szentimrey, Mónika Lakatos, Lilla Hoffmann

PREFACE

Homogenization of climate data series and spatial interpolation of climate data play a growing role in the meteorology and climatology. The data series are usually affected by inhomogeneities due to changes in the measurement conditions (relocations, instrumentation) therefore a direct analysis of the raw data series can lead to wrong conclusions about climate change. Reconstruction of meteorological fields and gridded databases require spatial interpolation methods.

The first seven Seminars for Homogenization and Quality Control in Climatological Databases as well as the first two Conferences on Spatial Interpolation Techniques in Climatology and Meteorology were held in Budapest hosted by the Hungarian Meteorological Service and supported by WMO. The specialty of both series was focusing on the mathematical methodology!

The 7th Seminar in 2011 was organized together with the final meeting of the COST Action ES0601: Advances, in Homogenization Methods of Climate Series: an integrated approach (HOME), while the 1st Conference on Spatial Interpolation was organized in 2004 in the frame of the COST Action 719: The Use of Geographic Information Systems in Climatology and Meteorology.

In 2014 the 8th Homogenization Seminar and the 3rd Conference on Spatial Interpolation were organized together considering certain theoretical and practical respects. Theoretically there is a strong connection between these topics since the homogenization and quality control procedures need spatial statistics and interpolation techniques for spatial comparison of data. On the other hand the spatial interpolation procedures (e.g. gridding) need homogeneous data series with high quality. Practically the CARPATCLIM project that was launched in 2010 and ended in 2013 is a good example for this problem. The main purpose of the project was to produce a gridded database for the Carpathian region based on homogenized data series. The experiences of this project may be useful for the implementation of gridded databases.

The WMO CCl set up team to support quality control and homogenization activities at NMHSs. The main task of the Task Team on Homogenisation (TT HOM) to provide guidance to Members on methodologies, standards and software required for quality control and homogenization of long term climate time-series. The results of the homogenization sessions can improve the content of the guidance is under preparation.

Gridded climate data derived from meteorological measurements are widely used in climate research, validation of global and regional climate models, in many applications in climate change impacts assessments and derivation of different climate products. The main aim of the C3S _311a Lot4: Copernicus Climate Change Service based on Surface in-situ Observations project is serving the different applications with gridded climate data. The meeting made a nice opportunity to exchange experiences on the recently used methods for homogenization and gridding. Special thanks to the C3S _311a Lot4 project for the support.

The Organizers

CONTENTS

INTRODUCTION ON HOMOGENIZATION, QUALITY CONTROL, SPATIAL INTERPOLATION, GRIDDING

Tamás Szentimrey

Hungarian Meteorological Service
szentimrey.t@met.hu

Abstract

In this paper we try to summarize the main topics of homogenization, quality control and spatial interpolation.

1 INTRODUCTION OF INTRODUCTION

1.1 Background of homogenization seminars and interpolation conferences

The first eight Seminars for Homogenization and Quality Control as well as the first three Conferences on Spatial Interpolation were held in Budapest and hosted by HMS and supported by WMO.

The specialty of both series was the **Mathematical Methodology**!

In 2014 the 8^{th} Homogenization Seminar and the 3^{rd} Interpolation Conference were organized together considering certain theoretical and practical aspects.

Theoretically there is a strong connection between these topics since the homogenization and quality control procedures need spatial statistics and interpolation techniques for spatial comparison of data. On the other hand the spatial interpolation procedures (e.g. gridding) require homogeneous, high quality data series to obtain good results.

1.2 The main topics of homogenization and quality control

- Theoretical, mathematical questions. There is not any exact mathematical theory of the homogenization.
- Relation of monthly and daily homogenization, mathematical formulation of homogenization for climate data series generally.
- Methods for homogenization and quality control (QC) of monthly data series, missing data completion.
- Spatial comparison of series, inhomogeneity detection, correction of series.
- Methods for homogenization and quality control (QC) of daily data series, missing data completion, examination of parallel measurements.
- Usage of metadata.
- Manual versus automatic methods.
- Theoretical evaluation and benchmark for methods, validation statistics.
- Applications of different homogenization and quality control methods, experiences with different meteorological variables.

1.3 The main topics of spatial interpolation

- Theoretical, mathematical questions.
- Interpolation formulas and loss functions depending on the spatial probability distribution of meteorological variables.
- Estimation and modelling of statistical parameters (e.g.: spatial trend, covariance or variogram) for interpolation formulas using spatiotemporal sample and auxiliary model variables (topography).
- Characterization, modelling of interpolation error.
- Real time data quality control (QC) procedures based on spatial comparison, interpolation.
- Use of auxiliary co-variables, background information (e.g.: forecast, satellite, radar data) for spatial interpolation, relation with data assimilation, reanalysis.
- Applications of different interpolation methods for the meteorological and climatological data, experiences with different meteorological variables.
- Gridding of data series, gridded databases.

1.4 Connection between various basic meteorological topics

In our conception the meteorological questions and topics cannot be treated separately. Therefore we present a block diagram (*Figure 1*) to illustrate the possible connection between various important meteorological topics.

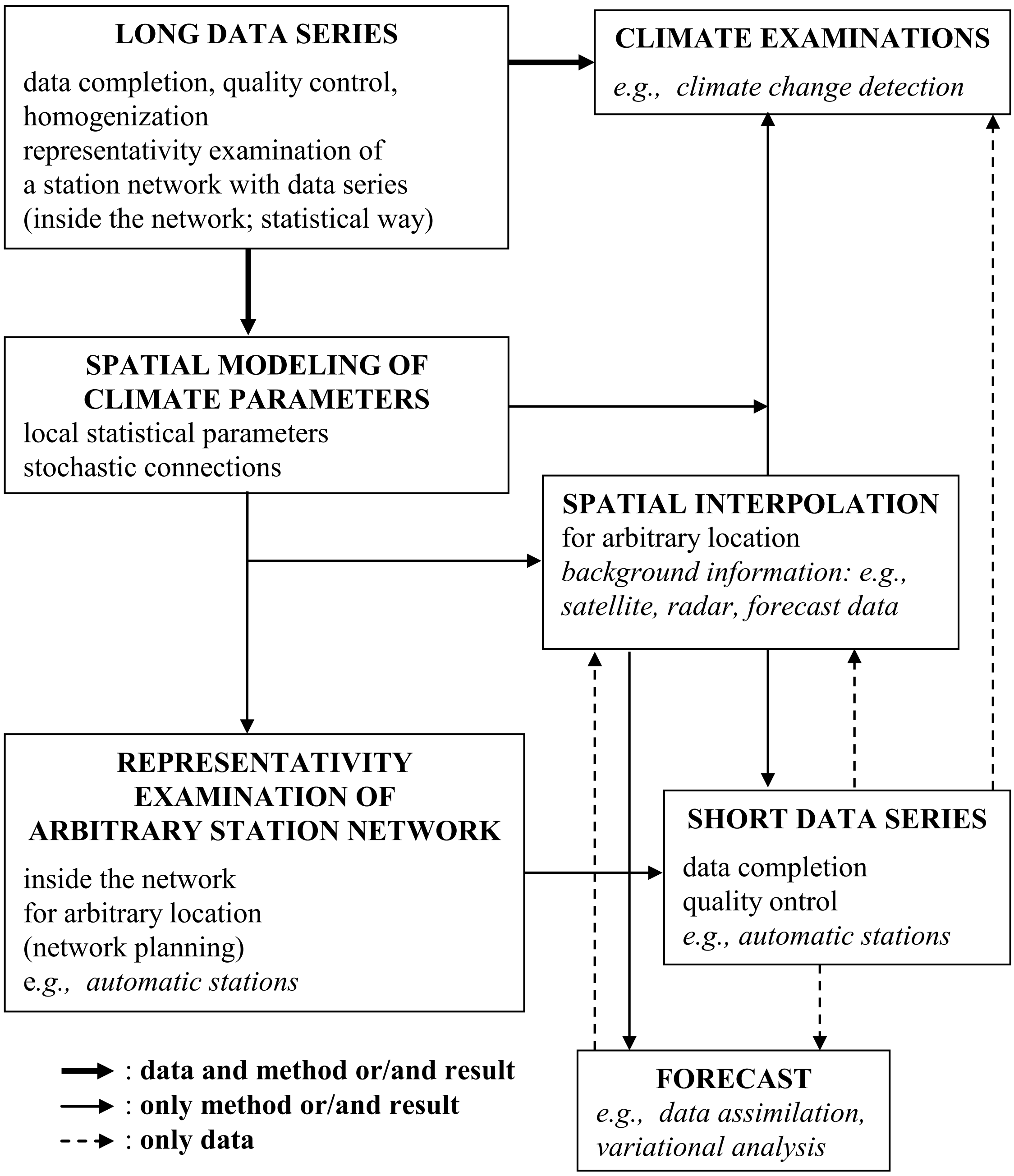


Fig. 1. Block diagram for the possible connection between various basic meteorological topics and systems

1.5 What is the mathematics of homogenization and spatial interpolation in meteorology?

First let us consider the abstract schema of the meteorological examinations. The initial stage is the meteorology that means the qualitative formulation of the given problem. The next stage is the mathematics in order to formulate the problem quantitatively. The third stage is to develop software on the basis of the mathematics. Finally the last stage is again the meteorology that is the application of the developed software and evaluation of the obtained results. In the practice however the mathematics is sometimes neglected.

Concerning our topic we have the following question. What is the mathematics of homogenization in meteorology? There are several methods and software for the homogenization of climate data series but unfortunately there does not exist any exact well elaborated mathematical theory of this problem. At the climatological examinations mainly the physical experiences are dominated while the mathematical formulation of the problems is neglected in general. We do not argue the importance of the physical aspects but the applied not too advanced mathematics is in contrast with the fact that the methods are declared to be based on the mathematical statistics. Moreover often there are some mathematical statements at the description of the methods in the papers – e. g. capability to correct the higher order moments – but without any proof and this way is contrary to the mathematical conventions of course. As we see the basic problem of the homogenization is the unreasonable dominance of the practical procedures over the theory and it is the main obstacle of the progress. As a consequence of this practice the exact evaluation of the methods is also very problematic or properly speaking it is unrealistic and the progress of the homogenization research activity is doubtful. Therefore we try to provide a general approach for the mathematical formulation of homogenization in accordance with the mathematical conventions. We believe the correct mathematical principles can promote understanding and clarifying the questions of homogenization in climatology.

Concerning our topic we have the following question. What kind of mathematics of spatial interpolation is adequate for meteorology? Nowadays the geostatistical interpolation methods built in GIS software are applied in meteorology. The mathematical basis of these methods is the geostatistics that is an exact but special part of the mathematical statistics. The speciality is connected with the assumption that the data are purely spatial. Consequently, as we see it, the geostatistical methods cannot efficiently use the meteorological data series while the data series make possible to obtain the necessary climate information for the interpolation in meteorology.

Modeling of the climate statistical parameters is a key issue to the interpolation of meteorological elements and that modeling can be based on the long data series. However the data series are usually affected by inhomogeneities (artificial shifts), due to changes in the measurement conditions (relocations, instrumentation) therefore the direct analysis of the raw data series can lead to wrong conclusions. In order to deal with this crucial problem many statistical homogenization procedures have been developed for detection and correction of these inhomogeneities.

2. MATHEMATICAL FORMULATION OF CLIMATE DATA HOMOGENIZATION

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

2.1 General mathematical formulation

Notation

Let us assume we have daily or monthly climate 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.

The appropriate theoretical cumulative distribution (CDF) functions are:

$$F_{1,t}(y) = \mathrm{P}(Y_1(t) < y)\ ,\quad F_{2,t}(y) = \mathrm{P}(Y_2(t) < y)\quad y \in (-\infty, \infty)\,,\ t = 1,2,..,n$$

It is very important to remark that as a consequence of some natural changes - e.g. annual cycle, climate change - the series of distribution functions $F_{1,t}(y)$, $F_{2,t}(y)$ $(t = 1,2,..,n)$ may change in time! In the statistical climatology the climate change is equivalent with the changing probability of the meteorological events. The inhomogeneity of data series can be defined on the basis of the distribution functions.

Definition 1

The merged series $Y_2(t)$ $(t = 1,2,..,T)$, $Y_1(t)$ $(t = T+1,..,n)$ is inhomogeneous, if the identity of the distribution functions $F_{2,t}(y) \equiv F_{1,t}(y)$ $(t = 1,2,..,T)$ is not true.

Definition 2

The aim of the 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 $Y_1(t)$ $(t = 1,2,..,T)$ have, i.e.:

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

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

Remark 1

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.

Theorem 1

Let us assume about the random variables Y_1, Y_2 and their distribution functions $F_1(y), F_2(y)$, that $\mathrm{P}(Y_j \in (a_j, b_j)) = 1$ and $F_j(y)$ is a strictly increasing continuous function on the interval (a_j, b_j) $(j=1,2)$. Then applying the transfer function $Y_{1,2h} = F_1^{-1}(F_2(Y_2))$ we obtain that the variable $Y_{1,2h}$ has the same distribution like Y_1 i.e. $\mathrm{P}(Y_{1,2h} < y) = \mathrm{P}(Y_1 < y) = F_1(y)$.

Definition 3

Transfer function: $F_{1,t}^{-1}(F_{2,t}(y))$ and quantile function: $F_{1,t}^{-1}(p)$.

Theoretical formulation of homogenization of $Y_2(t)$ $(t=1,2,..,T)$:

$Y_{1,2h}(t) = F_{1,t}^{-1}(F_{2,t}(Y_2(t)))$, then $\mathrm{P}(Y_{1,2h}(t) < y) = F_{1,t}(y)$.

2.2 Arising mathematical questions to be solved

Let us suppose the merged series is given that is,

$Y_2(t)$ $(t=1,2,..,T)$, $Y_1(t)$ $(t=T+1,..,n)$

In addition we suppose that the assumptions of the former theorem are fulfilled, consequently the theoretical correction or transfer formulas for the series elements are,

$$Y_{1,2h}(t) = F_{1,t}^{-1}(F_{2,t}(Y_2(t))) \quad (t=1,2,..,T) \tag{2}$$

However these transfer formulas are theoretical ones and if we want to apply them in the practice then a number of mathematical statistical estimation problems are arising. The most important problems are as follows.

- Estimation, detection of the change point(s) *T*.
- Estimation of the theoretical distribution functions $F_{1,t}(y)$, $F_{2,t}(y)$ $(t=1,2,..,T)$:
  - i, $F_{1,t}(y)$, $F_{2,t}(y)$ may change in time because of the climate change and the annual cycle, consequently the methodology of the use of the empirical distribution functions is very doubtful.
  - ii, There is no sample for $F_{1,t}(y)$ $(t=1,2,..,T)$ and $F_{2,t}(y)$ $(t=T+1,..,n)$ usually.

These mathematical problems are insolvable generally! Therefore only relative methods can be used with some model assumptions. In addition some simplifications are also necessary. Statistically speaking, some assumptions have to be made!

3. RELATION OF DAILY AND MONTHLY HOMOGENIZATION

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.

3.1 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?

4. OVERVIEW ON HOMOGENIZATION IN MEAN OF MONTHLY SERIES

This section considers some various theoretical aspects of monthly series homogenization.
In the practice the monthly series are homogenized in the mean mostly. The aim of these homogenization procedures is to detect the inhomogeneities of mean and to correct the series.

In connection with the such type of homogenization methods we have to give solutions for the following mathematical problems: relative models, statistical spatiotemporal modelling of the series, methodology for comparison of series, break point (changepoint) and outlier detection, methodology for correction of series, quality control procedures, missing data completion, usage of metadata, relation of daily and monthly homogenization, manual versus automatic methods, evaluation of methods (theoretical, benchmark).

In practice there are absolute and relative methods applied for homogenization. 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.

4.1 General structure of additive spatiotemporal models

If the data series are normally distributed (e.g. temperature) then the additive model can be used. 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,

$$X_j(t) = \mu(t) + E_j + IH_j(t) + \varepsilon_j(t) \qquad \left(j = 1,2,\ldots,N\,;\; t = 1,2,\ldots,n \right), \tag{3}$$

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. 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 normal distributed vector variables $\boldsymbol{\varepsilon}(t)=[\varepsilon_1(t),..,\varepsilon_N(t)]^{\mathrm{T}}\in N(\mathbf{0},\mathbf{C})$ $(t=1,...,n)$ are totally independent in time. The spatial covariance matrix $\mathbf{C}$ describes the spatial structure of the series.

If the data series are quasi lognormal distributed (e.g. precipitation) then the multiplicative model can be used that can be transformed into the additive one by certain logarithmic procedure.

4.2 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,

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

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 decrease the variance of noise term $\varepsilon_{Z_j}(t)$. The optimal weighting factors λ_{ji} that minimize the variance are determined by the spatial covariance matrices $\mathbf{C}$ uniquely.

4.3 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

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

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.

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

4.3.2 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, 2014) 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 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,

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

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 $(\hat{T}_{k-1}, \hat{T}_k]$.

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, 2014).

4.4 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, HOMER, ACMANT methods (*Caussinus* and *Mestre*, 2004) uses the standard least squares technique for joint estimation of the correction factors. Probably the generalized least squares estimation technique based on spatial covariance structure would be more efficient.

The MASH procedure (*Szentimrey*, 1999, 2014) 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!

4.5 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 numerous stations series must be examined together in general! Flexible automatic systems are necessary wherein the mechanic, labour-intensive procedures must be automated. But not pushing button systems! The problem is much more complex.

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,
- automatic usage of metadata.

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

4.6 Possibilities for evaluation of the methods

4.6.1 Theoretical evaluation

If want to obtain a real image of the methods, then the theoretical evaluation of their mathematical basis is indispensable.

4.6.2 Benchmark

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).

The creation of adequate benchmark dataset and the development of appropriate validation statistics are critical points and they need also strong theoretical mathematical background.

We remark that 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?

5. MATHEMATICAL OVERVIEW OF SPATIAL INTERPOLATION PROBLEM IN METEOROLOGY

According to the interpolation problem the unknown predictand $Z(\mathbf{s}_0,t)$ is estimated by use of the known predictors $Z(\mathbf{s}_i,t)$ $(i=1,...,M)$ where the location vectors $\mathbf{s}$ are the elements of the given space domain D and t is the time.

5.1 Additive model of spatial interpolation

The type of the adequate interpolation formula depends on the probability distribution of the meteorological variable. Assuming normal distribution (e.g. temperature) the additive (linear) formula is adequate.

5.1.1 Statistical parameters

In general the interpolation formulas have some unknown interpolation parameters which are known functions of certain statistical parameters. At the additive interpolation formulas the basic statistical parameters can be divided into two groups such as the local and the stochastic parameters. The local parameters are the expected values $\mathrm{E}(Z(\mathbf{s}_i,t))$ $(i=0,...,M)$. The stochastic parameters are the covariance or the variogram values belonging to the predictand and predictors such as,

$\mathbf{c}$: predictand-predictors covariance vector,

$\mathbf{C}$: predictors-predictors covariance matrix,

$\boldsymbol{\gamma}$: predictand-predictors variogram vector,

$\boldsymbol{\Gamma}$: predictors-predictors variogram matrix.

The covariance is preferred in mathematical statistics and meteorology while the variogram is preferred in geostatistics.

5.1.2 Linear meteorological model for expected values

At the statistical modeling of the meteorological elements we have to assume that the expected values of the variables are changing in space and in time alike. The spatial change means that the climate is different in the regions. The temporal change is the result of the possible global climate change. Consequently in case of linear modeling of expected values we assume that

$$\mathrm{E}(Z(\mathbf{s}_i,t)) = \mu(t) + E(\mathbf{s}_i) \quad (i = 0,..,M) \tag{6}$$

where $\mu(t)$ is the temporal trend or the climate change signal and $E(\mathbf{s})$ is the spatial trend.

5.1.3 Additive (Linear) Interpolation Formula

Assuming the linear model (6) the appropriate additive meteorological interpolation formula is as follows,

$$\hat{Z}(\mathbf{s}_0,t) = \lambda_0 + \sum_{i=1}^{M} \lambda_i \cdot Z(\mathbf{s}_i,t)$$

where $\sum_{i=1}^{M} \lambda_i = 1$ because of unknown $\mu(t)$.

The optimal interpolation parameters λ_0, λ_i $(i = 1,...,M)$ minimize the root-mean-square error, $RMSE(\mathbf{s}_0) = \sqrt{\mathrm{E}\left(\left(Z(\mathbf{s}_0,t) - \hat{Z}(\mathbf{s}_0,t)\right)^2\right)}$. (7)

These optimal parameters are known functions of statistical parameters!

The optimal constant term is: $\lambda_0 = \sum_{i=1}^{M} \lambda_i (E(\mathbf{s}_0) - E(\mathbf{s}_i))$

The vector of optimal weighting factors $\boldsymbol{\lambda}^{\mathrm{T}} = [\lambda_1,..,\lambda_M]$ can be written in covariance form

$$\boldsymbol{\lambda}^{\mathrm{T}} = \left(\mathbf{c}^{\mathrm{T}} + \mathbf{1}^{\mathrm{T}} \frac{\left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}\right)\mathbf{C}^{-1},$$

or equivalently in variogram form

$$\boldsymbol{\lambda}^{\mathrm{T}} = \left(\boldsymbol{\gamma}^{\mathrm{T}} + \mathbf{1}^{\mathrm{T}} \frac{\left(1 - \mathbf{1}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}\right)}{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\mathbf{1}}\right)\boldsymbol{\Gamma}^{-1}.$$

Consequently the unknown statistical parameters are the spatial trend differences $E(\mathbf{s}_0) - E(\mathbf{s}_i)$ $(i = 1,...,M)$ and the covariances $\mathbf{c}$, $\mathbf{C}$. In essence these parameters are climate parameters which in fact means we could interpolate optimally if we knew the climate.

Remark

Unfortunately inadequate formulas are often applied in the practice:

- Inverse Distance Weighting (IDW):
  $\lambda_0 = 0$ that is excluding spatial trend, and λ_i $(i = 1,...,M)$ based on distances are not optimal weighting factors.
- Ordinary kriging: $\lambda_0 = 0$ excludes the spatial trend.

Adequate formulas are in meteorology:

- Universal kriging formula,
- Regression (residual, detrended) kriging formula.

But in geostatistics modeling of statistical parameters is based on only the actual predictors.

5.1.4 Possibility for modeling of unknown statistical parameters in Meteorology

The special possibility in meteorology is to use the long meteorological data series for modeling of the climate statistical parameters in question. The data series make possible to know the climate in accordance with the fundament of statistical climatology!

The main difference between geostatistics and meteorology can be found in the amount of information being usable for modeling the statistical parameters. In geostatistics the usable information or the sample for modeling is only the actual predictors $Z(\mathbf{s}_i, t)$ $(i = 1,...,M)$ which belong to a fixed instant of time, that is a single realization in time. While in meteorology we have spatiotemporal data, namely the long data series which form a sample in time and space as well and make possible to model the climate statistical parameters in question. If the meteorological stations $\mathbf{S}_k$ $(k = 1,..,K)$ $(\mathbf{S} \in D)$ have long data series then spatial trend differences $E(\mathbf{S}_k) - E(\mathbf{S}_l)$ $(k,l = 1,...,K)$ as well as the covariances $\text{cov}(Z(\mathbf{S}_k), Z(\mathbf{S}_l))$ $(k,l = 1,...,K)$ can be estimated statistically. Consequently these parameters are essentially known and provide much more information for modeling than the predictors $Z(\mathbf{s}_i, t)$ $(i = 1,...,M)$ only. However nowadays unfortunately the geostatistical interpolation methods built in GIS software are applied in meteorology mostly.

5.1.5 Interpolation error RMSE

The optimal interpolation parameters λ_0, λ_i $(i = 1,...,M)$ minimize the root-mean-square error $RMSE(\mathbf{s}_0)$ according to the equation (7). The uncertainties of interpolation can be characterized quantitatively by this $RMSE(\mathbf{s}_0)$.

In the optimal case the minimal error,

$$RMSE(\mathbf{s}_0) = \sqrt{\left(D^2(\mathbf{s}_0) - \mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right) + \left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)^2 \cdot \frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}}$$

where standard deviation $D(\mathbf{s}_0) = \mathrm{D}(Z(\mathbf{s}_0, t))$ and $\mathbf{c}$ is the predictand-predictors covariance vector while $\mathbf{C}$ is the predictors-predictors covariance matrix. Consequently if we model the spatial covariance structure then we can also model the interpolation error $RMSE(\mathbf{s}_0)$ to characterize quantitatively the uncertainties of interpolation.

5.1.6 Real time Quality Control

Test schema of real time Quality Control (QC) procedure at additive, normal model is:

$$\frac{Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)}{RMSE(\mathbf{s}_0)} \in N(0,1),$$

where $Z(\mathbf{s}_0)$ is the predictand to be controlled, $\hat{Z}(\mathbf{s}_0)$ is the interpolated value and $RMSE(\mathbf{s}_0)$ is the modelled interpolation error.

6. INTERPOLATION WITH BACKGROUND INFORMATION

The background information e.g. forecast, satellite, radar data can be efficiently used to decrease the interpolation error. In this paper only the interpolation based on additive model or normal distribution is presented.

According to the section 5.1.3 let us assume that,

$Z(\mathbf{s}_0,t)$: predictand,

$\hat{Z}(\mathbf{s}_0,t)=\lambda_0+\sum_{i=1}^{M}\lambda_i Z(\mathbf{s}_i,t)$: interpolated predictand,

moreover there is given,

$\mathbf{G}=\{G(\mathbf{s},t) \mid \mathbf{s}\in D\}$: background information on a dense grid.

6.1 The principle of interpolation with background information

The interpolated predictand given $\mathbf{G}$ can be expressed as,

$$\hat{Z}_G(\mathbf{s}_0,t)=\hat{Z}(\mathbf{s}_0,t)+\mathrm{E}\left(Z(\mathbf{s}_0,t)-\hat{Z}(\mathbf{s}_0,t) \mid \mathbf{G}\right)$$

where $\mathrm{E}\left(Z(\mathbf{s}_0,t)-\hat{Z}(\mathbf{s}_0,t) \mid \mathbf{G}\right)$ is the conditional expectation of $Z(\mathbf{s}_0,t)-\hat{Z}(\mathbf{s}_0,t)$, given $\mathbf{G}$.

6.2 Reanalysis data, Data Assimilation

The reanalysis data are based on the data assimilation which procedure is in strong relationship with the methodology of interpolation with background information. The Bayes estimation theory is the mathematical background of the data assimilation and the following variational cost function has to be minimized in order to estimate the analysis field,

$$J(\mathbf{z})=(\mathbf{z}-\mathbf{g})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{z}-\mathbf{g})+(\mathbf{y}_0-\mathbf{F}\mathbf{z})^{\mathrm{T}}\mathbf{P}^{-1}(\mathbf{y}_0-\mathbf{F}\mathbf{z}) \qquad (8)$$

$\mathbf{z}$: analysis field, predictand (grid),
$\mathbf{g}$: given background field (forecast),
$\mathbf{y}_0$: given observations, predictors; $\mathbf{F}\mathbf{z}=\mathrm{E}(\mathbf{y}_0|\mathbf{z})$,
$\mathbf{Q}$: background error covariance matrix,
$\mathbf{P}$: observation error covariance matrix.

It can be proved that this procedure is essentially an interpolation with background information including a quality control part for the predictors.

However there are several problems with the reanalysis data in the practice:

i, Inhomogeneous predictor station data series are used.
ii, Few stations are used with little spatial representativity.
iii, There are also some problems with the data assimilation formula (8):

- Lack of good climate statistical parameters in matrix **Q.**
- Formula (8) includes an implicit assumption of $\mathrm{E}(\mathbf{z} \mid \mathbf{g})=\mathbf{g}$.

The detailed analysis of these problems can be found in the paper (*Szentimrey*, 2016).

7. GRIDDED DATABASES

We emphasize the importance of gridded databases based on observations with good quality!

The main topics connected with these databases are as follow.

- Homogenization of dense station data series.

- Interpolation, gridding of homogenized series.

- The methodology for comparison of gridded datasets?

There is also an important question that is the homogeneity of satellite datasets?

References

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

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, Bihari, Z., Lakatos, M., Szalai,S., 2011: Mathematical, methodological questions concerning the spatial interpolation of climate elements. Proceedings of the Second Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2009, *Időjárás* 115, 1-2, 1-11

Szentimrey, T. 2013: Theoretical questions of daily data homogenization, *Időjárás* Vol. 117. No. 1, January-March 2013. pp. 113-122.

Szentimrey,T.,2014: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p.69.

Szentimrey T., Lakatos M., Bihari Z., 2015: Mathematical questions of homogenization and quality control, Proceedings of the 8th Seminar for Homogenization and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation Techniques in Climatology and Meteorology, Budapest, Hungary, 2014, WCDMP-No. 84, pp. 5-22

Szentimrey T., Bihari Z., Lakatos M., 2015: Mathematical questions of spatial interpolation of climate variables, Proceedings of the 8th Seminar for Homogenization and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation Techniques in Climatology and Meteorology, Budapest, Hungary, 2014, WCDMP-No. 84, pp. 107-114

Szentimrey, T., 2016: Analysis of the data assimilation methods from the mathematical point of view. In: *Mathematical Problems in Meteorological Modelling*, Springer International Publishing, Switzerland, 193–205

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

ANALYSIS OF THE IMPACTS OF THE AUTOMATIZATION OF MEASUREMENT SYSTEMS USING PARALLEL MEASUREMENTS FROM GERMAN CLIMATE REFERENCE STATIONS

Lisa Hannak, Karsten Friedrich, Florian Imbery, Frank Kaspar

Deutscher Wetterdienst, National Climate Monitoring, Frankfurter Str. 135, 63067 Offenbach, Germany, lisa.hannak@dwd.de

Abstract

Meteorological parameters were measured systematically since more than a century. During this time, measurement systems were changed for example from manual observations to automatic measurements. Changing the measurement system can potentially affect the homogeneity of time series. To study the effects of the automatization of measurements, Germany's meteorological service DWD (Deutscher Wetterdienst) operates parallel measurements at climate reference stations since 2008. Observed meteorological parameters are air temperature, extreme temperatures, soil temperature, air pressure, relative humidity, sunshine duration, and precipitation. This presentation shows first results of the statistical analysis of the differences between the measurement systems and their effects on the homogeneity of time series.

1. INTRODUCTION

Parallel measurements can be used to analyze the performance of observations made with different screens to reduce the radiation effect on the measurements (*Brunet et al*., 2011; *Erell et al.,* 2005; *Brandsma and Van der Meulen*, 2008). These parallel measurements allow the direct comparison of different measurement conditions and help to find the best combination of measurement systems. Not only screens can affect the measurements, also the used sensor (including the calibration, the measurement method, and used material) should be compared to obviate errors caused by the sensor. *Doesken* (2005); *Legg* (2014); *Rennie et al.* (2014); *Baciu et al*. (2005); *Acquaotta et al.* (2016) analyzed different sensors for different kind of meteorological parameters. In most cases the parallel measurements are taken at nearby stations or are only available for a short overlapping time range.

In Germany, each set of parallel measurements of different sensors is taken at one measurement site therefore the distance between sensors is only a few meters and the overlapping time range is up to more than eight years. There are two types of German climate reference stations. Type I are stations where manual and automatic instruments are measuring in parallel. At climate reference stations type II, automatic sensors are operated in parallel over a sufficient long time to analyze the impact of changing automatic sensors (see *Kaspar et al.*, 2016). The objectives of the parallel measurements are to have a quality control of the measurements, to analyze the impact of changing the measurement system on the homogeneity of long time series, and to use the results from these climate reference stations for homogenization of time series of other stations in the measuring network. The impacts of changing manual measurement systems by automatic ones for the parameter air temperature are analyzed by *Kaspar et al.* (2016). At the traditional observing times the differences between automatic and manual observations were found to be small with a mean value close to zero and a small standard deviation. For that parameter, they do not expect any break in the time series caused by the transition of the measurements system. In the difference time series of daily maximum temperature, an annual cycle was found which is (at least partly) caused by a radiation effect of the automatic thermometer screen 'LAM 630', which is mostly used in the German meteorological network. Stations which do not use this kind of shelter do not show an annual cycle in the differences. The differences between automatic and manual measurements can be reduced by optimizing the position of the automatic sensors in the shelter.

In this study, we show results of the differences between measurements systems for the parameters soil temperature, relative humidity, air pressure, and daily sunshine duration.

2. DATA AND METHODS

The manual readings were done thrice a day at the traditional observing times (6:30 UTC, 13:30 UTC, 20:30 UTC). To analyze the comparability of automatic and manual measurement systems, automatic measurements at these points in time were used to calculate differences between automatic and manual observations. The differences were filtered, such that differences which are larger than four times the pseudo standard deviation or a tolerance value were excluded of the data. The pseudo standard deviation is less influenced by outlier and is therefore suitable to filter outlier (*Lanzante*, 1996). To calculate daily mean values, two different equations were used (equivalent to the operational routines today and in the past). The first one is based on the traditional observing times. The second one is an arithmetic mean over 24 hourly values which can only be used for the automatic measurements because of the larger temporal resolution. In this comparison the manual observations are treated as reference.

3. RESULTS

3. 1. Soil temperature in 0.05, 0.1, 0.2, 0.5, and 1 meter depth

Manually the soil temperature is measured with mercury-in-glass thermometers which are in a pipe in the ground and have contact to the ground in the bottom of the pipe. To read the value of the thermometers in 0.05, 0.1, and 0.2 m depth, the thermometer reaches outside the ground and is sloped to reduce reading errors. The thermometers in 0.5 and 1 m depth are lifted outside the pipe to enable the reading of the value. The rest of the time, these thermometers are inside the pipe and do not reach outside the surface.

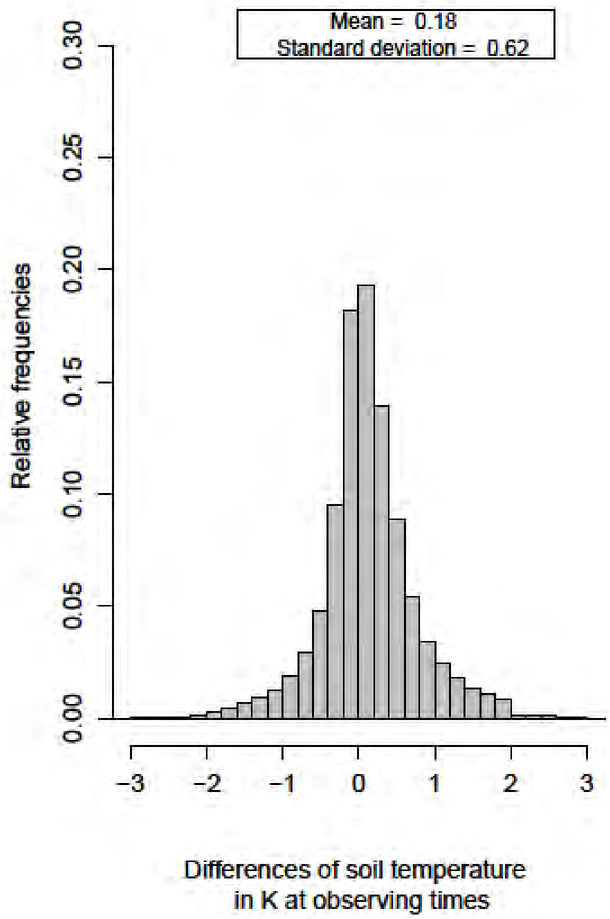


Fig. 1. Histogram of differences of soil temperature in 0.05m depth in K between automatic and manual observations measured at traditional observing times. Differences of all climate reference stations were used.

The soil temperature is measured automatically with Pt100 sensors. These sensors are platinum resistance thermometers and are fixed on a stainless steel framework. Differently to the manual instrument, the automatic sensors are inserted into the ground horizontally.

Figure 1 shows the histogram of the differences between automatic and manual measurements in 0.05 m depth. The distribution is asymmetric and the mean is 0.18 K, which means that the automatic measurements are mostly warmer than manual observations. Also the standard deviation is relatively large which is caused by an annual cycle in the differences. *Figure 2* shows the mean differences of soil temperature in 0.05 m depth split into observing times, seasons, and stations. In most stations the differences are highly positive during spring and summer and smaller in autumn and winter during midday (13:30 UTC). During the evening, there are stations with negative differences during spring and summer (Fichtelberg, Frankfurt, Schleswig, Aachen) but there also exists stations with positive differences (Görlitz, Hohenpeißenberg, Potsdam, Aachen-Orsbach). There is no clear pattern in all of these stations. There might be a radiation effect on the measurements because in most stations the differences are larger during summer and midday where strong solar radiation is present.

In deeper depths, the seasonal cycle is smaller in terms of the size of the amplitude. So this effect is mostly caused by processes which are stronger close to the surface.

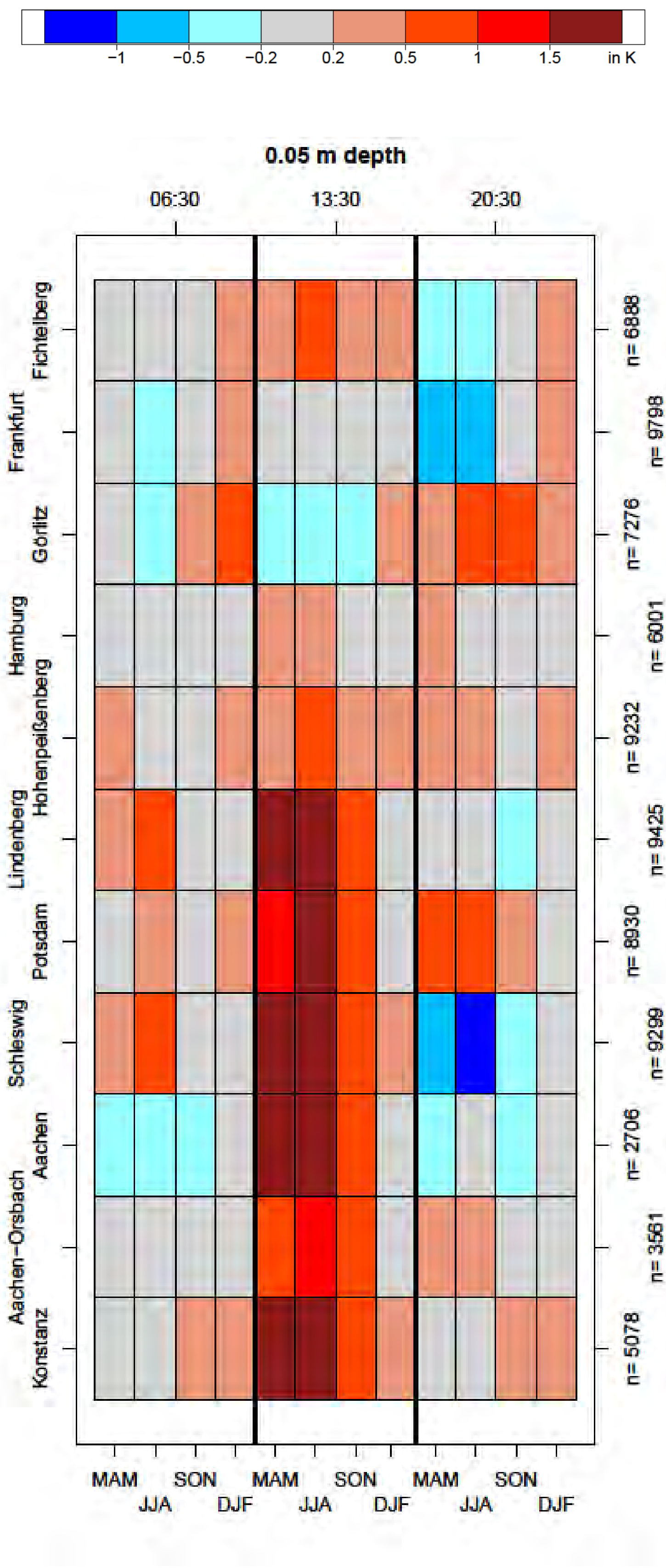


Fig. 2. Mean differences between automatic and manual observations split in seasons (spring=MAM, summer=JJA, autumn=SON, winter=DJF), observing times (6:30 UTC, 13:30 UTC, 20:30 UTC), and stations. The number of measurements of each station is shown on the right.

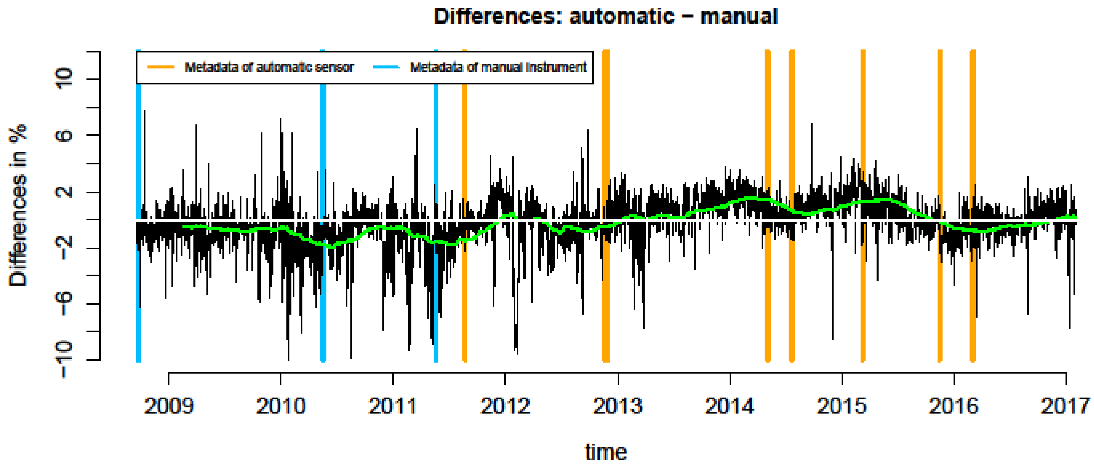


Fig. 3. Difference time series of relative humidity in % (Lindenberg). Metadata information of manual instrument (blue) and automatic sensor (orange) are represented as vertical lines. The green line represents the moving averaged over 150 days.

3.2 Relative humidity

The parameter relative humidity is manually measured with a psychrometer which is mostly positioned in a Stevenson shelter and read at the traditional observing times. The automatic sensor changed at the beginning of the year 2016. Before 2016 the relative humidity is automatically measured with a HMP45D-sensor (the sensor Hygromer MP100 was installed at the station Brocken). Now, the sensor type EE33 is used. So, for this parameter we have

parallel measurements for manual and automatic sensors and for the two automatic sensors. The automatic sensors are mostly installed in the lamellar shelter 'LAM 630'.

The distribution of the differences of relative humidity between automatic and manual observations is symmetric, the mean is close to zero and the standard deviation is smaller than 3 % relative humidity. Based on this analysis, it can be assumed that the measurement systems are comparable.

Figure 3 shows that there are some features which we do not capture by the analysis of the distribution of the differences. In the difference time series of Lindenberg, there are trend periods which are partly up to several months long. The direction of this trend periods is not uniform such that the drifts can be positive or negative.

Another feature in the differences of automatic and manual observations is a radiation effect of the lamellar shelter 'LAM 630' on the automatic observations. Especially the new sensor EE33 is affected by this radiation effect. This sensor has an integrated heating system to prevent condensation on the sensor. Due to condensation, the formerly used sensor HMP45D shows problems to measure relative humidity changes after periods of saturation. With the integrated heating system, the EE33 shows better results and reacts faster on relative humidity changes. To calculate the relative humidity, an additional temperature sensor is necessary. This additional sensor is positioned in the south-east of the lamellar shelter 'LAM 630'. At this position the radiation effect is strongest. An overestimation of temperature measurements leads to an underestimation of relative humidity for the sensor EE33.

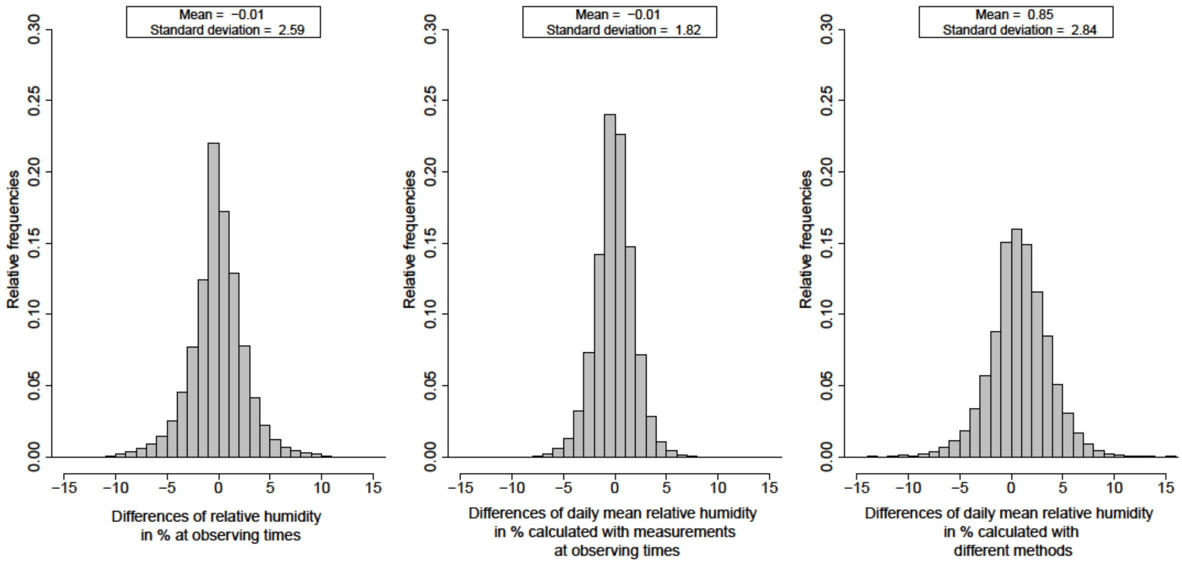


Fig. 4. Histogram of differences of relative humidity in % between automatic and manual observations measured at traditional observing times (left), differences of daily mean values calculated with the traditional equation (center), and differences of daily mean values calculated with different equations (right). Differences of all climate reference staions were used.

The differences between daily mean automatic and manual measurements using the traditional equation for both measurements are small with a mean value close to zero. Using the arithmetic mean over 24 hourly values for the automatic measurements to calculate daily mean values and the traditional equation for manual observations, higher differences are more likely than using the same equation for both measurement systems and the standard deviation of the differences is larger than using the traditional equation for both observations (*Figure 4*). It is advisable to use the traditional equation also for the automatic measurements to prevent a break in the time series.

3.3. Pressure

The meteorological parameter air pressure is measured manually with a mercury barometer at the traditional observing times. The measurement uncertainty is 0.3 hPa (the reading accuracy is 0.1 hPa) (*Löffler*, 2012). Automatic measurements of air pressure are performed with a digital barometer which has a measurement uncertainty of 0.15 hPa (included are the effects of linearity, calibration uncertainty and hysteresis) (*Löffler*, 2012). The differences between the measurement systems at the traditional observing times are small, such that the mean value is close to zero and the standard deviation is 0.23 hPa (*Figure 5*). Consequently the differences of daily mean values calculated with the traditional equation are small as well. The mean value of the differences between daily mean automatic and manual measurements calculated with the two different equations (arithmetic mean over 24 hourly values for automatic measurements and the traditional equation for manual measurements) is close to zero but the standard deviation is over twice as large as using the same equation. To prevent breaks in the time series of daily mean air pressure, the traditional equation should be used for automatic measurements to analyze long time changes.

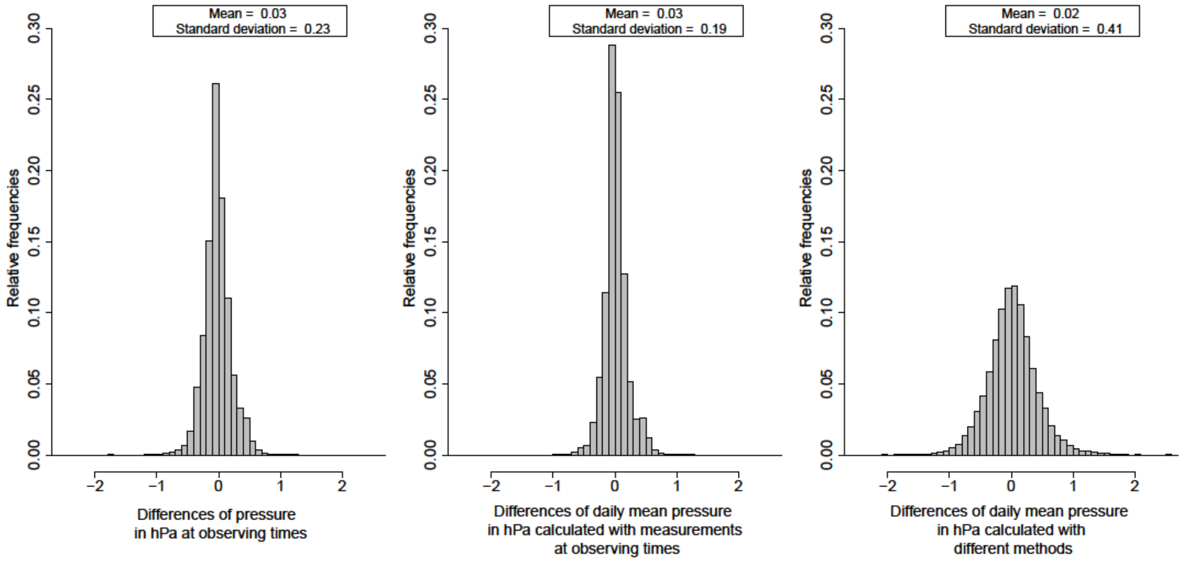


Fig. 5. Histogram of differences of air pressure in hPa between automatic and manual observations measured at traditional observing times (left), differences of daily mean values calculated with the traditional equation (center), and differences of daily mean values calculated with different equations (right). Differences of all climate reference staions were used.

3.4. Daily sunshine duration

To measure the daily sunshine duration manually, a Campbell-Stokes instrument is used. A glass marble focuses the sunlight and burns a line into a paper. The length of this line represents the time range when the sun was shining. The daily sunshine duration is automatically measured by a SONIe. This instruments works with a photo diode and measures the solar radiation. Is the solar radiation higher than a critical value, the result can be interpreted as sunny moment (*Löffler*, 2012).

The differences between automatic and manual measurements show a negative mean value. The manual observations are in most cases higher than the automatic, especially during summer. *Legg* (2014) analyzes that the manual readings are overestimated when there are frequent changes between sunny and cloudy conditions. The burned mark on the paper is not sharp such that the interpretation is subjective. Also the standard deviation is higher in summer than in winter which can be caused by the different day time lengths. In summer the days are longer and the potential error can be larger than in winter times.

4. CONCLUSIONS AND OUTLOOK

Changes in the measurements systems should be reduced to a small number but in the process of automatization, measurement systems have to be changed. Understanding the impacts of changing measurement systems is important for climate data. The differences of soil temperature are characterized by a seasonal cycle with mostly higher values for automatic measurements in summer during midday and smaller differences in winter. The amplitude of such a seasonal cycle is highest close to the surface and lessens in deeper depths.

In averaged, the differences between automatic and manual observations for relative humidity are small. Compared to the manual observations, the automatic measurements show drift periods which can be several months long. Furthermore, the formerly used automatic sensor HMP45D has problems after periods of saturation. This sensor reacts slowly on relative humidity changes after periods of high relative humidity. With the new automatic sensor EE33, condensation on the sensor can be prevented (because of the integrated heating system). The new sensor has a radiation effect caused by the lamellar shelter 'LAM 630'. To improve the measurements, the temperature sensor which is used to correct the heating effect from the heating system of the EE33 should not be placed in the south of the lamellar shelter.

The differences in air pressure measurements are mostly small. Here, we do not have indications for an artificial break in the time series caused by the change of the measurement system. For daily sunshine duration, the difference time series are characterized by a seasonal cycle with mostly negative values during summer which means that the automatic measurements are smaller. This can be explained by the reading accuracy of the manual measurements which results in most cases in an overestimation of sunshine duration.

Comparing the different equation to calculate daily mean values, the standard deviation when using different kind of equations is larger than when using the traditional equation for manual and automatic measurements. It is advisable to use the traditional equation for modern instruments as well to prevent large errors in individual cases.

As a next step, parallel measurements for precipitation are analyzed. Furthermore the parallel measurements are used to monitor the quality of data and to identify breaks which are caused by sensor defects or changes in the measurements conditions. These breaks should be corrected if necessary.

References

Acquaotta, F., Fratianni, S., and Venema, V. (2016). Assessment of parallel precipitation measurements networks in Piedmont, Italy. *International Journal of Climatology*.

Baciu, M., Copaciu, V., Breza, T., Cheval, S., and Pescaru, I. V. (2005). Preliminary results obtained following the intercomparison of the meteorological parameters provided by automatic and classical stations in Romania. In WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005).

Brandsma, T. and Van der Meulen, J. (2008). Thermometer screen intercomparison in De Bilt (the Netherlands) – Part II: Description and modeling of mean temperature differences and extremes. *International Journal of Climatology*, 28(3):389–400.

Brunet, M., Asin, J., Sigró, J., Bañón, M., García, F., Aguilar, E., Palenzuela, J. E., Peterson, T. C., and Jones, P. (2011). The minimization of the screen bias from ancient western mediterranean air temperature records: an exploratory statistical analysis. *International Journal of Climatology*, 31(12):1879–1895.

Doesken, N. J. (2005). The national weather service mmts (maximum-minimum temperature system)–20 years after. In Preprint, 15th Applied Climatology Conference. Amer. Meteor. Soc., Savanna, GA.

Erell, E., Leal, V., and Maldonado, E. (2005). Measurement of air temperature in the presence of a large radiant flux: an assessment of passively ventilated thermometer screens. *Boundary-layer meteorology*, 114(1):205–231.

Kaspar, F., Hannak, L., and Schreiber, K.-J. (2016). Climate reference stations in Germany: Status, parallel measurements and homogeneity of temperature time series. *Advances in Science and Research*, 13:163–171.

Lanzante, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology*, 16(11):1197–1226.

Legg, T. (2014). Comparison of daily sunshine duration recorded by Campbell–Stokes and Kipp and Zonen sensors. *Weather*, 69(10):264–267.

Löffler, H. (2012). Meteorologische Bodenmesstechnik (vormals Instrumentenkunde), volume 3. Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main.

Rennie, J., Lawrimore, J., Gleason, B., Thorne, P., Morice, C., Menne, M., Williams, C., Almeida, W. G., Christy, J., Flannery, M., et al. (2014). The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. *Geoscience Data Journal*, 1(2):75–102.

TIME SERIES HOMOGENISATION WITH OPTIMAL SEGMENTATION AND ANOVA CORRECTION: PAST, PRESENT AND FUTURE

Peter Domonkos

Tortosa, Spain, dpeterfree@gmail.com

Abstract

In the past two decades new, mathematically sophisticated methods have been developed for the joint detection and joint adjustment of multiple inhomogeneities in climatic time series. In 2010-2011, during the blind international tests of COST ES0601 (HOME), the Optimal Segmentation for multiple break detection and the ANOVA Correction model for the joint adjustments of biases justified their power. The interactive method HOMER, one representative of this method family, became the principally recommended method of HOME, while the automatic method ACMANT, other representative of the method family, became also a recommended method after having produced the best test results among the tested methods. However, HOMER is yet known and applied only in Europe, while ACMANT is less known until now. This study looks through the past, present and future of Optimal Segmentation and ANOVA Correction methods, analyses the possible causes of the delay in the spread of their practical use in climate research, and makes recommendations.

1. INTRODUCTION

The homogenisation of climatic time series is a research topic on the border of climatology and statistics. The increasing demand for the more precise estimates of climate change and climate variability and the improvement of computer science in the recent decades have accelerated developments in this research area, and several pieces of the new findings have surprised even the best experts. Some 15-20 years ago it was generally believed that homogenisation with statistical tests generally makes climate data more reliable for climate variability studies, anything is the chosen homogenisation method (*Peterson et al.,* 1998; *Aguilar et al.,* 2003). However, tests with a realistic benchmark dataset of HOME showed that the accuracy of homogenisation is strongly varied according to homogenisation method, dataset properties and the time resolution of homogenisation.

The closing study of HOME (*Venema et al.* 2012) reports five important conclusions of test experiments among others: i) The reliability and accuracy of time series homogenisation highly depends on the selected statistical methods; ii) Methods that do not presume the homogeneity of reference series function perform better than other methods; iii) Absolute homogenisation (i.e. homogenisation without filtering climate signal through time series comparisons) in automatic mode can worsen data quality rather than improving it; iv) Tested methods including tools for the joint detection and joint correction of multiple inhomogeneities, namely PRODIGE, ACMANT and MASH, outperform the other methods; v) Reliability and accuracy of the homogenisation products generally decrease with the increasing time resolution of homogenisation. Note that beyond PRODIGE, ACMANT and MASH, the HOME closing report recommended also the use of Craddock test for its excellent test results, and the PHA-USHCN method, which although showed somewhat more modest results than multiple break homogenisation methods, it seemed to be the best prepared for the homogenisation of huge and spatially dense datasets. Note also that the renewed version of Climatol homogenisation method (*Guijarro*, 2014) often shows competitive efficiencies with multiple break methods. The wide spread of method efficiencies found by HOME and some other recent researches shows that it is crucially important to foster the use of effective homogenisation methods in practical climate research. It is the main aim of this study.

The organisation of the paper is as follows. In the next section, the selection of the methods of Optimal Segmentation and ANOVA Correction is reasoned, and the methods are described. In section 3 some steps of the past development in the inclusion of this methodology in modern homogenisation methods is recalled, and the two main representatives of the method family, i.e. HOMER and ACMANT are briefly presented. Section 4 shows a review about the present strategies in the practical time series homogenisation, and presents how the modern and effective homogenisation methods are adapted in recently published climate studies of impacted scientific journals. In section 5, some problems related to the use of HOMER and ACMANT are discussed, and recommendations are given.

2. METHODS

2.1. Multiple break homogenisation

Climate time series often contain more than one non-climatic breaks. For instance, in surface air temperature series the usual frequency of breaks is 1 per 20 year or higher according to direct and indirect experiences (*Domonkos*, 2011a; *Venema et al.* 2012). Therefore, one of the important characteristics of a homogenisation method is its approach to eliminate the accumulated effect of multiple breaks. Most of the homogenisation methods focus on the detection and correction of breaks one-by-one, and organises the detection and correction of multiple breaks into a hierarchic or semi-hierarchic procedure. These methods are referred as single break methods, their most known representatives are SNHT (*Alexandersson*, 1986), RHtests (*Wang et al.,* 2007) and PHA – USHCN (*Menne and Williams*, 2009). The other, much smaller group of homogenisation methods treat the multiple break problem with the joint search of breaks occurring in a given time series and joint calculation of adjustments for all the breaks within all the time series examined together. These methods are referred as multiple break homogenisation methods, and their representatives are MASH (*Szentimrey*, 1999), PRODIGE (*Caussinus and Mestre*, 2004), ACMANT (*Domonkos*, 2011b) and HOMER (*Mestre et al.,* 2013). Single break methods have limited potential to provide accurate homogenisation results for the simple fact that in a hierarchic organisation of a procedure including stochastic decisions the errors of the previous steps are carried over to the later steps of the procedure. The use of single break homogenisation methods should, therefore, be changed up with the use of better performing multiple break methods.

Note, however, that the accuracy of homogenisation depends on other properties of the homogenisation methods than their approach to the multiple break problem, e.g., the methodology of the time series comparisons, any iterations, the organisation of the work between varied time scales (annual, monthly, daily) also impact the efficiency. The analysis of all factors influencing the efficiency of homogenisation is beyond the scope of this study. In the following part of this section, the known multiple break techniques are briefly described and the selection of the Optimal Segmentation and ANOVA Correction as most preferred techniques is reasoned.

The use of Optimal Segmentation for inhomogeneity detection (see its description in Sect. 2.2) is introduced by PRODIGE to the climate time series homogenisation. It applies the best fitting step function providing the theoretically possible best approach for the characterisation of the inhomogeneities in time series with pre-set number of breaks. The method includes a semi-empirical algorithm for the determination of the number of breaks. It is based on a relatively simple model and formulas and its computational time demand is low. The flexible structure of the model was favourable for the development of the bivariate detection for annual means and summer – winter differences (included in ACMANT and HOMER), which is an especially powerful tool for the homogenisation of temperature series of mid- or high-latitude origin.

ANOVA Correction (see its description in Sect. 2.3-2.4) was also introduced by PRODIGE to the climate time series homogenisation. It provides the optimal estimation of adjustment terms when the break positions are known. It is based on the minimisation of the residual variance after adjustments. The computational demand of the method is moderated.

MASH offers alternative tools for multiple break homogenisation. In MASH, the most likely break structure is selected by hypothesis testing, examining all the possible combinations of breaks. Then the adjustments terms are deduced from confidence intervals for the estimated break sizes generated during the hypothesis testing. Note that although these adjustment estimations are based on multiple comparisons between time series, the method does not use all pieces of information together as ANOVA Correction. An important part of the philosophy of MASH is the iterative improvement of the quality of time series. The computational time demand of MASH is 10^2 – 10^3 times higher than that of ACMANT, which makes somewhat difficult to test MASH on large datasets.

Joint Segmentation (*Picard et al.* 2011) performs a concerted search for all the breaks in the time series homogenised together. The method is based on an iterative and partly modified application of Optimal Segmentation and ANOVA Correction. Regarding the efficiency of this method, the higher level organisation of break search is promising, but as it includes iterations and empirical parameterisation, the method should be tested on various kinds of homogenisation tasks, which has not been done yet. Recently it was reported that Joint Segmentation sometimes destroys true climate signals (*Gubler et al.,* 2017). Note here that the original development of was done for biostatistical application (*Picard et al.,* 2011), which is different in its nature from climate data homogenisation, so that the adaptation of the method to climate studies should be checked before its use.

Li and Lund (2012) developed the method Minimum Description Length, which is an alternative tool for searching the most likely break structure in time series. Instead of examining all the possible combinations, they apply a genetic algorithm for finding the optimal solution. The method includes the estimation of autocorrelations from the time series under homogenisation, which might seem valuable for the use of a more realistic model setup, but the inclusion of the elevated number of parameters also includes additional error sources. This method never has participated in international tests, and has not yet been used in practical climate research.

Hereafter the methods Optimal Segmentation and ANOVA Corrections are focused in this study, these techniques are preferred principally for their proven high efficiency. *Domonkos* (2011a) showed with a large number of experiments that Optimal Segmentation provides the smallest residual root mean squared error (RMSE) and smallest residual trend bias among the break detection methods widely used in climate data homogenisation, although the exceedance of the efficiency is not big relative to some other detection methods.

Further test experiments with the COST ES0601 (HOME) contributions showed that the performance of any homogenisation method can be improved with the inclusion of ANOVA Correction (*Domonkos et al.,* 2011). In that experiment, detected break lists of the examined homogenisation methods were treated as input for the ANOVA Correction, then the residual errors with and without ANOVA Correction were compared (*Figure 1*).

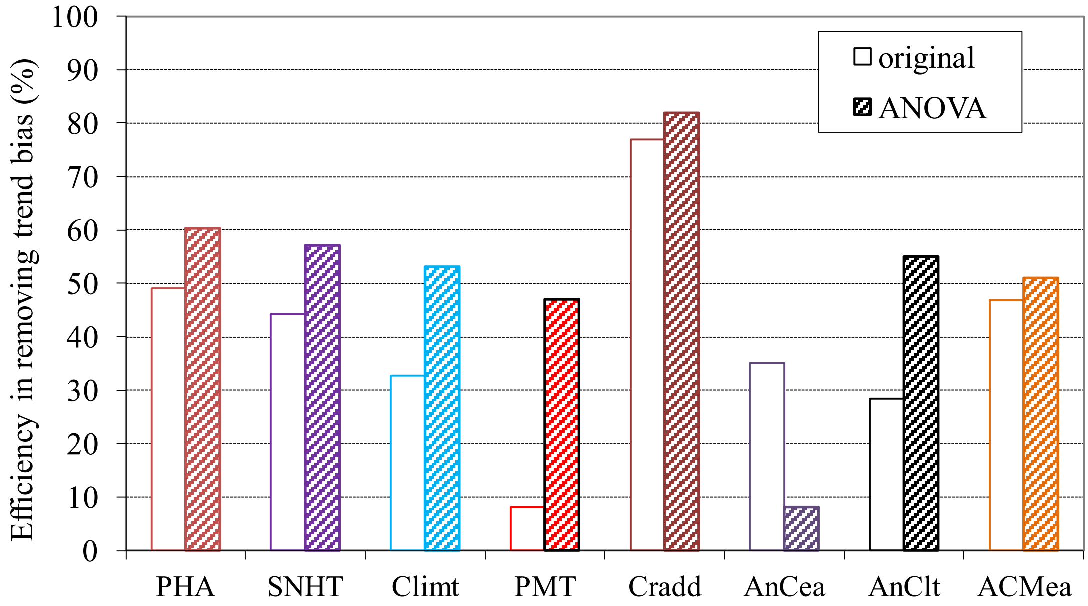


Fig. 1. Improvement of efficiency of HOME participated homogenisation methods without ANOVA Correction in their original algorithm, by supplying the algorithm with ANOVA Correction. Climt = Climatol, Cradd = Craddock test, AnCea = Anclim early, AnClt = AnClim late, ACMea = ACMANT early.

Unfortunately, we could not examine MASH, as it did not provide a compatible break list. We found increasing error with ANOVA Correction only for the early version of AnClim, but likely due to inconsequencies in the detected break list (i.e. AnClim early did not use the same break list in its homogenisation, as which were published and used together with ANOVA Correction, Venema, V. personal communication).

2.2. Optimal Segmentation for the joint detection of breaks within time series

Let consider the dataset of N station series ($s = 1,2,\ldots N$) with n observations in each:

$$\mathbf{X_s} = x_{s,1},\ x_{s,2},\ \ldots\ x_{s,n} \qquad (1)$$

Before starting the homogenisation of candidate series s, differences or ratios relative to the series of the neighbouring stations are constructed, in order to remove the regionally common climate signal from the temporal evolution of the data. These relative time series are denoted with $\mathbf{T_s}$. $\mathbf{T_s}$ series can be derived either by pairwise comparisons or by the use of composite reference series (*Mestre et al.,* 2013; *Domonkos and Coll*, 2017). Index s is shown hereafter only when it needs for the understanding.

Let suppose that **T** has K breaks (k = 1,2,…K) with unknown timings of **H** = h_1, h_2,… h_K. These breaks split the time series to K+1 homogeneous sections (steps). The concept of optimal step function fitting is to minimise the internal distance (d) within steps and maximise the external distance (e) between steps with the examination of all possible **H** (Eqs. 2-4), then the break timings of the minimal sum of d is retained as optimal solution.

$$d(t_i) = t_i - \overline{\mathbf{T_k}} \quad \text{where} \quad i \in k \tag{2}$$

$$e(\overline{\mathbf{T_k}}) = l_k(\overline{\mathbf{T_k}} - \overline{\mathbf{T}}) \tag{3}$$

$$\min_{[h_1, h_2, \ldots h_K]} \left\{ \sum_{k=0}^{K} \sum_{i=h_k+1}^{h_{k+1}} (d(t)_{k,i})^2 \right\} \tag{4}$$

In the equations, l denotes the length of constant section, upper stroke denotes section average, and in case of no label shown, it shows the average for the whole series. The first step is labelled with 0 and the last point of the last step appears as break K +1 only for simplicity. It can be seen that Eqs. (2-4) provide the best approach to the true variation of **T** for a given K by definition, either all of the inhomogeneities are sudden shifts, or when the shape of one or more inhomogeneities are different. These features are the main source of the power of the Optimal Segmentation. However, the best fitting step function depends on the generally unknown K. The sum of the internal distances decreases with increasing K, and it arrives to zero when $K = N - 1$. Therefore, in a good searching algorithm, low summarised d with relatively low K must be preferred. The semi-empirical Caussinus – Lyazrhi formula was developed for finding good balance between the minimization of d and that of K (*Caussinus and Lyazrhi*, 1997).

$$Z = \ln\left\{1 - \frac{\sum_{k=0}^{K} l_k e(\overline{\mathbf{T_k}})}{\sum_{i=1}^{n} (t_i - \overline{\mathbf{T}})}\right\} + \frac{2K}{n-1}\ln(n) \tag{5}$$

$$l_k = h_{k+1} - h_k \tag{6}$$

It can be seen that Z decreases with increasing e and with decreasing K. The optimal solution is provided by the minimal Z for all possible K and **H**.

2.3. ANOVA Correction model for spatially constant climate within network

An observed climatic value can be considered the sum of climate effect (u), station effect (v) and white noise (ε).

$$\mathbf{X} = \mathbf{U} + \mathbf{V} + \boldsymbol{\varepsilon} \tag{7}$$

In the simplest model, the climate is spatially constant within the network homogenised together. (Time series of high geographical distance in their origin or of low spatial correlations are not recommended to homogenise together.) If the time series is homogeneous, the station effect is constant for the whole series, while inhomogeneities are represented as changes in **U**. The noise is the composition of the difference in the local weather from the spatial mean weather and the non-systematic errors of climate observation. The optimal estimations for **U** and **V** can be obtained by the minimisation of the residual noise after bias adjustments, presuming that the break positions are known (*Caussinus and Mestre*, 2004). The practical solution of this task is provided by Eqs (8-9). Estimations are marked with apostrophe in the equations.

$$Nu_i^{'} + \sum_{s=1}^{N} v_s^{'} = \sum_{s=1}^{N} x_{s,i} \quad \text{for every } i \in [1,n] \tag{8}$$

$$\sum_{i=h_k+1}^{h_{k+1}} u_i^{'} + l_k v_{s,k}^{'} = \sum_{i=h_k+1}^{h_{k+1}} x_{s,k} \quad \text{for every steps of every } s \tag{9}$$

2.4. ANOVA Correction model for spatially changing climate

The true climate is not constant spatially, and it can be considered in the ANOVA Correction model by giving spatially varied weights (w) to the observed values in Eqs. (9), forming Eqs. (10-11). In this model each site (s*) has a specific climate.

$$\sum_{s=1}^{N} w_s u_{s*,i}^{'} + \sum_{s=1}^{N} w_s v_s^{'} = \sum_{s=1}^{N} w_s x_{s,i} \tag{10}$$

$$\sum_{i=h_k+1}^{h_{k+1}} u_{s*,i}^{'} + l_k v_{s,k}^{'} = \sum_{i=h_k+1}^{h_{k+1}} x_{s,k} \tag{11}$$

The optimal weights are provided by ordinary kriging (*Szentimrey*, 2010). As the climate is defined separately for each location, one equation system must be solved for each time series of the network, and it multiplies the computational time demand with N. Varied length of time series within network may complicate more the calculations, as the weights must be re-calculated each time when the set of available station series changes. For the latter reason, I suggest the simplification of the weight calculation, i.e. the application of squared spatial correlations as weights provides somewhat suboptimal, but still good solution. According to my knowledge, weighted ANOVA model has not been included yet in homogenisation methods. I have made some comparative experiments between the use of common ANOVA Correction and weighted ANOVA (not shown), and I found that weighted ANOVA results in a very small but consistent improvement in the accuracy of homogenisation. The experienced improvement is small (~1%), likely because the more precise consideration of climate does not touch the main error source, which is the imperfectness of detected breaks.

Note that the use of the name "ANOVA" is sometimes criticised, as this name is already used for with the widely applied variance analysis in multifactor processes. Therefore I suggest to name the presented method with any of "ANOVA model", "ANOVA Correction" or "weighted ANOVA" (when that is applicable) instead of the bare acronym ANOVA.

3. PAST: THE DEVELOPMENT OF HOMER AND ACMANT

3.1. Historical review

Some 45 years ago, US statistician Douglas M. Hawkins invented the Dynamic Programming algorithm for the fast identification of the optimal segmentation among all possible segmentations of a time series (*Hawkins*, 1972). The core idea of Dynamic Programming is that once the optimal break structure is calculated for a given section of the time series, that result can be adopted in determining the optimal break structure for any other section of the time series including the given section. Dynamic Programming provides a striking decline of computational time demand, e.g., for a time series of 100 observations the time demand is ~10^{26} times lower with Dynamic Programming than with repeating all the calculations for each combinatorically possible case. It was historically the first, and likely also the most important step towards the feasible and mathematically correct solution of multiple break homogenisation. However, Dynamic Programming was published in a statistical journal and did not gain interest among climatologists at that time. In addition, Hawkins did not provide suggestions for the assessment of the number of segments, while in climate data homogenisation both the number of segments (breaks) and their positions are unknown. Various statistics can be used for assessing the optimal number of segments, among which the Caussinus – Lyazrhi criterion (*Caussinus and Lyazrhi*, 1997) turned out to be the most viable, and at the time of its publication, in the last years of the 20^{th} century, started the development and adaptation of multiple break homogenisation techniques in climatology.

Henry Caussinus and Olivier Mestre developed PRODIGE (*Caussinus and Mestre*, 2004) adapting all of Optimal Segmentation with Dynamic Programming, Caussinus – Lyazrhi criterion and ANOVA Correction in their method. PRODIGE is an interactive homogenisation method, makes pairwise comparison between time series to separate inhomogeneities from climate signal, and metadata can be used together with the visual comparison of the results of pairwise comparisons. PRODIGE was one of the most successful methods of HOME.

Domonkos (2008, 2011a) tested a large number of break detection methods, and found that Optimal Segmentation generally provides the lowest residual RMSE and lowest trend bias among the tested methods. During HOME, I developed my own homogenisation method, ACMANT (*Domonkos*, 2011b). In ACMANT (= Adapted Caussinus – Mestre Algorithm for the homogenisation of Networks of Temperature series), Optimal Segmentation and ANOVA Correction are adapted, just because of the high efficiency of these techniques were already proven that time. Then in the last year of HOME, the interactive homogenisation method HOMER was composed from the quality control units of Climatol homogenisation method (*Guijarro*, 2011), Joint Segmentation, and from the best units of PRODIGE and ACMANT (*Mestre* et al. 2013).

The development of HOMER and ACMANT was accompanied by supplementary studies about the properties of the methodology included. The efficiency of ANOVA Correction was tested (*Domonkos et al.*, 2011 and *Figure 1* of this study), the appropriateness of the Caussinus – Lyazrhi criterion was analysed (*Lindau and Venema*, 2013) and the accuracy of the detected break positions of Optimal Segmentation was also tested (*Lindau and Venema*, 2016).

In spite of several common approaches within PRODIGE and ACMANT, they differ in many ways, first of all ACMANT uses composite reference series for the time series comparison, it is fully automatic, and includes the bivariate detection of breaks for annual means and for summer – winter differences. As the most striking difference between PRODIGE and ACMANT (and also between HOMER and ACMANT) is the way of time series comparison, the long-standing debate about the best way of time series comparisons is briefly discussed here. Experts' opinions are diverse at this point. Szentimrey suggests the use of composite reference series weighted by ordinary kriging (*Szentimrey*, 2010), while he keeps on using another and unique strategy in MASH, i.e. multiple comparison with varied reference series. I believe that for interactive homogenisation with metadata, the pairwise comparison is the best, as it helps to the in-depth analysis of the data, while in automatic procedures the use of composite reference series is better, as it facilitates the highest possible signal to noise ratio. Note, however, that the practical efficiency might depend more on the iterations included in the homogenisation methods than on the way of time series comparisons, as the potential impact of large inhomogeneities in reference series can be attenuated with proper iterations. All of the HOME recommended homogenisation methods (except Craddock-test) include some iterations.

3.2. HOMER

HOMER is a monthly homogenisation method for the homogenisation of both additive and multiplicative climatic elements. It offers various tools for the effective homogenisation, first of all the same units included in PRODIGE. All pieces of the detection results are displayed in an improved graphical interface, and additional checks of break positions can be done with Joint Segmentation and ACMANT detection. ACMANT detection can help also in detecting season dependent inhomogeneities. In HOMER, user can freely decide about the repetitions of detection – correction cycles, and also about the combination of the detection methods, therefore the application of the method needs some skill. Several World Meteorological Organisation (WMO) sponsored trainings have been organised to teach climatologists how to use HOMER.

3.3. ACMANT

After HOME, I continued the development of ACMANT in two main directions: i) The scale of homogenisation tasks can be solved with ACMANT has been widened; ii) The accuracy of homogenisation has been improved. ACMANTv2 (*Domonkos*, 2014) already included both temperature and precipitation homogenisation, offered different programs for temperature biases with quasi-sinusoid annual cycle (mean and maximum temperatures of mid and high latitudes) and for other temperature data, and provided the downscaling of homogenisation results to daily scale. In ACMANTv3 (*Domonkos and Coll*, 2017), the calculation of monthly temperature adjustment terms with irregular seasonality is provided when that is applicable, the position of large-size breaks are calculated with daily precision, and efficiency improving novelties of ACMANTv3 are the ensemble pre-homogenisation and the application of ordinary kriging for setting the weights of the reference composites.

ACMANT is an easy-to-use, automatic method, in addition, it can be used for homogenising together data records of varied periods, and large data gaps are tolerated.

ACMANTv3 offers the full completion of homogenised datasets to a predefined period. The method has been shown and taught to the participants in several WMO sponsored homogenisation trainings.

I often test the performance of ACMANT with an own developed test dataset. This dataset includes the homogeneous products of earlier dataset developments (HOME benchmark and the US daily temperature data base (*Willett et al*., 2014; *Killick*, 2016). It has 21 segments, each segment is characterised with specific inhomogeneity properties and spatial correlations. These tests (*Figure 2*) show the continuous improvement of efficiency from version to version.

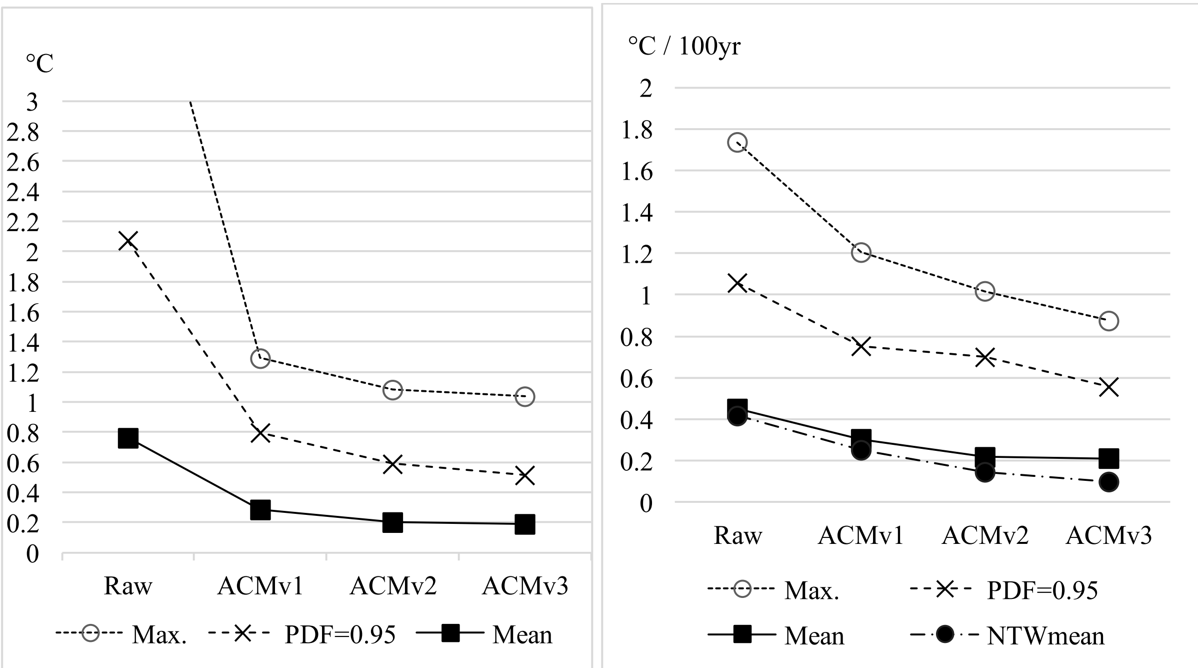


Fig. 2. Raw data error and residual errors after homogenisation with ACMANTv1, ACMANTv2 and ACMANTv3, using a test dataset of 300 networks. Left panel: monthly RMSE, right panel: trend bias. PDF = probability distribution function, NTWmean = network mean trend bias. See more details in Domonkos and Coll (2017).

Beyond the good results of the own made tests, international test experiments show that ACMANT often produces the smallest residual RMSE and smallest residual trend bias among the tested methods (*Killick*, 2016; *Guijarro* et al., 2016).

4. PRESENT: REVIEW OF RECENT PRACTICES IN CLIMATE RESEARCH

4.1. Material and methods

I have reviewed the inclusion of statistical homogenisation in climate studies published between January 2013 and May 2017 in high-impacted scientific journals. I examined the publications in International Journal of Climatology (IJoC), Theoretical and Applied Climatology (TAAC), Climate of the Past and in all journals of the American Meteorological Society (AMS). I split the studies to European and non-European groups, and split also to two groups according to the date of publication, i.e. to publications of 2013-2014 and to those of 2015-2017. Thus temporal and geographical comparisons can be made about the spread of the application of homogenisation methods. I also checked the reasoning of method selection and the ways of giving references to the source studies of the method development and method testing studies.

The frequency of the use of each HOME recommended method (HOMER, ACMANT, MASH, Craddock test, PHA-USHCN), as well as the frequencies of SNHT and RHtests applications were checked. PRODIGE applications were merged with HOMER applications and the CUSUM method with the Craddock test.

Notes: i) It is likely that I have not found all relevant studies, but I intended not to leave out any; ii) If more than one statistical homogenisation methods are applied in a study, each of the methods is taken consideration for the frequency statistics, but methods offering homogenisation on monthly scale are considered only; iii) In the geographical classification, the area of the study is the prioritised factor, and the nationality of authors is the secondary. This classification remained ambiguous only for one-two studies due to the mixed area of raw data and mixed nationality of author group; iv) The date of the publication is the date of the printed publication, except for studies which had only electronic publications until May 2017.

4.2. Statistical results

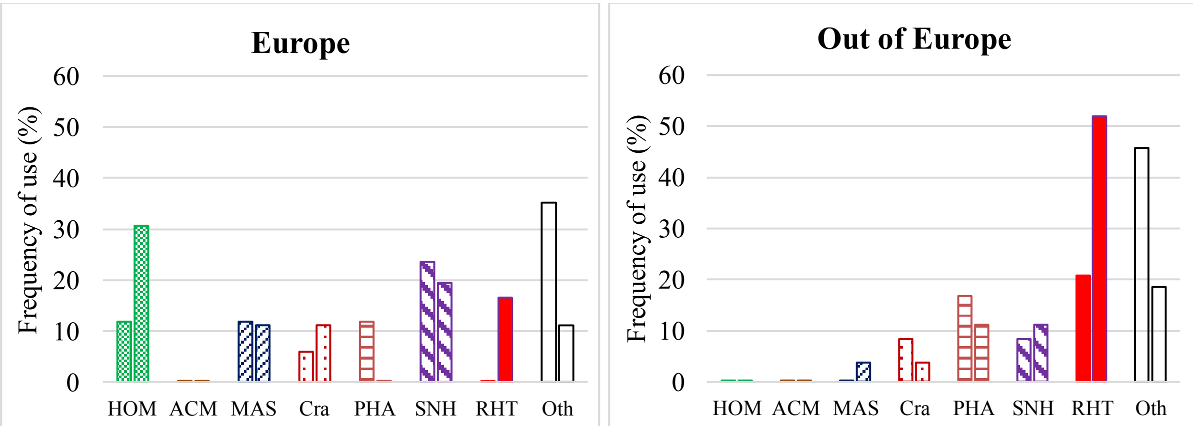


Fig. 3. Frequency of the use of homogenisation methods. For each panel and each method, left columns: 2013-2014; right columns: 2015-2017. Homogenisation methods are denoted with the first 3 letters of their name, Oth = other method.

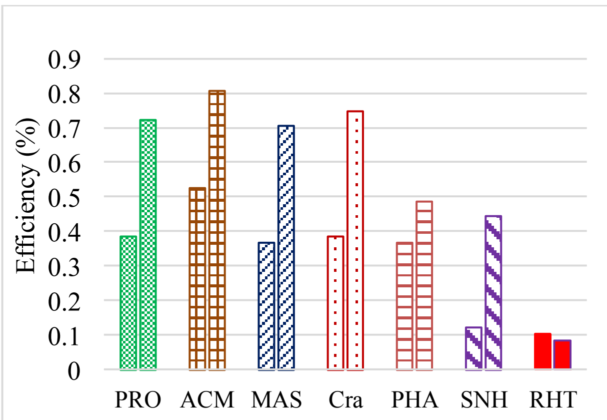


Fig. 4. Efficiency of homogenisation methods in the removal of raw data errors from the HOME temperature benchmark dataset. Left columns: centred monthly RMSE, right columns: trend bias. (Source: Venema et al., 2012)

In the examined journals, I have found 104 cases of applying and referring to a known statistical homogenisation method. The frequency of the application of statistical homogenisation methods for Europe is much higher than for any other continent, namely in 2013-14 17 European applications and 24 out of Europe applications were published, while between 2015-17 36 European applications and 27 out of Europe applications. Note that the number of European studies is particularly high for the latest period, partly because the gap between the dates of electronic and printed publications in the European journals IJoC and TAAC is as large as ~1 year.

The panels of *Figure 3* and *Figure 4* together allow to make direct comparisons between the temporal and geographical distribution of the method applications, as well as between the method efficiencies and the tendency in the frequencies of method applications. It can be seen that the ratio of the use of HOME recommended methods is much higher for Europe than for other continents, i.e. between 2015-17 it is 53% for Europe, while only 15% for non-European studies. Analysing further the results, we can find two significant and very positive features, but at least three significant and very negative features:

- The frequency of the use of HOMER increases fast in Europe, and was the most frequently applied method in Europe between 2015-2017.
- The use of "other methods", which mostly have not even been tested, has become markedly rarer recently, everywhere.
- The European development of multiple break homogenisation has not spread over the European borders yet, at least not in climate research applications.
- ACMANT, in spite of its excellent test results and easy-to-use construction, has not been used yet in climate studies of impacted journals.
- The frequency of the application of RHtests, in spite of its known theoretical weaknesses and relatively poor test results, increases fast in all over the world, and it is the most used method outside of Europe.

Globally, RHtests, HOMER and SNHT were the three most used methods between 2015 and 2017 with frequencies of 31.7%, 17.5% and 15.9%, respectively. Most of the HOMER applications were published in IJoC.

4.3. Reasoning of method selection in climate studies

As it was shown in the previous section, a quite large proportion of the method selections is not supported by the results of international tests. It may be useful to know and understand the reasoning of climate researchers' method selections.

The selection of RHtests is most commonly reasoned by the fact that it is published by and accessible via the homepage of WMO expert team ETCCDI (Expert Team on Climate Change Detection and Indices), hence it is considered the WMO recommended homogenisation method by many users. I found even an example where RHtests was referred as "standard method", and another example where the reference of method description of RHtests pointed to the WMO guidelines on homogenisation released in 2003 (WCDMP 53), suggesting in this way that RHtests is one of the methods or the only method recommended by WMO. RHtests is published under ETCCDI without showing any comment about other homogenisation methods. WMO has another expert team, Task Team on Homogenization (TT-HOM), and this body would be the competent to recommend and make accessible homogenisation methods. However, at this time (June 2017) TT-HOM web pages do not include downloadable software, either recommendations. European project HOME still has living web pages (the project terminated in 2011), and HOMER is downloadable from its homepage, while the other HOME recommended softwares for monthly homogenisation are not accessible there either. Information on several homogenisation methods is accessible via a link from the HOME homepage to a TT-HOM page, but that page does not show clear, efficiency related recommendations either.

Further finding is that climate researchers often trust more on statistical methods whose detected breaks mostly can be validated with metadata, than on methods detecting more breaks. It is a misconception, as metadata lists are generally incomplete, and in any case, the frequency of detected breaks has no direct relation to the accuracy of homogenisation.

A relatively new tendency is that homogenisation packages offering wider scale service (homogenisation of any variable in any time scale, additional quality control, etc.) are most recommended and most likely to be used than other homogenisation packages, therefore several software developers have recently widened the service of their homogenisation packages. It is not necessarily a positive change, as the observation of each climatic variable has specific problems, and if the solution of such problems are not treated adequately during the homogenisation, the reliability of the data might worsen instead of achieving quality improvement.

Finally, there is a very painful experience with AMS journals that they are fully separated from communicating anything about the European development of multiple break homogenisation. Readers of AMS journals never could have learnt from their journal about HOMER, ACMANT or MASH. I found one only exception, a 4-page conference summary mentioned HOMER in the Bulletin of the American Meteorological Society (BAMS) without giving any detail about the method. Another 4-page summary about HOME project was published in BAMS in 2012, but that summary report is not part of the electronic archives of BAMS, thus it is not available for the wider readership. Although the HOME closing report (*Venema et al*., 2012) sometimes referred in AMS journals, I have not found there any incident when any of its five main conclusions (shown in Sect. 1) was mentioned in context with the reference.

Summarising the findings of this section, it can be concluded that the adaptation of the new and valuable scientific achievements of time series homogenisation into the practical climate research is slow and loaded with contradictions. Poor accessibility to the most relevant information about method efficiencies, asymmetric promotion of some arbitrary selected methods, users' misconceptions, forced widening of the service of homogenisation packages and some geopolitical (?) separation of AMS journals from considering the results of European methodological developments make it difficult to select the most effective homogenisation methods in practical climate research.

5. TASKS FOR THE FUTURE

As it was shown in the previous section, we cannot be satisfied with the present practice of time series homogenisation in climate research, so we must think about what we could do better. One question is what we, method developers, could do better, and another question is how other actors of the climate research community could promote the most effective dissemination of the best science.

5.1. Problems with HOMER

We have good, better and even better homogenisation methods, but none of the methods is perfect. The following problems may influence the use of HOMER: i) Joint Segmentation does not separate adequately inhomogeneities from climate shifts; ii) Reported that ACMANT detection tends to give too many breaks; iii) The combined use of varied detection methods has never been tested; iv) Missing data tolerance is limited.

These problems should be examined and solved for offering an even better HOMER to climate researchers. However, one particular problem around any future development of HOMER is that its lead developer, Olivier Mestre has other job than homogenisation development, and thus likely he will not return to develop HOMER. Although I am a contributor to HOMER, I cannot continue that development either, because I use other program language (FORTRAN) than the language of HOMER (R). Therefore, the help of volunteers is needed for the further development of HOMER. Alternatively, we can give clear instructions how to use HOMER safely, what I try to do here:

- Use Pairwise Detection (PD) always as principal tool;
- Use PD both for annual and seasonal series;
- Use Joint Detection (JD) only to decide about ambiguous pieces of detection results of PD, don't shift break positions for better fitting with JD results, as far as JD is not adequately tested;
- Use ACMANT detection only to decide about ambiguous pieces of detection results of PD

Following these instructions, HOMER functions with the same or higher efficiency than PRODIGE, so that it can be used safely. Note that I recommend the use of HOMER for networks of up to 12-15 series and when metadata is available, as for larger networks the probability of committing subjective errors is increased and the importance of active metadata use is relatively little. When the spatial density of time series allows to form larger networks, I recommend first the use of ACMANT.

5.2. Problems with ACMANT

The following characteristics of ACMANT sometimes considered to be weaknesses:

i) ACMANT tends to detect too many breaks and make unnecessary adjustments;

ii) Metadata cannot be used together with ACMANT;

iii) Automatic selection of partner series is not offered;

iv) ACMANT can be used only for homogenising temperature and precipitation;

v) Homogenisation of higher moments and that of probability distribution function (PDF) are not included;

vi) Graphical output is not provided;

vii) The inclusion of weighted ANOVA could improve the accuracy, so that ACMANTv3 is still suboptimal.

I continue the software development, and I plan to step forward in a few issues. I think of publishing ACMANTv4 in 2018. Here the potential issues for improvements are discussed briefly.

i) The frequency of detected breaks will unlikely change significantly in later ACMANT versions. First, because small-size breaks have minor effect on the accuracy of the homogenisation, and thus it is more useful to optimise directly the more robust efficiency measures of residual RMSE and residual trend bias than the break frequency. Second, the experience with unnecessary adjustments is connected to the use of perfectly homogeneous time series in test datasets, but perfectly homogeneous time series are unlikely to occur in true observational datasets, as dozens of factors may cause changes in the conditions of the climate observations.

ii) The original philosophy of ACMANT development is that it is recommended first of all for the homogenisation of spatially dense datasets where the potential role of metadata is limited due to the large number of spatially comparable data. However, a small improvement likely can be realised with inbuilt metadata use, and I am determined to provide such option in the next ACMANT version.

iii) ACMANTv4 will include the option of automatic network formation, and will be prepared to homogenise input datasets of up to 5000 time series.

iv) I will examine the possibilities of applying ACMANT programs to other climatic elements than temperature and precipitation, but I remain careful, as I consider the reliability of homogenisation products to be more important than to offer the homogenisation of everything.

v) Only the means can be homogenised in fully automatic mode. In each time series homogenisation, one step is the selection of the reference period. This reference period is a section of the time series to which the data of the other sections are adjusted during the homogenisation, while the data of the reference period are usually kept unchanged. In homogenising the means only, there is a convention to use the last homogeneous section of the time series as reference period, and it works well even if the data in the last section of the time series are biased. However, this convention cannot be extended to the homogenisation of higher moments or to PDF homogenisation, as observational errors of the last homogeneous section may change the statistical properties of the homogenisation product in various ways including unreasoned changes in the temporal variability of the means. Therefore, in the homogenisation of higher moments and PDF homogenisation, the selection of reference period needs the manual check of the data of the candidate series and some expertise. ACMANTv4 will remain fully automatic, without offering PDF homogenisation.

vi) Unfortunately, I am not very familiar with the edition of graphical outputs, so I cannot offer improvement at this point. I believe that the important results of automatic homognisation can be presented in data tables.

vii) ACMANTv4 will surely include weighted ANOVA and likely also some other methodological modifications aiming to improve efficiency further.

5.3. More efficiency tests are needed

The main source of our knowledge about the efficiency of homogenisation methods is test experiments with artificial test datasets. The homogenisation community has three internationally announced test projects until now, HOME (2007-2011), US daily temperature homogenisation program of Kate Willett and Rachel Killick (2014-2015), and a Spanish national project, MULTITEST (2015-2017). Among these three, HOME tests were the best organised, but its benchmark dataset was small: the participated methods were tested only with 111 temperature time series and 111 precipitation time series. Although only automatic homogenisation methods can be tested in very large test datasets, we need such tests, as high variety of inhomogeneity scenarios and diversity also in the climatic properties are needed for obtaining reliable conclusions. It is the main goal of MULTITEST, but as a modest national project, its program has remained smaller and partly oversimplified relative to the scientific needs. While thousands of millions of euros are spent worldwide for climate observations and data archiving, the financial support of data quality control and time series homogenisation is practically zero. MULTITEST project functions with 0.4% of the budget of the European project HOME, and it obviously limits the possibilities of a project. And all this scarcity is in spite of the great importance of how climatic data are treated with quality control and time series homogenisation. Only some small mistakes in a homogenisation procedure are enough to turn data homogenisation into data ruining.

One might doubt that the use of artificially developed test datasets is a reliable source of knowledge, as we might fail in reproducing the true properties of observed climatic data. While it is true that reproductions cannot be perfect, we can produce good approaches. Such projects should be international projects just for minimising the risk of mistaken approaches, to provide great diversity of tests and to obtain and maintain the necessary international attention. Although interactive homogenisation methods cannot be tested in very large test datasets, the statistical procedures included in them can, so that we could gain useful information also about the efficiency of interactive methods by tests on large-size, artificially developed test.

The next international tests of homogenisation methods will likely be done under the International Surface Temperature Initiative (ISTI), as its benchmark dataset is under preparation.

5.4. Knowledge transfer

One principal source of knowledge transfer in science is the scientific literature. However, when a topic is well known only to very few people, the selection of the best science is not always automatic by scientific journals. Therefore, scientific bodies should orient researchers and journal editors with disseminating knowledge. In our case, the most competent scientific body is TT-HOM. As a minimum, the most relevant information about method efficiencies should be easily accessible directly from the homepage of TT-HOM. In addition, if one method is accessible from a WMO website or international project website, all the other recommended methods should be accessible from there.

Anybody may publish his own standpoints and opinions disregarding the standpoints of competent scientific bodies and disregarding international test results, also the ETCCDI may do that. The only thing that should be avoided is to publish private opinions in a way that they might be seemed standpoints of professional consensus for readers.

Optimally, knowledge transfer should include organised teaching of data quality control and time series homogenisation. Misconceptions and misunderstandings are unfortunately frequent around the topic of time series homogenisation, and although often even experts' views differ, we should try to acquire financial support and establish organised teaching about time series homogenisation.

6. SUMMARY AND CONCLUSIONS

Theoretical and practical aspects of the inclusion of Optimal Segmentation and ANOVA Correction in time series homogenisation have been analysed. It was found that in spite of their outstanding theoretical values and excellent test results, their spread in practical climate research is slow and loaded with contradictions. The practical selection of homogenisation methods often shows indications of free market effects and lobby effects, which are generally unfavourable from scientific aspect, although it is true that free market effects give some positive inspiration to the methodological development.

Observed climatic datasets have huge value both in scientific and financial sense. Statistical treatments as time series homogenisation can significantly increase or decrease this value according to what we do with the data. Therefore, the selection of applicable homogenisation methods should be based on well concerted and highly professional decisions. In this study I gave recommendations what the homogenisation community should do to assure high level professionality in the practical treatments of observational datasets.

Finally, as according to our present knowledge HOMER and ACMANT give the most accurate homogenisation results in a large scale of homogenisation tasks, I do everything possible to put these methods into their right place in practical data treatments.

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.

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

Caussinus H, 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, Mestre O (2004) Detection and correction of artificial shifts in climate series. *J Roy Stat Soc C* 53: 405-425. doi: 10.1111/j.1467-9876.2004.05155.x.

Domonkos P (2008) Testing of homogenisation methods: purposes, tools and problems of implementation. Proc 5th Seminar for Homogenisation and Quality Control in Climatological Databases. (Ed Lakatos M, Szentimrey T, Bihari Z, 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.

Domonkos P (2014) The ACMANT2 software package. In: Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques In Climatology and Meteorology (Ed Lakatos M, Szentimrey T, Marton A) WMO, WCDMP-84: 46-72.

Domonkos P, Coll J (2017) Homogenisation of temperature and precipitation time series with ACMANT3: Method description and efficiency tests. *Int J Climatol* 37: 1910-1921. doi: 10.1002/joc.4822.

Domonkos P, Venema V, Mestre O (2011) Efficiencies of homogenisation methods: our present knowledge and its limitation. In: Seventh Seminar for Homogenisation and Quality Control in Climatological Databases (Ed Lakatos M, Szentimrey T, Vincze E), WCDMP-78, WMO, Geneva, 19-32.

Gubler S, Hunziker S, Begert M, Croci-Maspoli M, Konzelmann T, Brönnimann S, Schwierz C, Oria C, Rosas G (2017) The influence of station density on climate data homogenization. *Int J Climatol*, doi: 10.1002/joc.5114 (early view).

Guijarro JA (2014) User's Gide to Climatol. http://www.meteobal.com/climatol/climatol-guide.pdf.

Guijarro JA, López JA, Aguilar E, Domonkos P, Brunet M (2016) Benchmarking homogenization computer packages: First results of the MULTITEST project. In: 16th Ann Meeting of the European Meteorol Soc, Trieste Italy EMS2016-325.

Hawkins DM (1972) On the choice of segments in piecewise approximation. *J Inst Math* Appl 9: 250-256.

Killick RE (2016) Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data. PhD thesis, University of Exeter, UK. https://ore.exeter.ac.uk/repository/handle/10871/23095.

Li S, Lund R (2012) Multiple changepoint detection via genetic algorithms. *J Clim* 25: 674-686. doi: 10.1175/2011JCLI4055.1.

Lindau R, Venema V (2013) On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records. *Időjárás* Q J Hung Meteorol Serv 117: 1-34.

Lindau R, Venema V (2016) The uncertainty of break positions detected by homogenization algorithms in climate records. *Int J Climatol* 36: 576-589. doi: 10.1002/joc.4366.

Menne MJ, Williams CN (2009) Homogenization of temperature series via pairwise comparisons. *J Clim* 22: 1700–1717. https://doi.org/10.1175/2008JCLI2263.1.

Mestre O, Domonkos P, Picard F, Auer I, Robin S, Lebarbier E, Böhm R, Aguilar E, Guijarro J, Vertacnik G, Klancar M, Dubuisson B, Štěpánek P (2013) HOMER: homogenization software in R – methods and applications. *Időjárás* Q J Hung Meteorol Serv 117: 47–67.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Böhm 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.

Picard F, Lebarbier E, Hoebeke M, Rigaill G, Thiam B, Robin S (2011) Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics* 12: 413-428. doi: https://doi.org/10.1093/biostatistics/kxq076.

Szentimrey T (1999) Multiple Analysis of Series for Homogenization (MASH). Proc 2nd Seminar for Homogenization of Surface Climatological Data (Ed Szalai S, Szentimrey T, Szinell Cs). WMO WCDMP 41: 27-46.

Szentimrey T (2010) Methodological questions of series comparison. In: 6th Seminar for Homogenization and Quality Control in Climatological Databases (Ed Lakatos M, Szentimrey T, Bihari Z, Szalai S). WMO WCDMP-76: 1-7.

Venema V, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey T, Štěpánek 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, Duran MP, Likso T, Esteban P, Brandsma T (2012) Benchmarking monthly homogenization algorithms. *Clim Past* 8: 89–115. doi: 10.5194/cp-8-89-2012.

Wang XL, Wen QH, Wu Y (2007) Penalized maximal t test for detecting undocumented mean change in climate data series. *J Appl Meteorol Climatol* 46: 916-931. doi: 10.1175/JAM2504.1.

Willett KM, Williams CN, Jolliffe I, Lund R, Alexander L, Brönniman S, Vincent LA, Easterbrook S, Venema V, Berry D, Warren R, Lopardo G, Auchmann R, Aguilar E, Menne M, Gallagher C, Hausfather Z, Thorarinsdottir T, Thorne PW (2014) A framework for benchmarking of homogenisation algorithm performance on the global scale, *Geosci Instrum, Meth Data Sys* 3: 187-200. doi: 10.5194/gi-3-187-2014.

COMPARISON OF HOMOGENIZATION PACKAGES APPLIED TO MONTHLY SERIES OF TEMPERATURE AND PRECIPITATION: THE MULTITEST PROJECT

José A. Guijarro[1], José A. López[1], Enric Aguilar[2], Peter Domonkos, Victor K.C. Venema[3], Javier Sigró[2] and Manola Brunet[2]

[1]State Meteorological Agency (AEMET), Balearic Islands Office, Spain
<jguijarrop@aemet.es>
[2]Universitat Rovira i Virgili, Tarragona, Spain
[3]Meteorological Institute, University of Bonn, Germany

1. INTRODUCTION

It is well known that observational climatic series are exposed to unwanted alterations due to changes in the observational practices, instrumentation, relocations or changes in the surroundings of the stations. Many methods have been proposed to remove these perturbations from the series and leave the climate signal only, and the successful COST Action ES0601 "HOME" provided interesting inter-comparison results to understand the strengths and weaknesses of many of them (*Venema et al*., 2012).

However, some methods, implemented in computer packages, have been upgraded to new versions (in part as a result of the fruitful discussions maintained along the COST Action), and therefore new inter-comparisons are needed to evaluate their performance. Yet repeating such an effort, which involved the work of dozens of researchers along five years, is not foreseen in a near future. Hence, the only practical alternative is to implement a benchmarking system to test the homogenization packages in a completely automatic way. The drawbacks are that only packages able to be run in this mode can be tested, with default parameters in those that can be tuned to different climatic variables, and that the added value of manual homogenization is lost.. But on the other side, the methods can be tested on a high number of networks with varying characteristics, which enhance the knowledge about its applicability to different climatic zones.

The results of a preliminary automatic comparison on synthetic monthly temperature series can still be found at http://www.climatol.eu/DARE/testhomog.html, but the Spanish project MULTITEST (Multiple verification of automatic software homogenizing monthly temperature and precipitation series) aims at updating and improving those benchmarking experiments in various ways:

- More realistic temperature networks.
- Inclusion of precipitation networks with different climatic characteristics (Temperate, Mediterranean and Monsoonal).
- More realistic inhomogeneities.
- Comparison of more homogenization packages.

The details of the benchmarking implementation, the packages tested and the results obtained so far are explained in the following sections, ending with some conclusions and prospects of future work.

2. BENCHMARKING METHODOLOGY

2.1. General procedure

Several "master" networks were generated consisting in 100 series containing 720 monthly values (equivalent to 60 years of data). For each of them, for each of several different inhomogeneous settings (experiments), and for each chosen homogenization package, 100 tests were done by:

- Randomly sampling a subset of 10 series (true solution). (Some supplemental tests were done with 20, 40 and 80 series.)
- Inserting inhomogeneities into them (problem series).
- Homogenizing them with a backward adjustment (results).
- Comparing the results with the true solutions, computing Root Mean Squared Errors (RMSE), trend differences, and other metrics.

2.2. Homogeneous synthetic series

Six homogeneous master networks were generated, three for temperatures and three for precipitations. The temperature networks were built in the following way:

- Random locations were assigned to 100 points in a 4x3° lon-lat geographic domain.
- Mean monthly homogenized temperatures from Valladolid (Duero basin, Spain) were assigned to the first point, located near the center of the domain.
- Its closest point was assigned the same series plus white noise from a normal distribution with mean = 0 and standard deviation = 1.5, multiplied by a constant C. The rests of the series were computed with the same procedure, in order of minimum distance to any other already assigned series.
- Three different constant coefficients were used: C = 0.18, 0.30 and 0.65, yielding three master networks with decreasing correlation between stations, which were called Tm1, Tm2 and Tm3 (*Figure 1*, left column).
- All series were shifted to account for simulated elevation, a 2 °C/100yr trend was added, and their annual oscillation were varied in ±20%.

The three monthly master precipitation networks were generated taking as models real precipitation series from three different climatic regions from which, after their homogenization by Climatol 3.0, derive variograms, gamma coefficients and frequencies of zeros, which were used to compute synthetic series by means of the R package "gstat", preserving the spatial correlation structure. The names assigned to this networks, type of simulated climates and data used to model them were:

- PEir (Atlantic temperate): 198 Irish precipitations series (1941-2010).
- PMca (Mediterranean): 107 Majorcan precipitation series (1951-2015).
- PInd (Monsoonal): 64 SW India series from 0.5° resolution gridded monthly precipitations from the Global Precipitation Climate Center (Schneider, 2015).

Figure 1 (right column) shows the cross-correlations of these monthly data, computed on the first differences of the series.

2.3. Added inhomogeneities

In a first stage, inhomogeneities were applied to the synthetic homogeneous series. We used five different settings with increasing difficulty and realism:

- i) Big shifts in half of the series.
- ii) The same with a strong seasonality.
- iii) Short term platforms and local trends.
- iv) Random number of shifts with random size and location in all series.
- v) The same plus seasonality of random amplitude.

For setting ‘i’, the shifts have a size of 2°C; series 1 to 3 have fixed position and signs (“-,-”, “+,+” and “-,+” respectively); series 4 and 5 have only one shift, but with random sign and position. For the case with seasonal cycle (ii) the size of the shifts is 1.5 or 2°C and the seasonal cycle is a sinusoidal function with an amplitude of 2°C. The size of the platforms and linear gradients used for setting ‘iii’ are 2°C, with random lengths within certain limits to avoid overlapping. The number of shifts in settings ‘iv’ and ‘v’ was taken from a Poisson distribution with a mean of 5 every 100 years. The shifts were applied as deviations from the baseline, additive for temperature, with an amplitude drawn from the standard normal distribution N(0, 1), and multiplicative for precipitation, with factors taken from N(1, 0.2) without any seasonal perturbation (only setting ‘iv’ was applied to this variable). Seasonality amplitudes were taken from N(0, 0.7). In all cases, the last 10 years are always kept untouched. Sample examples of these five settings can be seen in *Figure 2*.

2.4. Tested homogenization packages

Most commonly used homogenization programs that could be run in completely automatic mode were tested, namely:

- Climatol 3.0 (*Guijarro*, 2016), with constant and variable corrections.
- ACMANT 3.0 (*Domonkos*, 2015), versions for temperature and precipitation (sinusoidal and irregular seasonalities).
- MASH 3.03 (*Szentimrey*, 2007), with constant corrections.
- RHtestsV4 (*Wang & Feng*, 2013), absolute and relative (average series were given as references), with or without quantile adjustment.
- USHCN v52d (*Menne & Williams*, 2005), which makes constant corrections.
- HOMER 2.6 (*Mestre et al.,* 2013), with different iteration strategies.

Tests were run on a Linux PC by means of bash scripts. USHCN was compiled on a Linux computer and could be run natively. Climatol, RHtestsV4 and HOMER are implemented in R (R Core Team, 2015), and hence could also be easily run, although HOMER could not be automated by a simple redirection of the input from a bash script and the utility “expect” needed to be used to provide automatic responses to the questions of the program.

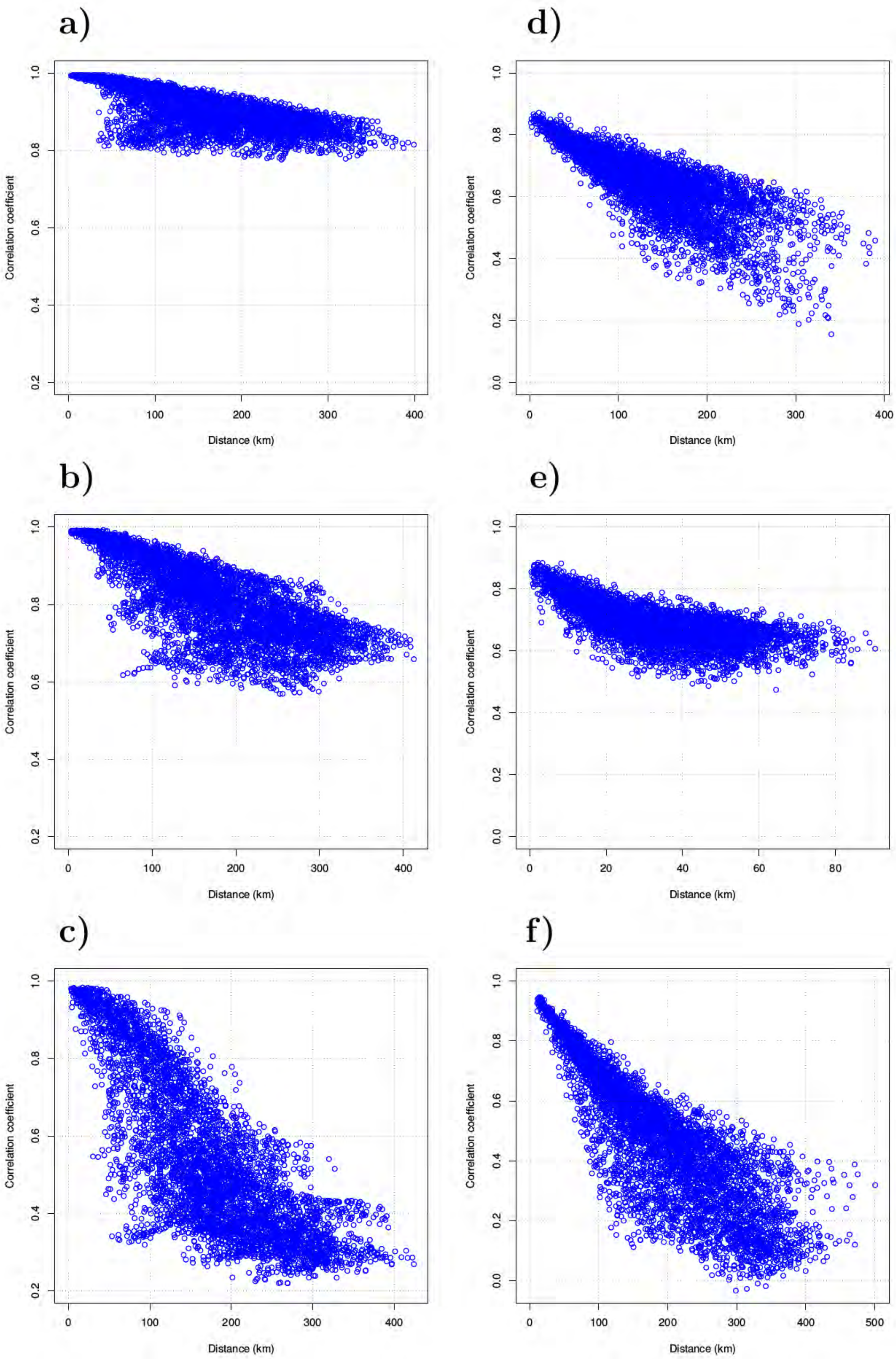


Fig. 1. Correlograms of the first differences of the master networks. Temperatures in the left column: a) Tm1, b) Tm2 and c) Tm3. Precipitations in the right column: d) PEir, e) Pmca, f) Pind. (Vertical axis ranges from 0.2 to 1.0 in the left column and from 0.0 to 1.0 in the right one).

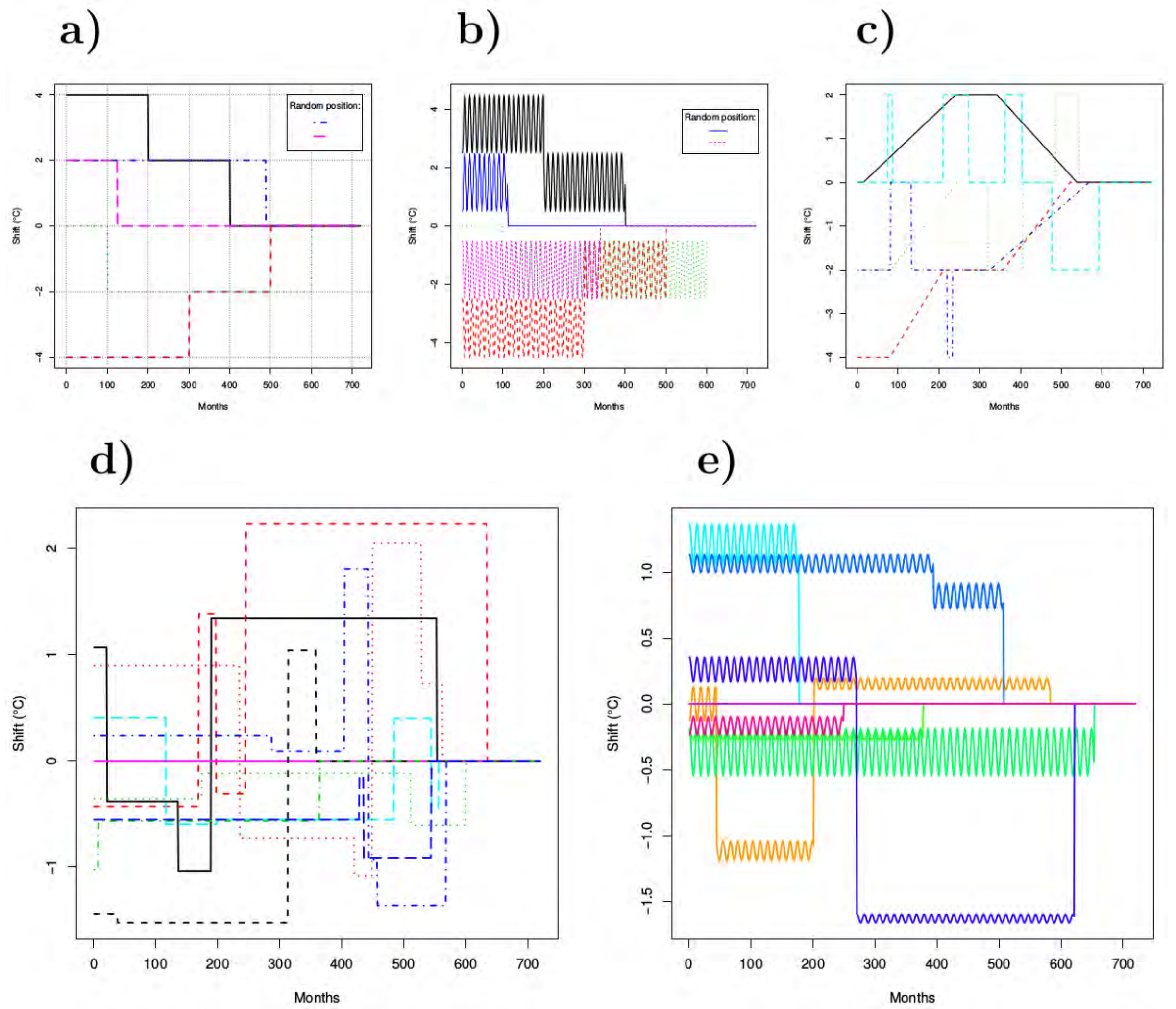


Fig. 2. Examples of inhomogeneities with increasing complexity introduced in 10 series sample networks: a) Small number of big constant shifts in five of the series. b) The same, but with a strong seasonal variation in the shifts. c) Local trends and short term "platforms". d) Random number of shifts with random constant amplitude can affect all 10 series. e) The same, but with a seasonal variation in the shifts of random amplitude. The last 10 years of the series are kept homogeneous in settings a) to c), and only the 5 last years in d) and e), to allow a reliable backward reconstruction of the homogenized solutions.

ACMANT and MASH are Windows executables, but could be run on Linux through "wine" (a Linux implementation of the Windows Application Interface). This was straightforward in the case of ACMANT, but MASH automatic procedures are implemented in DOS batch scripts that gave errors of incorrect file specification in wine when "*." were copied or moved between directories, so these DOS batch scripts had to be translated into Linux bash versions.

Whenever a method stop with an error condition (sometimes simply stating that the problem series was homogeneous, sometimes due to a limitation of the software), the problem series was taken as the solution provided by the tested method. This procedure allowed the unsupervised run of hundreds of tests in a continuous flow, but in some experiments HOMER gave more serious errors that aborted the process, giving incomplete solutions.

3. RESULTS AND DISCUSSION

The performance of the different methods in each experiment is evaluated mainly by looking at the Root Mean Squared Errors (RMSE) of the solutions returned by the different software packages, and also by comparing their trends with those of the original series. Box-plot figures allow an easy comparison of the performances of the methods between them and with respect to the problems (inhomogeneous networks) they had to solve.

3.1. Temperatures

Figure 3 shows the RMSE box-plots of the methods for the five different settings displayed in *Figure 2*, using the network Tm2 (intermediate level of cross-correlations). Horizontal axis labels are: “Inh” (Inhomogeneous, problem series), “cl1” (Climatol with constant correction, “Cl1” (Climatol with variable correction), “A3i” (ACMANT with irregular seasonality), “A3s” (ACMANT with sinusoidal seasonality), “MSH” (MASH), “RHa” and “RHA” (RHtestsV4 in absolute mode, without and with quantile matching correction), “RHr” and “RHR” (RHtestsV4 with reference series, without and with quantile matching correction), “US1” (USHCN), “Hoa”, “Hob” and “Hom” (HOMER, with different iterative approaches).

It is clear that absolute homogenization without strong metadata support should be avoided. (An exception will be discussed later on). All other relative homogenization methods provide results clearly better than the inhomogeneous (“Inh”) problem, although with a different performance degree. In particular, comparison of results from experiments “a” (where RHtestsV4 gives the best results) and “b” show a clear improvement in the methods that are able to correct a seasonally variable bias, the lowest RMSE being in this case achieved by ACMANT and HOMER. As to experiment “c” (short term platforms and local trends), ACMANT is also ranking the best, closely followed by MASH, Climatol and RHtestsV4 without quantile adjustment.

The lower row in *Figure 3* displays the results for the more realistic experiments. When the random inhomogeneities do not have a seasonal cycle (“d”), the lowest mean RMSE correspond to ACMANT sinusoidal, Climatol (with varying correction ranking better than the constant correction version!) and USHCN, followed not too far away by the other relative methods. And when more realism is added by imposing a sinusoidal seasonal cycle of the inhomogeneities (“e”), ACMANT (sinusoidal and irregular) is still ranking the best, followed by Climatol with variable correction and, more distantly, HOMER, USHCN, MASH and RHtestsV4 with quantile adjustment.

Figure 4 shows these last more realistic results in the middle left box-plots (4b), which can now be compared with the same results when the sample problems are drawn from the better (Tm1) and worse (Tm2) correlated master series (4a and 4c respectively). The performance of the methods decay with a decreasing level of cross-correlation in the networks as could be expected, but the ranking is quite constant in 4a and 4b. When the correlations between stations are worse (4c), ACMANT is still the method producing the lower RMSE, while the other methods give more similar results. It is worth mentioning that in this latter case the quantile adjustment worsens the performance of RHtestsV4, and that, even in this lower correlated scenario, all relative methods return series more homogeneous than the problem networks.

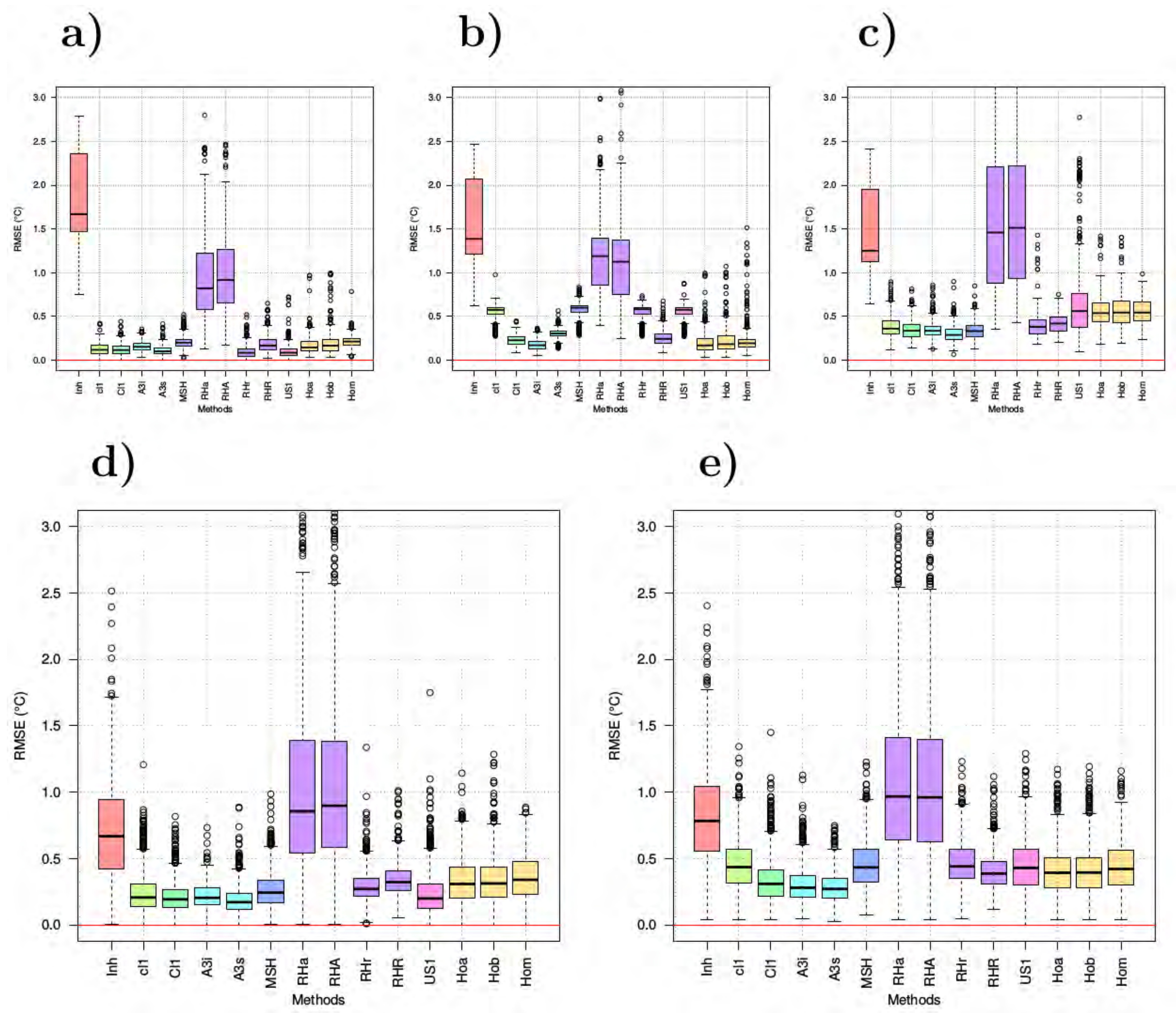


Fig. 3. RMSE of the solutions of the tested software packages for the five types of experiments displayed in Figure 2 using the master network with an intermediate level of cross-correlations (Tm2). Each box of the upper row contains results from 2000 series, and potentially up to 4000 series in the lower row, but in this case an undetermined number of homogeneous series have been excluded from the evaluation. (RMSE axis has been set constant from 0 to 3, so some outliers may lay out the figures).

The right column of *Figure 4* (d, e, f) show the box-plots of the trends errors, whose dispersions are lower than the original and unbiased, except for the absolute homogenization and for the HOMER results. These have also a lower trend dispersion than the problem series, but are negatively biased in all three temperature networks. The lower dispersion of trend errors is achieved by ACMANT, followed by USHCN, MASH and Climatol in the best correlated network (4d). ACMANT is still showing the best results with moderate correlations (4e), but in the worse correlated networks (4f) the dispersions are very similar between the methods. RHtestsV4 show dispersions very similar irrespectively of the correlation degree, while USHCN is the less robust, changing from one of the lower dispersions in 4d to one of the worse in 4f.

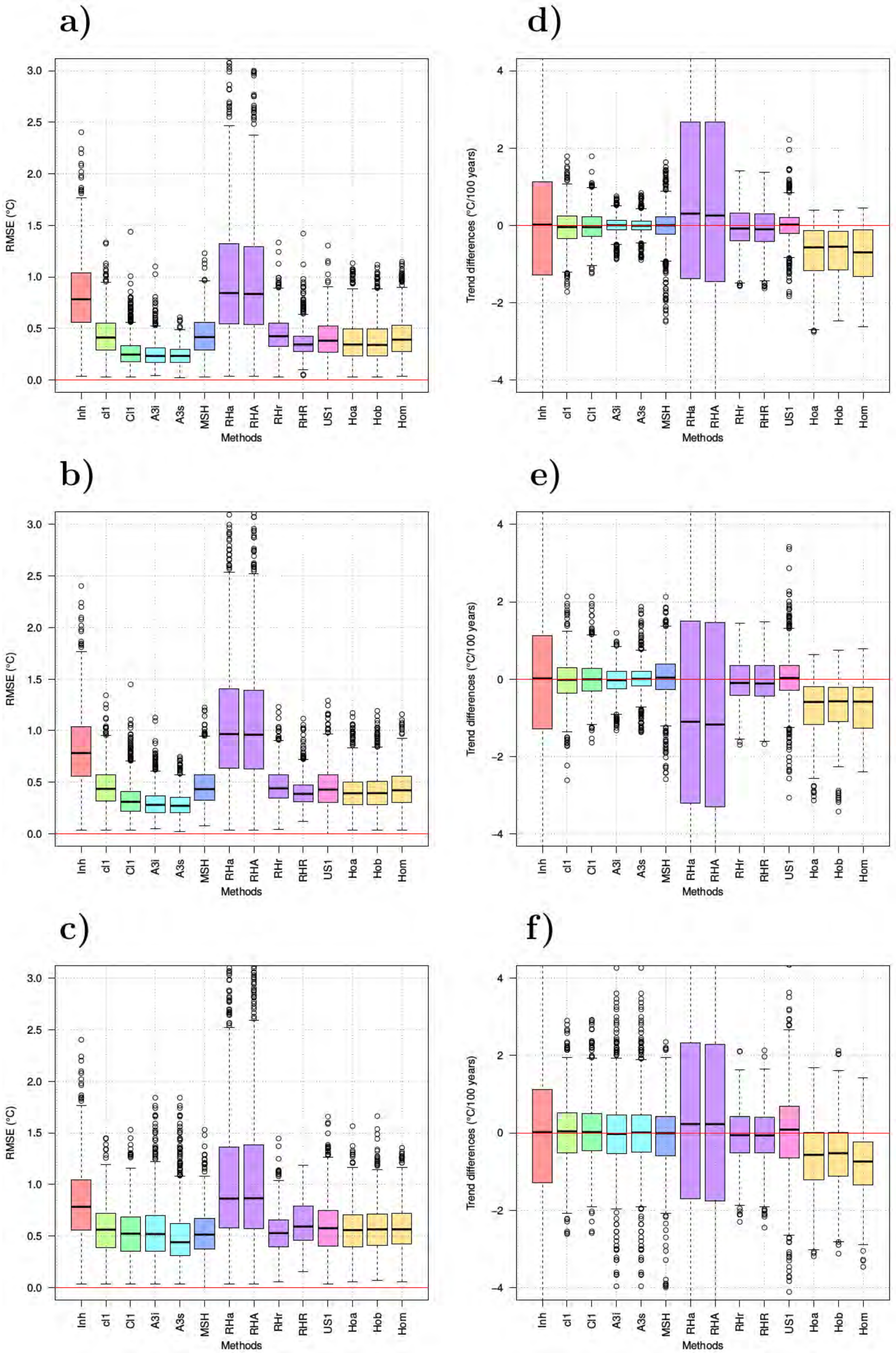


Fig. 4. The left column show the RMSE of the solutions provided by the methods for experiment v (the most realistic) in networks with decreasing level of correlations, a (Tm1), b (Tm2) and c (Tm3). Also for these three master networks, trend errors are shown in the right column (d, e, f).

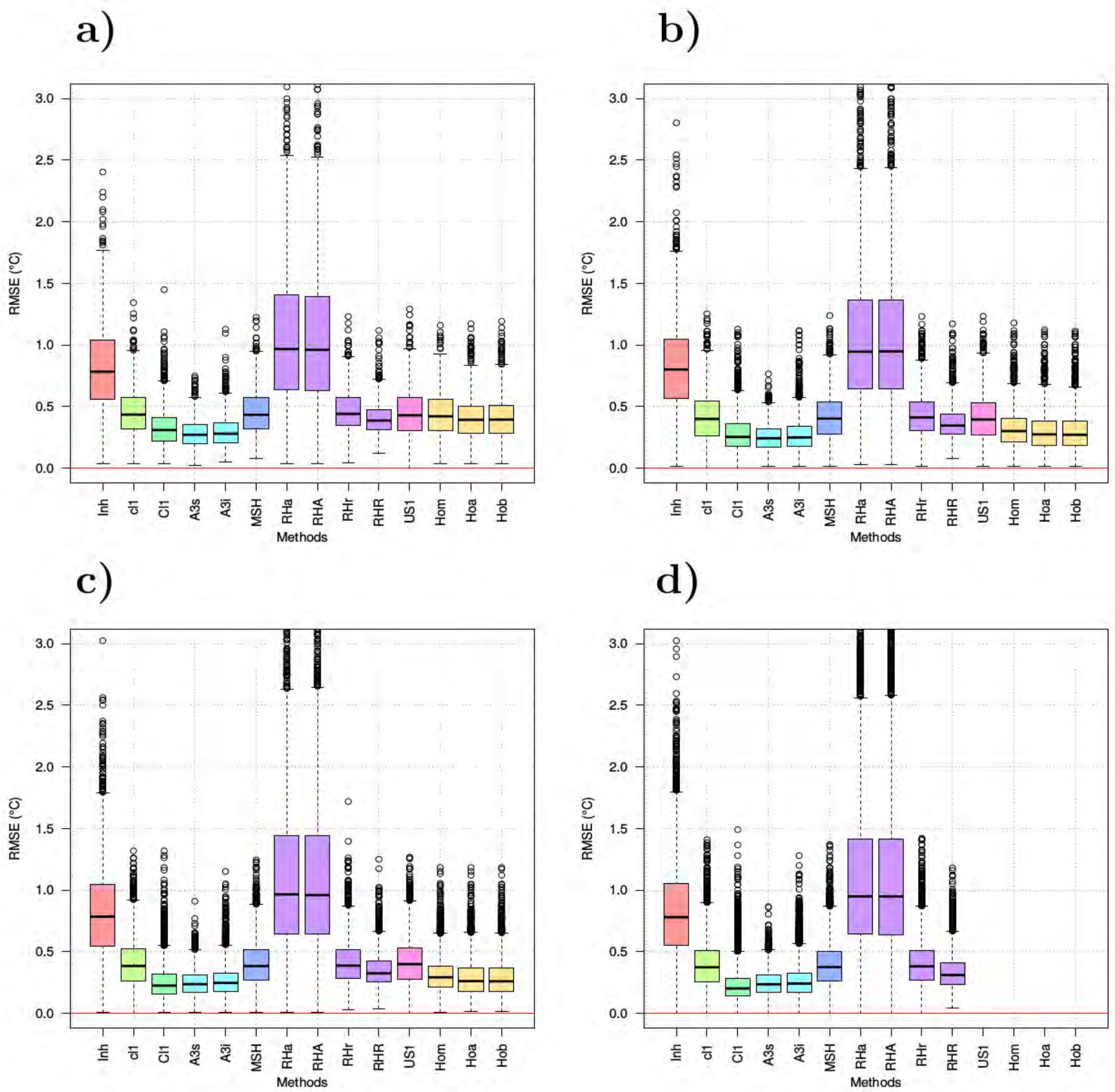


Fig. 5. RMSE of the methods increasing the number of series in each problem sample from 10 (a) to 20 (b), 40 (c) and 80 (c). (In this latter case we could not test USHCN because we had it compiled with a limit of 40 series, and HOMER aborted with errors and produced no results.)

The effect of increasing the number of series in the problem was tested on the master network with intermediate cross-correlations (Tm2) with the more realistic settings (v: random number of shifts of random amplitude with random seasonality) by increasing the sample size from 10 to 20, 40 and 80 series, although in this latter case no results could be obtained from HOMER due to unresolved errors, and USHCN could not be tested because our compilation was done with a limit of 40 series and could not be repeated because the only computer that allowed a successful compilation is no longer available. The box-plots of the RMSE obtained with this increasing number of series can be seen in *Figure 5*.

Results seem to improve when more series are available, especially in HOMER (although limited to the 40 series sample size). The lower RMSE are achieved in all cases by both versions of ACMANT and by Climatol with variable correction (the default); in this order when samples contain 10 or 20 series, and in the reverse order when 40 or 80 series are involved. The errors in the trends of the series also improve with an increased sample size (*Figure 6*), and in the case of HOMER, the worrying bias found in the 10 series samples vanishes as the number of series increases to 20 and 40.

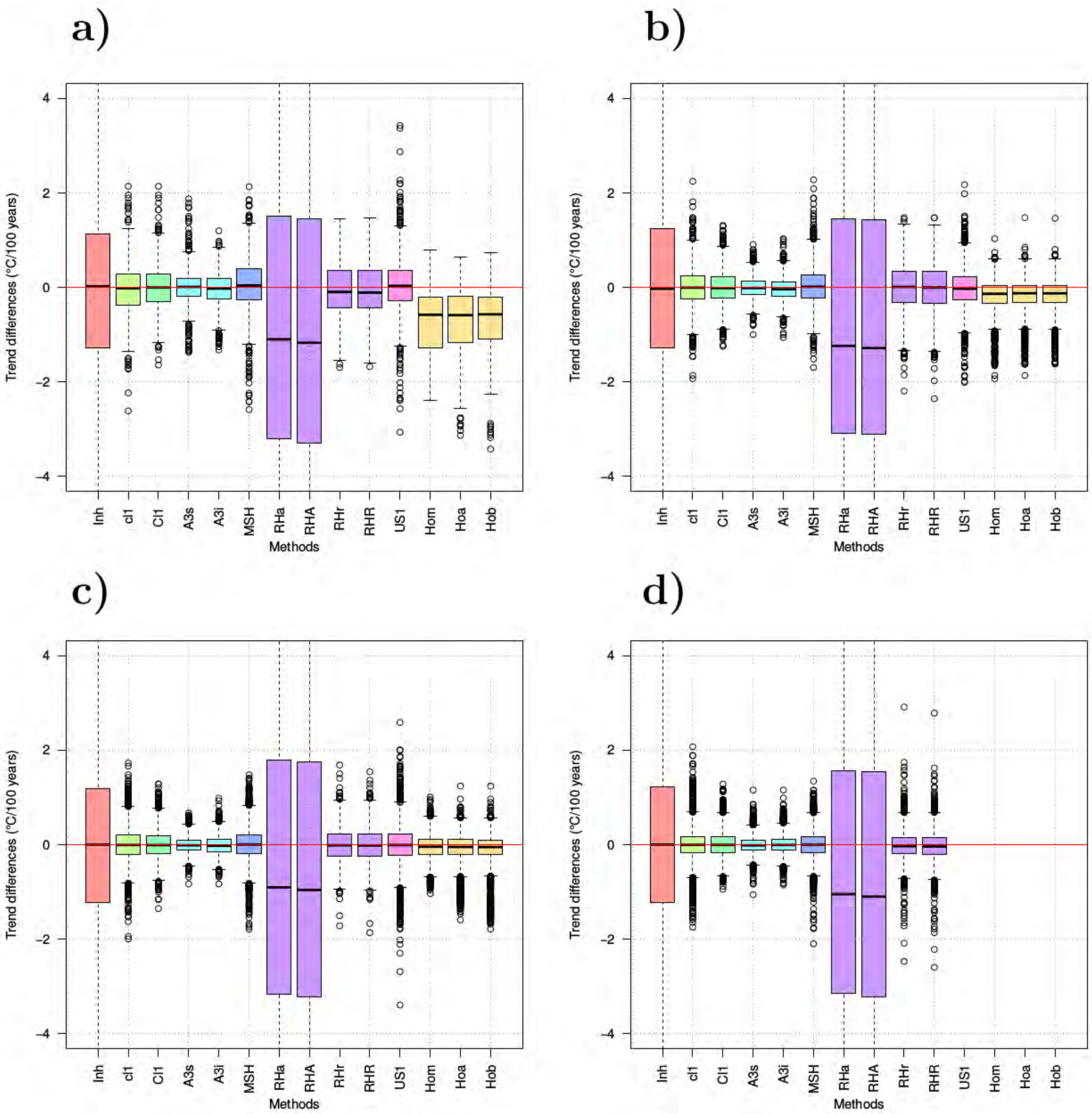


Fig. 6. Trend errors of the methods obtained with an increasing number of series in each problem sample from 10 (a) to 20 (b), 40 (c) and 80 (d). (USHCN and HOMER could not be run in the latter case.)

Another aspect worth to be tested is the performance of the methods when there is a simultaneous shift in a many of the series. It is expected that relative homogenization methods will fail to detect such inhomogeneities when they affect most or all of the series, since those changes will be attributed to real climate variability. Therefore, an additional very simple experiment was performed by introducing a simultaneous shift of 2°C in the middle of the series in 4, 7 and all 10 series of the problem samples. *Figure 7* shows the RMSE and trend errors when 40 and 70% of the series are affected (HOMER did not return any results due to unresolved errors). In the first case (upper row of the figure) all tested methods improved the problem series, although RHtestsV4 results are much worse than the others (*Figure 7a*). The higher reduction in RMSE correspond to USHCN, probably due to its pairwise detection strategy, followed closely by ACMANT. These methods also gave the best unbiased trend corrections (*Figure 7b*), although some outliers appear in the ACMANT trends. On the contrary, *Figures 7c* and *7d* show that when a majority of the series suffer a simultaneous shift (70% in this case), only RHtestsV4 makes a significant reduction on RMSE and, to a lesser extent, in trend errors. (In this case the absolute homogenization over-corrects the trends and introduces an important negative bias.) When all series in the samples present this big simultaneous shift, results (not shown) are similar to those with 70% affected, except that RHtestsV4 relative is also unable to reduce the errors (as expected), and in this case quantile adjustments introduce huge errors. The big magnitude of the introduced simultaneous shifts has allowed absolute homogenization to give relative good results, but real simultaneous changes in the conditions of observation are not expected to produce such big shifts.

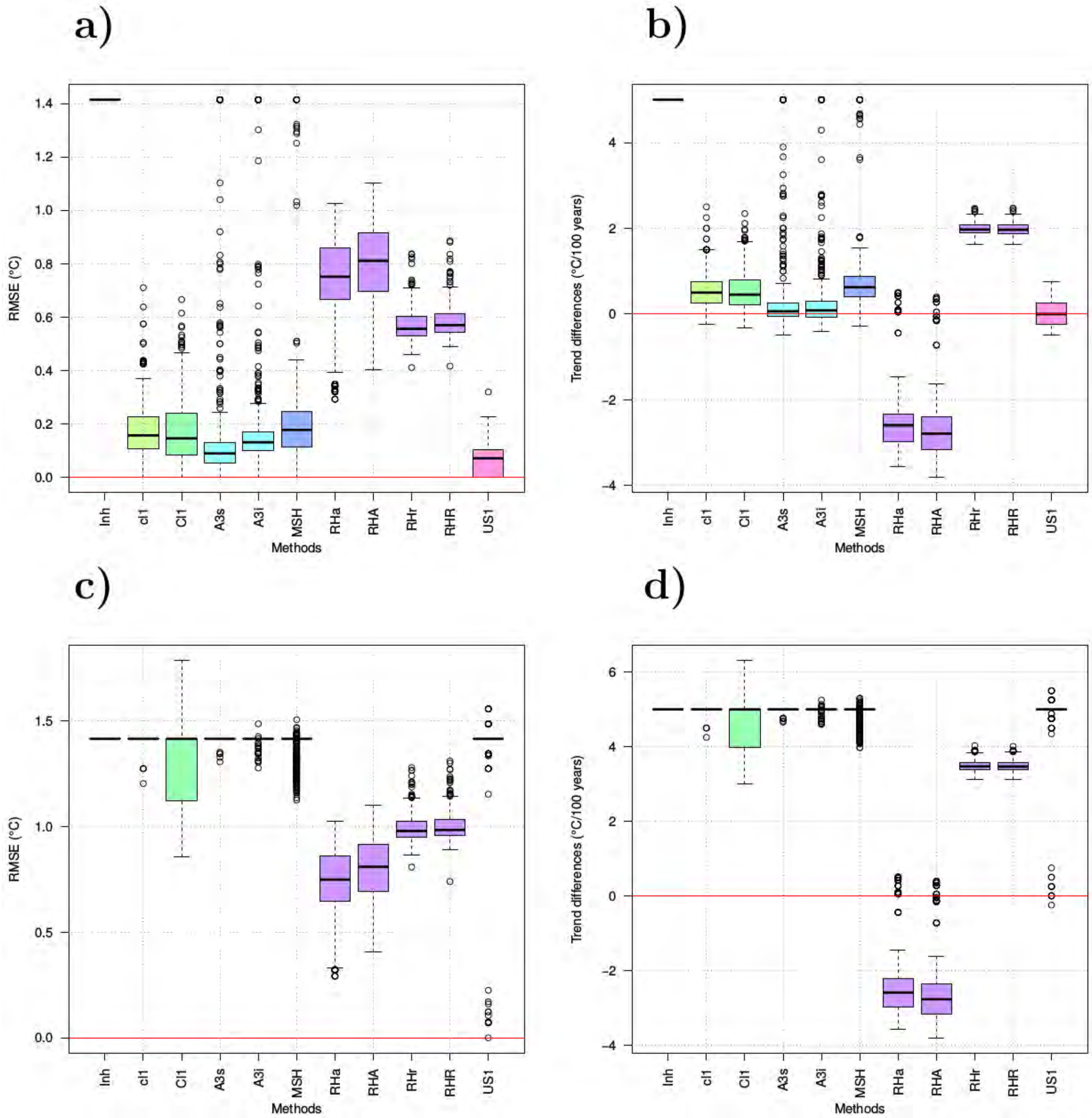


Fig. 7. Boxplots of RMSE (a and c) and trend errors (b and d) when a simultaneous shift of 2°C is introduced in the middle of the series in 40% (upper row) and 70% (lower row) of them. (HOMER gave errors in this experiment and did not produce any results).

3.2. Precipitations

Precipitation tests were performed only with random N(1, 0.2) factors affecting a random number of series at random locations (excluding the last five years of the series) on 10 series samples drawn from the three master networks PEir (Atlantic temperate), PMca (Mediterranean) and PInd (monsoonal). *Figure 8* displays box-plots of RMSE (left column) and trend errors (right column) showing that the homogenization of precipitation series is much more problematic than those of temperature, which can be attributed to their much higher variability, both in time and space, the presence of zeros and its biased probability distribution. There is a general reduction of RMSE in the Atlantic temperate precipitation case (*Figure 8a*), but much less in the Mediterranean (*8b*) and monsoonal (*8c*) climates, in which some methods even gave greater errors than the problem series (especially RHtestsV4 relative with quantile adjustment).

Without quantile adjustment, RHtestsV4 relative performed quite well in the Atlantic and Mediterranean climates, with error reductions similar to HOMER and Climatol, and not too far from ACMANP (the precipitation version of ACMANT), which gave the best results in all three climates. MASH results improve in the most difficult monsoonal climate (*Figure 7c*), where it performs as well as HOMER.

Trend errors (*Figures 8d, e, f*) are also reduced after the homogenization, except for the absolute mode. Biases are small in general, although the greatest (positive) deviations appear in the "easier" Atlantic precipitations with RHtestsV4 and HOMER. This latter method produced negative biases in all three Tm1, Tm2 and Tm3 temperature master networks (*Figures 4d, e, f*), but in this case biases are positive, less noticeable in the Mediterranean precipitation and inexistent in the monsoonal climate case.

Climatol was tested with three different settings: ratio normalization (cl1) and full standardization of cubic root (Cr1) or log(x+1) (Cl1) transformed data. These two last methods produced much worse results that the simpler first setting, probably due to an amplification of errors when undoing the transformation of the data. Therefore, it seems better to disregard the use of the transformation options included in this package, although many other packages use transformations when dealing with multiplicative model variables.

The poor performance found in these precipitation tests does not dismiss applying homogenization procedures to precipitation series, since the best methods did improve the problem series, and shifts greater than those drawn form a N(1, 0.2) may appear in real cases.

4. SUMMARY AND FUTURE WORK

Thousands of tests have been performed by applying six of the most used homogenization computer packages to hundreds of sample networks of monthly temperatures and precipitations with different characteristics and affected by a variety of prescribed inhomogeneities. Results have allowed a comparison of the performance of the tested methods under different circumstances, although only an automated procedure could be applied. Therefore, the packages were run with default parameterizations, and better results could perhaps be obtained by tuning them to every kind of network. Moreover, some methods were not devised to be operated automatically, but were included in this intercomparison project because of their high number of users. That was the case with RHtestsV4, which homogenizes the series one by one and it is the user who must provide a suitable homogeneous reference. Also HOMER is intended to be applied by experienced users who must take subjective decisions.

The simple unrealistic experiments help in detecting the strengths and weaknesses of the methods, while more realistic problem networks give results more useful to evaluate what can be expected when applied to real problems. The ranking of the methods is relatively consistent, but changes appear between different networks and types of inhomogeneities, hence the importance of showing results from networks that can be representative of different real climates and varying station densities. Overall results indicate that ACMANT and Climatol produced the most reliable results in this study, with MASH and USHCN following with also remarkable efficiency metrics, but user preferences may be driven by other characteristics of the software packages, as presented in comparative tables at http://www.climatol.eu/tt-hom/index.html. Moreover, as no significance testing of the differences between methods has been done, the rankings are informative and do not account for the uncertainty associated to the benchmarking process itself.

Automatic testing of homogenization computer packages is the only feasible way of updating the evaluation of the performance of new methods or new versions of existing packages, as far as they can be run in unattended mode. These intercomparisons will be valuable not only for users, but also to the developers of the tested packages, who can see how their algorithms behave under varied climate conditions.

Work under progress include tests with seasonalities other than sinusoidal, and with shifts concentrated over a short period for a high proportion of series. Also the unexpected behavior of RHtestsV4 and HOMER trend biases in some experiments must be investigated to understand their causes and discard possible flaws in the testing scripts.

Finally, tests will be performed on a longer and more realistic benchmark, with varying number of missing data along time, similar to that used in the COST Action ES0601, and biased inhomogeneities producing abnormal trends in the networks will also be considered.

At the end of the project, all results and scripts will be accessible in a web page to allow reproducibility. This benchmarking infrastructure will remain operative after the end of the project, allowing future updates of intercomparison exercises, both of the currently tested packages and any other willing to be added. The most straightforward way to widen the number of tested methods would be the involvement of the developers by providing suitable automatic scripts to apply their software to the problem networks. This way would not only facilitate the time consuming task of studying their way of operation, input and output file formats and other requirements, but their operation procedures would be directly designed by the developers, minimizing the possibilities of applying them incorrectly.

Acknowledgments

Project MULTITEST (CGL2014-52901-P) is funded by the Spanish Ministry of Economy and Competitiveness. Manola Brunet is also supported by the European Union-funded project "Uncertainties in Ensembles of Regional Reanalyses" (UERRA, FP7-SPACE-2013-1 project number 607193). Victor Venema is also supported by the DFG project Daily HUME (VE 366 - 8).

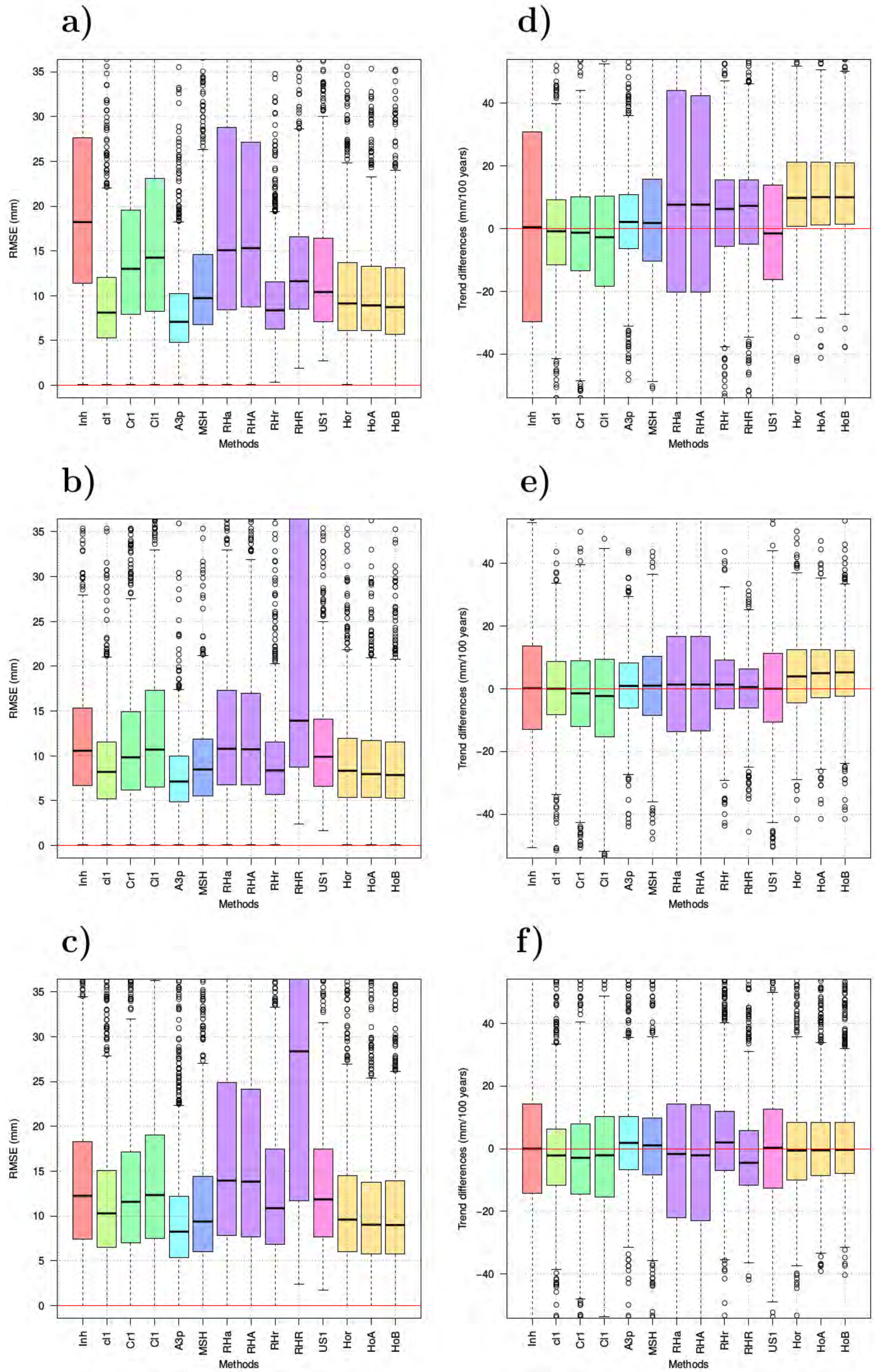


Fig. 8. RMSE (a, b, c) and trend errors (d, e, f) of the methods applied to Atlantic temperate (upper row), Mediterranean (middle row) and monsoonal (bottom row) simulated precipitation series.

Our thanks go to Met Éireann for for providing the Irish monthly precipitation series that served as model to synthesize the network of Atlantic Temperate precipitations. Mallorca monthly precipitations were taken from AEMET data bases, and monthly precipitations from SW India, gridded at 0.5° resolution, were obtained from the Global Precipitation Climate Center (GPCC).

References

Domonkos P (2015): Homogenization of precipitation time series with ACMANT. *Theor. Appl. Climatol*., 122:303-314.

Guijarro JA (2016): Package 'climatol'. https://cran.r-project.org/web/packages/climatol/climatol.pdf (Accessed in June 2017).

López JA, Guijarro JA, Aguilar E, Domonkos P, Brunet M (2016): Una propuesta metodológica para la generación de redes de precipitación simuladas a partir de redes de precipitación observadas en el marco del proyecto MULTITEST. In Olcina J, Rico AM, Moltó E (eds.): "Clima, sociedad, riesgos y ordenación del territorio", Universidad de Alicante (Spain), *Asociación Española de Climatología*, ISBN 978-84-16724-19-2, pp. 183-194.

Menne MJ, Williams CN Jr (2005): Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate* 18:4271-4286.

Mestre O, Domonkos P, Picard F, Auer I, Robin S, Lebarbier E, Böhm R, Aguilar E, Guijarro J, Vertachnik G, Klancar M, Dubuisson B, Stepanek P (2013): HOMER: a homogenization software - methods and applications. *Időjárás*, 117:47-67.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (Accessed in June 2017).

Schneider U; Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Ziese M (2015): GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD_GPCC/FD_M_V7_050.

Szentimrey T (2007): Manual of homogenization software MASHv3.02. Hungarian Meteorological Service, 65 pp.

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.

Wang XL, Feng Y (2013): RHtestsV4 User Manual. http://etccdi.pacificclimate.org/software.shtml, 29 pp. (Accessed in June 2017).

SOME THEORETICAL QUESTIONS AND DEVELOPMENT OF MASH FOR HOMOGENIZATION OF STANDARD DEVIATION

Tamás Szentimrey

Hungarian Meteorological Service
szentimrey.t@met.hu

Abstract

There are several methods and software for homogenization of climate data series but there is not any exact mathematical theory of the homogenization. Therefore we have to formulate some questions of homogenization in accordance with the mathematical conventions. The basic question is the mathematical definition of the inhomogeneity and the aim of homogenization. It is necessary to clarify that the homogenization of climate data series is a distribution problem instead of a regression one. Another problem is the relation of monthly and daily data series homogenization.

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. In the practice the monthly series are homogenized in the mean only, while there exist some trials to homogenize the daily series also in higher order moments. These procedures are based on a popular assumption that is the correction of mean is sufficient for monthly series, and 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 inhomogeneity in the higher order moments. However, this assumption is false, since it can be proved if there is a common inhomogeneity in the standard deviation of daily data then we have the same inhomogeneity in monthly data. Therefore we develop a mathematical procedure for the homogenization of mean and standard deviation together. Building of this procedure in our software MASH (Multiple Analysis of Series for Homogenization; Szentimrey) is ongoing and it is based on the examination of monthly series and the monthly results are applied for the homogenization of daily series. We remark if the data are normally distributed then the homogenization of mean and standard deviation is sufficient since in case of normal distribution if the first two moments are homogenous then the higher order moments are also homogeneous.

1. INTRODUCTION

In our conception the meteorological questions and topics cannot be treated separately. Therefore we present a block diagram (*Figure 1*) to illustrate the possible connection between various important meteorological topics. The software MASH (Multiple Analysis of Series for Homogenization; *Szentimrey*, 1999, 2014) and MISH (Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey* and *Bihari*, 2014) were developed by us. These software were applied also in CARPATCLIM project (http://www.carpatclim-eu.org).

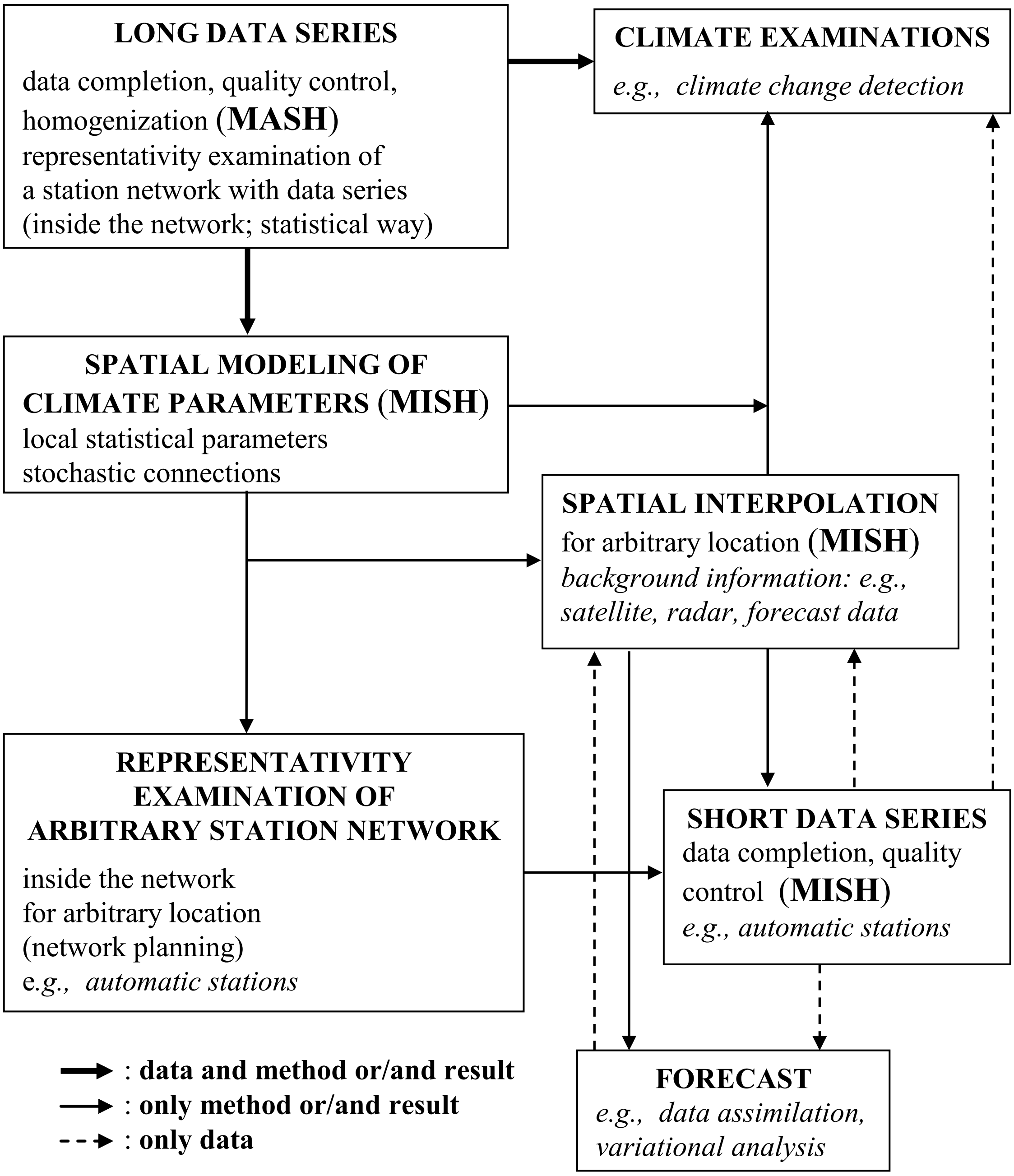


Fig. 1. Block diagram for the possible connection between various basic meteorological topics and systems

2. MATHEMATICAL FORMULATION OF CLIMATE DATA HOMOGENIZATION

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

2.1 General mathematical formulation

Notation

Let us assume we have daily or monthly climate 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.

The appropriate theoretical cumulative distribution (CDF) functions are:

$$F_{1,t}(y)=\mathrm{P}(Y_1(t)<y)\ ,\quad F_{2,t}(y)=\mathrm{P}(Y_2(t)<y)\quad y\in(-\infty,\infty)\,,\ t=1,2,..,n$$

It is very important to remark that as a consequence of some natural changes - e.g. annual cycle, climate change - the series of distribution functions $F_{1,t}(y)$, $F_{2,t}(y)$ $(t=1,2,..,n)$ may change in time! In the statistical climatology the climate change is equivalent with the changing probability of the meteorological events. The inhomogeneity of data series can be defined on the basis of the distribution functions.

Definition 1

The merged series $Y_2(t)\ (t=1,2,..,T)$, $Y_1(t)\ (t=T+1,..,n)$ is inhomogeneous, if the identity of the distribution functions $F_{2,t}(y)\equiv F_{1,t}(y)\ (t=1,2,..,T)$ is not true.

Definition 2

The aim of the 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 $Y_1(t)\ (t=1,2,..,T)$ have, i.e.:

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

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

Remark 1

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.

Theorem 1

Let us assume about the random variables Y_1, Y_2 and their distribution functions $F_1(y), F_2(y)$, that $\mathrm{P}(Y_j \in (a_j, b_j)) = 1$ and $F_j(y)$ is a strictly increasing continuous function on the interval (a_j, b_j) $(j = 1,2)$. Then applying the transfer function $Y_{1,2h} = F_1^{-1}(F_2(Y_2))$ we obtain that the variable $Y_{1,2h}$ has the same distribution like Y_1 i.e. $\mathrm{P}(Y_{1,2h} < y) = \mathrm{P}(Y_1 < y) = F_1(y)$.

Definition 3

Transfer function: $F_{1,t}^{-1}(F_{2,t}(y))$ and quantile function: $F_{1,t}^{-1}(p)$.

Theoretical formulation of homogenization of $Y_2(t)$ $(t = 1,2,..,T)$:

$Y_{1,2h}(t) = F_{1,t}^{-1}(F_{2,t}(Y_2(t)))$, then $\mathrm{P}(Y_{1,2h}(t) < y) = F_{1,t}(y)$.

Remark 2

The basis of the Quantile Matching methods can be integrated into the general theory. However these methods developed in practice mainly for daily data are very weak empiric methods. It is not real mathematics! These methods have good heuristics with poor mathematics.

2.2 Arising mathematical questions to be solved

Let us suppose the merged series is given that is,

$Y_2(t)$ $(t = 1,2,..,T)$, $Y_1(t)$ $(t = T+1,..,n)$

In addition we suppose that the assumptions of the former theorem are fulfilled, consequently the theoretical correction or transfer formulas for the series elements are,

$$Y_{1,2h}(t) = F_{1,t}^{-1}(F_{2,t}(Y_2(t))) \quad (t = 1,2,..,T) \tag{2}$$

However these transfer formulas are theoretical ones and if we want to apply them in the practice then a number of mathematical statistical estimation problems are arising. The most important problems are as follows.

- Estimation, detection of the change point(s) *T*.
- Estimation of the theoretical distribution functions $F_{1,t}(y)$, $F_{2,t}(y)$ $(t = 1,2,..,T)$:
  - i, $F_{1,t}(y)$, $F_{2,t}(y)$ may change in time because of the climate change and the annual cycle, consequently the methodology of the use of the empirical distribution functions is very doubtful.
  - ii, There is no sample for $F_{1,t}(y)$ $(t = 1,2,..,T)$ and $F_{2,t}(y)$ $(t = T+1,..,n)$ usually.

These mathematical problems are insolvable generally! Therefore only relative methods can be used with some model assumptions. In addition some simplifications are also necessary. Statistically speaking, some assumptions have to be made!

2. 3 Mathematical formulation for normal distribution

The homogenization problem is very complicated in general case however in case of normal distribution a much simpler mathematical formula can be obtained. We emphasize that the normal distribution is a special case but it is basic one in the mathematical statistics as well as in the meteorology. For example the normal distribution model can be accepted for the temperature variables in general.

Theorem 2

Let us assume the data series have normal distribution that is,
$Y_1(t) \in N(E_1(t), D_1(t))$, $\quad Y_2(t) \in N(E_2(t), D_2(t)) \quad (t = 1,2,..,n)$,

where $\mathrm{E}(Y_1(t)) = E_1(t)$, $\mathrm{E}(Y_2(t)) = E_2(t)$ are the means or expected values and $\mathrm{D}(Y_1(t)) = D_1(t)$, $\mathrm{D}(Y_2(t)) = D_2(t)$ are the standard deviations.
Then the transfer formula of homogenization:

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

2.4 Mathematical questions in case of normal distribution to be solved

In case of normal distribution according to the *Theorem 2* we have a much simpler transfer formula for correction than the general form (2), that is,

$$Y_{1,2h}(t) = E_1(t) + \frac{D_1(t)}{D_2(t)}(Y_2(t) - E_2(t)) \quad (t = 1,2,..,T) \tag{3}$$

This formula is a simple linear one that means if the data series have normal distribution it is sufficient to homogenize the means and standard deviations only that is equivalent with the homogenization of the first two moments. We emphasize that the normal distribution is a basic model in the mathematical statistics as well as in the meteorology and there is no “tail distribution” problem at this important distribution according to the *Theorem 2*! At the normal distribution if the means and standard deviations are homogenous then the higher order moments are also homogeneous and there is not any inhomogeneity in the tails of the distributions. It is in contrast with the popular assumption based on parallel measurements as it is very likely the inhomogeneity in the tails of the distributions at the daily data series. As regards the parallel measurements a mathematical examination for them will be presented at Section 2.5.

Returning to the formula (3) although it is much simpler than (2), there are still a number of mathematical statistical estimation problems to be solved as follows.

- Estimation, detection of the change point(s) *T*.
- Estimation of the statistical parameters $E_1(t), D_1(t), E_2(t), D_2(t) \quad (t = 1,2,..,T)$:
  - i, $E_1(t), D_1(t), E_2(t), D_2(t)$ may change in time because of the climate change and the annual cycle.
  - ii, There is no sample for $E_1(t), D_1(t) \ (t = 1,..,T)$ and $E_2(t), D_2(t) \ (t = T+1,..,n)$ usually.

However these mathematical problems are still very complicated! Therefore only relative methods can be used with some model assumptions. In addition some simplifications are also necessary.

The most often applied transfer formula in the practice can be obtained from the formula (3) with the following simplifications,

a, $D_2(t)/D_1(t) = D_{21}, \quad E_2(t) - E_1(t) = E_{21} \quad (t = 1,2,..,T)$

b, $D_{21} = 1 \;, \quad E_2(t) - E_1(t) = E_{21} \quad (t = 1,2,..,T)$

Then the transfer formula is,

$Y_{1,2h}(t) = Y_2(t) - E_{21} \qquad (t = 1,2,..,T)$

This is the homogenization in mean applied in the practice mostly (Section 3).

2.5 Mathematical examinations of parallel measurements

On the one hand the daily data series are very important for studying extremes. On the other hand there is a popular assumption based on parallel measurements and some physical considerations as it is very likely the inhomogeneity in the tails of the distributions at the daily data series. What is the reason of this assumption?

Essentially the reason is 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 can be found in the paper (*Szentimrey et al.*, 2015).

3. RELATION OF DAILY AND MONTHLY HOMOGENIZATION

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.

3.1 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?

3.2 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 (E)).
   Assumption: homogeneity in higher order moments (e.g. standard deviation (D)).
2. Homogenization of daily series:
   There is a trial to homogenize also in higher order moments.
   (Quantile Matching, Spline methods)
   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? According to Section 3.3 the correct answer is that this model cannot be accepted.
- Why are not used the monthly correction factors for daily homogenization? It seems to lose some valuable information obtained during the monthly homogenization.

3.3 Problem of inhomogeneity of the standard deviation

There is also a popular assumption applied in the practice 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 a common inhomogeneity in the standard deviation of daily data, we may have the same inhomogeneity in monthly data.

Theorem 3

Let us assume $Y(t)\ (t=1,..,30)$ are daily data and the monthly mean is $\overline{Y}=\frac{1}{30}\sum_{t=1}^{30}Y(t)$.

Monthly variable for examination of standard deviation (D): $S=\sqrt{\frac{1}{29}\sum_{t=2}^{30}\left(Y(t)-Y(t-1)\right)^2}$

Let us introduce some inhomogeneity of the mean (E) and the standard deviation (D) for the daily data by a linear function:

$$Y_{ih}(t)=\alpha\cdot\left(Y(t)-\mathrm{E}(Y(t))\right)+\mathrm{E}(Y(t))+\beta \qquad (t=1,..,30)$$

Then the expected values and the standard deviations are:

$$\mathrm{E}(Y_{ih}(t))=\mathrm{E}(Y(t))+\beta\ ,\quad \mathrm{D}(Y_{ih}(t))=\alpha\cdot\mathrm{D}(Y(t)) \qquad (t=1,..,30)$$

The appropriate monthly variables: $\overline{Y}_{ih}=\frac{1}{30}\sum_{t=1}^{30}Y_{ih}(t)$, $\quad S_{ih}=\sqrt{\frac{1}{29}\sum_{t=2}^{30}\left(Y_{ih}(t)-Y_{ih}(t-1)\right)^2}$

i, Then the monthly mean is also inhomogeneous in mean (E) and standard deviation (D) with the same measure like the daily values:

$$\mathrm{E}(\overline{Y}_{ih})=\mathrm{E}(\overline{Y})+\beta \qquad \text{and} \qquad \mathrm{D}(\overline{Y}_{ih})=\alpha\cdot\mathrm{D}(\overline{Y})\ .$$

ii, Moreover variable S_{ih} can be used to estimate the inhomogeneity of standard deviation (*D*):

$$\mathrm{E}(S_{ih}) = \alpha \cdot \mathrm{E}(S)$$

3.4 An alternative procedure developed in MASH

We suggest an alternative procedure to homogenize both the daily and the monthly series.

The steps of the procedure in case of quasi normal distribution (additive model, e.g. temperature) are as follows.

1. Homogenization of monthly series $S(t)$, $\overline{Y}(t)$.
   Homogenization of series $S(t)$ by multiplicative model: break points detection, estimation of inhomogeneity of standard deviation (*D*).
   Correction of standard deviation of series $\overline{Y}(t)$.
   Homogenization of corrected series $\overline{Y}(t)$ by additive model: break points detection, estimation of the inhomogeneity of mean (*E*).
   Assumption: homogeneity in higher order (>2) moments. This assumption is always right in case of normal distribution according to *Theorem 2.*
2. Homogenization of daily series.
   Homogenization of mean and standard deviation on the basis of the monthly results. The used monthly information are the break points and the monthly corrections of the mean (*E*) and standard deviation (*D*). The correction is based on the transfer formula (3) considering *Theorem 3*. If the daily data are normally distributed then there is no inhomogeneity in the higher order moments according to *Theorem 2.*

3.4 Example for the inhomogeneity in mean (*E*) and standard deviation (*D*)

The presented example is based on the daily maximum temperature series of station Miskolc in Hungary for the period 1901-2015. During the beginning period 1901-1908 the measurement was in Réaumur while the further values were in Celsius. Therefore there is some inhomogeneity in the period 1901-1908 since $\mathrm{Re} = 0.8 \cdot$ºC. Consequently we have inhomogeneity both in mean (*E*) and in standard deviation (*D*):

$$\mathrm{E}(Y_{ih}(t)) = 0.8 \cdot \mathrm{E}(Y(t)),\ \mathrm{D}(Y_{ih}(t)) = 0.8 \cdot \mathrm{D}(Y(t)).$$

In the *Figure 2.* at the annual mean series can be detected this phenomenon.

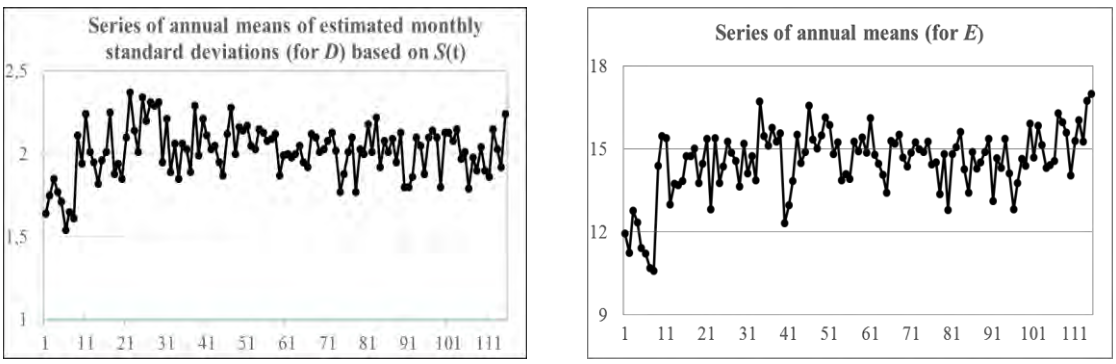


Fig. 2. Annual mean series of monthly series $S(t)$, $\overline{Y}(t)$

5. SOFTWARE MASH

(Multiple Analysis of Series for Homogenization; *Szentimrey* 1999, 2008, 2014)

5.1 General comments

The new version MASHv4.01 is under development. The MASH system is based on the homogenization of monthly series derived from daily series. The procedures depend on the distribution of climate elements.

Quasi normal distribution (e.g. temperature)

Beside the monthly mean series another type monthly series are also derived. These series are homogenized by multiplicative model for standard deviation (D). The monthly mean series corrected in standard deviation are homogenized by additive model for mean (E).

Quasi lognormal distribution (e.g. precipitation)

Monthly mean or sum series are homogenized by multiplicative model.

5.2 The most important properties of MASH

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 or multiplicative model can be used depending on the distribution.
- 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.

Remark

The aim of MASH is not the full automation and we are sceptic in such an aspect. However our intention is to obtain such a flexible automatic system wherein the mechanic, labour-intensive procedures are automated, moreover the operating process can be controlled simply and the accidental mistakes can be corrected easily. The basic idea of this conception is to control the results via the verification tables generated automatically during the automatic procedures.

Some MASH specialty

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

Our MISH-MASH software can be downloaded from:
http://www.met.hu/en/omsz/rendezvenyek/homogenizationand_interpolation/software/

5.3 Some results for homogenization in mean (*E*) and standard deviation (*D*)

15 Hungarian July mean temperature series (1901-2015) were homogenized by MASH in mean (*E*) and standard deviation (*D*). Some verification tables of MASH are presented on the following Figures.

```
Test Statistics for St. Deviation (D) Before Homogenization
Null hypothesis: the series are homogeneous in D.
Critical value (significance level 0.01):  28.00
Test statistics (TSB) can be compared to the critical value.
The larger TSB values are more suspicious!

Series      TSB     Series      TSB     Series      TSB
   7     201.40       8      168.65       13     126.68
   9     123.38       4      121.03       14      94.02
  12      83.32       2       78.07        5      63.54
   6      58.47      11       44.14       15      43.91
  10      32.60       1       25.54        3      17.14

AVERAGE:  85.46


Test Statistics for Mean (E) Before Homogenization
Null hypothesis: the series are homogeneous in E.
Critical value (significance level 0.05):  21.76
Test statistics (TSB) can be compared to the critical value.
The larger TSB values are more suspicious!

Series      TSB     Series      TSB     Series      TSB
  12    1674.66       7      388.59        8     237.88
   3     230.71      10      224.70        5     211.41
   6     188.81      11      154.68       14     125.35
   4      82.50       9       72.61       15      57.57
   1      53.55      13       49.91        2      32.95

AVERAGE: 252.39
```

Fig. 3. Test statistics for mean (*E*) and standard deviation (*D*) before homogenization

```
Test Statistics for St. Deviation (D) After Homogenization
Null hypothesis: the series are homogeneous in D.
Critical value (significance level 0.01):  28.00
Test statistics (TSA) can be compared to the critical value.
The larger TSA values are more suspicious!

Series      TSA     Series      TSA     Series      TSA
  13      36.93      14       32.50        4      32.29
   8      26.56      12       25.73        7      23.87
   9      23.71       5       23.37        2      22.18
   1      19.85       3       19.70       11      18.49
   6      16.55      10       16.55       15      14.82

AVERAGE:  23.54


Test Statistics for Mean (E) After Homogenization
Null hypothesis: the series are homogeneous in E.
Critical value (significance level 0.05):  21.76
Test statistics (TSA) can be compared to the critical value.
The larger TSA values are more suspicious!

Series      TSA     Series      TSA     Series      TSA
   5      25.55       3       23.24       13      21.64
  14      21.19       7       19.43        9      18.53
   6      18.02      15       16.98        8      16.61
  12      16.49      11       16.25        4      15.70
   2      14.29      10       13.28        1      11.69

AVERAGE:  17.93
```

Fig. 4. Test statistics for mean (*E*) and standard deviation (*D*) after homogenization

The estimated inhomogeneities can be characterized by the following statistics.

i, For standard deviation (multiplicative model):

$$IHD = \frac{100}{n}\sum_{t=1}^{n}\left|IHD(t)-1\right| \quad \text{where} \quad D_{ih}(t) = D(t)\cdot IHD(t) \quad (t = 1,...,n)$$

and $D_{ih}(t), D(t)$ are the standard deviations before and after homogenization.

ii, For mean (additive model):

$$IHE = \frac{1}{n}\sum_{t=1}^{n}\left|IHE(t)\right| \quad \text{where} \quad E_{ih}(t) = E(t) + IHE(t) \quad (t = 1,...,n)$$

and $E_{ih}(t), E(t)$ are the means before and after homogenization.

Estimated Inhomogeneities for St. Deviation (D)(%)

Series	IHD	Series	IHD	Series	IHD
8	8.05	9	7.98	4	6.73
12	4.88	7	4.08	11	3.59
6	3.33	2	2.43	15	2.22
5	2.16	13	2.02	10	1.70
1	1.57	14	1.34	3	0.54

AVERAGE: 3.51

Estimated Inhomogeneities for Mean (E) (°C)

Series	IHE	Series	IHE	Series	IHE
3	0.80	8	0.55	15	0.53
7	0.52	12	0.48	10	0.48
14	0.31	6	0.31	5	0.29
11	0.24	1	0.23	4	0.14
9	0.13	2	0.09	13	0.08

AVERAGE: 0.35

Fig. 5. Characterization of inhomogeneities for mean (*E*) and standard deviation (*D*)

References

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

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. and CARPATCLIM Homogenization-Interpolation Team, 2012: Final report on quality control and data homogenization measures applied per country, including QC protocols and measures to determine the achieved increase in data quality. *Deliverable D1.12 of CARPATCLIM*, homepage: http://www.carpatclim-eu.org/pages/deliverables/

Szentimrey, T. 2013: Theoretical questions of daily data homogenization, *Időjárás* Vol. 117. No. 1, January-March 2013. pp. 113-122.

Szentimrey,T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60.

Szentimrey,T.,2014: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p.69.

Szentimrey T., Lakatos M., Bihari Z., 2015: Mathematical questions of homogenization and quality control, Proceedings of the 8th Seminar for Homogenization and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation Techniques in Climatology and Meteorology, Budapest, Hungary, 2014, WCDMP-No. 84, pp. 5-22

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

HOMOGENISATION OF DAILY STATION DATA IN ENGLAND AND WALES

Kay Shelton[1*], Sarah Warren[1], Richard Davis[2], Duncan Faulkner[1]

[1]JBA Consulting, Skipton, UK
[2]Environment Agency, Reading, UK
*(Kay.Shelton@JBAConsulting.com)

Abstract

The Environment Agency (EA) in England is planning on developing a gridded dataset of historical potential evapotranspiration (PET) across England and Wales for the period 1961 onwards, based on observed climate data. The dataset will be used to validate and calibrate existing surface exchange schemes that provide inputs to operational hydrological rainfall-runoff models and numerical groundwater models.

The calculation of PET, using the Penman-Monteith equation, requires temperature, humidity, radiation and wind speed information. Spatial and temporal discontinuities have been identified in PET calculated using standard methodologies within the UK. For example, inhomogeneities have been identified in the wind speed observations currently used within these calculations.

The first step in deriving the gridded PET dataset is the generation of homogeneous time series of relevant meteorological data for each available station across England and Wales. To ensure robustness and reproducibility of the final homogenised dataset, the fully-automatic homogenisation package Multiple Analysis of Series for Homogenization (MASH) is used. The use of an automatic homogenisation package reduces the subjectivity in the homogenisation process and reduces the possibility for the introduction of errors due to inexperienced data homogenisers. MASH also allows daily data, for both additive (e.g. temperature) and cumulative (e.g. windspeed or precipitation) meteorological variables, to be homogenised using a single consistent process.

The following meteorological station data are homogenised: 0900 UTC 2m air temperature, 0900 UTC 2m dewpoint temperature, 24-hour maximum 2m air temperature, 24-hour minimum 2m air temperature, 24-hour average 10m wind speed, and 24-hour sunshine duration. Complete calendar year data are available for the period 1961 to 2013. For each variable, network coverage varies greatly throughout this period, with a mixture of record lengths across the network for each variable. Due to the frequent presence of missing data, and the existence of numerous sites with short records, the networks selected for homogenisation are limited to stations with record lengths of at least 15 years (or 20 years for some variables). This yields network sizes of 300-400 stations for each of the temperature variables, 250 for sunshine and 150 for wind speed.

Results from the homogenisation of completed variables will be presented using the verification scores determined within the MASH process, along with some of the challenges faced and lessons learned during the experience.

Questions will also be posed regarding how best to proceed with integrating future updates of meteorological data into the homogenised time series, and how to handle the homogenisation of shorter record sites.

1. INTRODUCTION

The Environment Agency (EA) in England is developing a gridded dataset of historical potential evapotranspiration (PET) across England and Wales for the period 1961 onwards, based on observed climate station data. The dataset will be used to validate and calibrate existing surface exchange schemes that provide inputs to operational hydrological models.

The calculation of PET, using the FAO56 Penman-Monteith method (*Allen et al.,* 1998), requires temperature, humidity, radiation and wind speed information. Spatial and temporal discontinuities have been identified in existing inputs into PET datasets calculated using standard methodologies within the UK (*Sharpe*, 2012). For example, inhomogeneities have been identified in wind speed observations currently used within these calculations. The derivation of a new PET dataset aims to minimise spatial and temporal inhomogeneities in the final gridded PET dataset as much as possible. Temporal inhomogeneities in the station data are minimised by generating homogeneous time series of relevant meteorological data. From these homogeneous station time series, gridded datasets will be derived using spatial interpolation to a regular grid, from which the resulting gridded PET dataset will be calculated.

This paper focusses on the first stage of deriving the new gridded PET dataset, namely the homogenisation of meteorological data for each available station across England and Wales

2. DATA AND METHODOLOGY

As noted in Section 0, the gridded PET dataset under development covers England and Wales, as shown in *Fig. 1.* This domain includes surface elevations up to 1,085 m, the peak of Snowdon in northwest Wales, with several areas of high elevation spread across the domain. Most notably, elevations over 300 m are common in areas of the Pennines, oriented approximately north-south in the north of England, the Lake District in the northwest of England, the North Yorkshire Moors in northeast England, Exmoor and Dartmoor in southwest England and most of inland Wales.

To ensure an accurate representation of PET around the domain, the climate stations used in this study must reflect the orographic features identified above and the underlying topography of the domain by including adequate coverage by both inland and coastal stations.

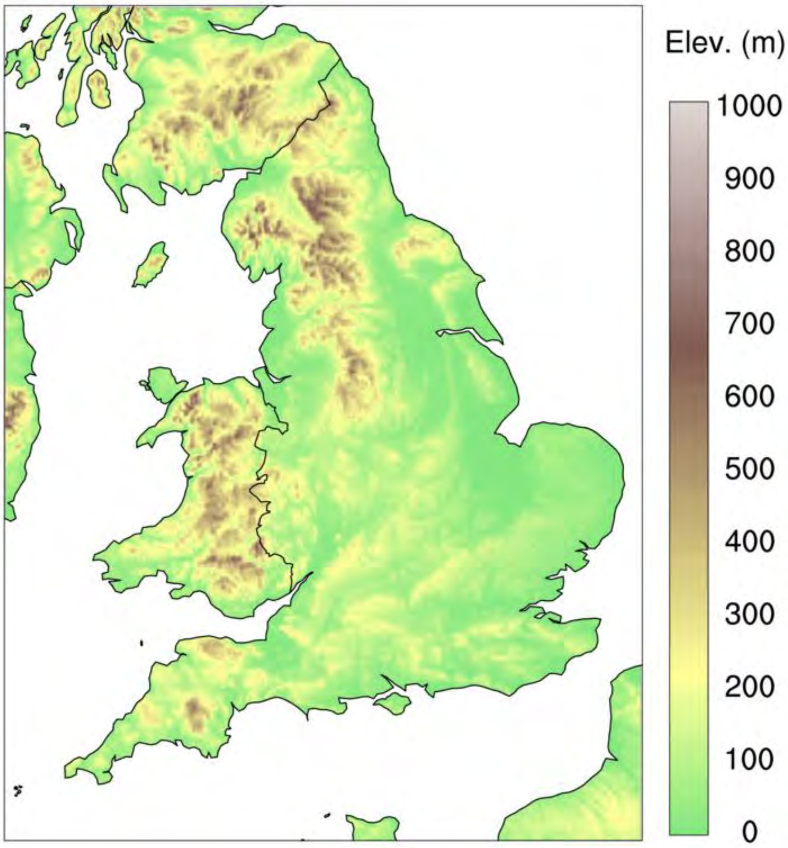


Fig. 1. Topography of England and Wales based on the NOAA ETOPO1 dataset (Amante and Eakins 2009).

2.1 Data

Daily synoptic and climate station data, archived by the UK Meteorological Office (UKMO), for the period 1961-2013 are used in this study, however, complete calendar year data are only available for the period 1961 to 2012. The meteorological station data, and associated station network sizes, used in this study are summarised in *Table 1*. For each meteorological variable, network size and coverage varies greatly throughout the 52-year period, with a mixture of station record lengths across the network for each variable.

The UKMO has used the Met Office Integrated Data Archive System (MIDAS) to store station data since 1997. The MIDAS archiving process includes both automatic and manual quality control (QC) mechanisms, which provides a means of flagging suspect or corrected data. Data prior to 1997 is also stored in MIDAS, however these data have not passed through the MIDAS QC processing; historical QC processes may have been applied to these data. In light of the inconsistent QC processes applied to the data throughout the full period of record used in this study, the data are used "as is", taking no account of any MIDAS QC flags.

Table 1. Meteorological variables and associated number of stations in the network based on the proportion of non-missing data. The minimum record length threshold applied to the variables for the homogenised networks is indicated by the number of stations, in italics.

Meteorological variable	Number of stations with any data	Number of stations with at least 20 years of 90% complete	Number of stations with at least 15 years of 90% complete
0900 UTC 2 m air temperature (AIRTEMP)	891	277	*387*
0900 UTC 2 m dewpoint temperature (DEWPOINT)	861	234	*326*
24-hour maximum 2 m air temperature (AIRMAX)	941	*347*	457
24-hour minimum 2 m air temperature (AIRMIN)	939	*349*	461
24-hour sunshine duration (SUNSHINE)	543	188	*248*
24-hour average 10 m wind speed (WIND)	368	96	*143*

AIRTEMP and DEWPOINT are only available for a limited number of stations (~50) prior to 1972. Additionally, in the period 1994-1999 the number of stations with available DEWPOINT is limited; ~35% stations have data during this period. The DEWPOINT observations are derived either from directly measured dewpoint temperature, or calculated from the wetbulb temperature. During the period 1994-1999, dewpoint temperature measurements derived from the wetbulb temperature were not archived; the wetbulb temperature was archived.

SUNSHINE observations across the network are derived from a mixture of manual observations from Campbell-Stokes recorders (CS) and automatic observations from Kipp and Zonen pyranometers (AUTO). The type of instrument used varies by site and date; at selected sites both instruments are operational for an overlapping period. Based on the overlapping observations for both instruments at several sites around the entire UK, *Legg* (2014) derived a means of converting daily AUTO observations to their CS-equivalent. This methodology uses a quadratic regression relationship where the coefficients are derived for each month using daily data. For the current study, it was decided that converting the CS observations to their AUTO-equivalent would be preferable, as in future, it is likely that the number of stations with AUTO instruments will exceed those with CS instruments. The AUTO-equivalent observations are then derived from the CS observations by inverting the regression equation of *Legg* (2014). While this is not the ideal approach (a new regression relationship and associated coefficients should be determined to derive AUTO-equivalent observations from CS observations), revising the methodology is beyond the scope of this study.

The World Meteorological Organisation recommends observations of windspeed be made at a height of 10 m; where this is not possible, the observations should be made at an effective height of 10 m, to account for the influence of surrounding topography and obstacles. Several stations are found to have incomplete mast height information in the digital station metadata accompanying the observation data. For a few of these sites, additional mast height metadata are retrieved from the paper station records held at the National Meteorological Archive at the UKMO in Exeter. For any remaining stations with incomplete mast height metadata, the effective mast height is assumed to be 10 m.

2.2 Homogenisation software selection.

Several climate data homogenisation software packages evaluated in the European Cooperation in Science and Technology (COST) program Action ES0601 (COST-HOME) project (*Venema et al.,* 2012) possess aspects suited to the current study. Packages are ranked on their performance in the *Venema, et al.* (2012) benchmarking study, and based on their ability to meet the following criteria: the homogenisation process used in the current study needs to be

- repeatable (to allow consistent future updates),
- objective (to minimise errors introduced by inexperienced users),
- automatic (to minimise time spent on the process and keep costs down),
- appropriate to all climate variables (to allow application of a single consistent process),
- applicable to daily data.

To ensure robustness and reproducibility of the final homogenised dataset, the fully-automatic climate data homogenisation package Multiple Analysis of Series for Homogenization (MASH; *Szentimrey*, 2014) is used. In addition to meeting all the criteria above, and performing consistently well in the benchmarking study, MASH allows both additive and cumulative meteorological variables to be homogenised using a single consistent process. MASH also includes built in tools for assessing the performance of the homogenisation process, as well as performing QC on the homogenised daily series at the end of the homogenisation process.

2.3 Application of MASH

The application of MASH to the climate data station networks for each of the meteorological variables identified in *Table 1* places restrictions on which stations are included in the network, due to the way MASH infills missing data prior to running the iterative homogenisation process. Therefore, a minimum record length threshold is imposed for each variable using the criterion that a given year of data must be at least 90% complete. The minimum record length applied to each meteorological variable is determined by the number of repetitions of the infilling missing data processes in MASH required before homogenisation can be performed. The initial minimum record length is 15 years; if after 20 repetitions all missing data is not filled and homogenisation cannot begin, the minimum record length is increased to 20 years. The network size available for homogenisation decreases as the minimum record length increases for all variables, therefore a shorter minimum record length is desired to maximise network coverage across the domain and at higher elevations.

The success of the missing data infilling process in MASH is heavily dependent on the proportion of missing data across the entire network in any given year. Consequently, for AIRTEMP and DEWPOINT, where the number of stations with any data in the 1961-1971 period is limited, MASH is unsuccessful in filling missing data. Therefore, for AIRTEMP and DEWPOINT, the period for homogenisation is restricted to the 1972-2012 period.

As noted in Section 0, MASH includes several tools for assessing the performance of the homogenisation process. The most important of these tools is the Test Statistic (TS), which is used internally within MASH for the identification and correction of break points, but can also be used to determine how well the homogenisation process has performed. The null hypothesis that a data series is homogeneous is tested with the TS through comparison of the

TS value to a critical value (defined by the data series length and the desired statistical significance level). Therefore, TS is a measure of the inhomogeneity within an individual station data series; large TS values indicate greater inhomogeneity in the data series, while TS values equal to or less than the critical value are considered homogeneous. For each station, TS values are calculated for the missing data-filled series before the homogenisation process begins (TS before; TSB) and again after the homogenisation process has completed (TS after; TSA). Comparison of these two values (TSB and TSA) with each other, and in comparison to the critical value, provides information of the improvement in homogeneity of each station data series individually, but also for the network as a whole.

Following the detection and estimation of the magnitude of inhomogeneities in monthly station data series, the daily station data is homogenised through linear interpolation of the monthly inhomogeneity corrections. The homogenised daily data then passes through the MASH QC process to identify and correct any outliers. Statistics of the magnitude and frequency of errors detected and corrected during this process provide a useful overview of the daily station data, which can be useful for identification of particularly problematic stations.

3. RESULTS

Table 2 details the size of network for each meteorological variable upon completion of the MASH homogenisation process, along with the number of repetitions required to successfully infill all missing data. Note, WIND is still undergoing homogenisation.

The following sections present selected results from applying MASH in the current study, highlighting some specific issues and challenges associated with the data used in this study.

Table 2. The number of stations remaining in the network, for each meteorological variable, following homogenisation with MASH, and the number of repetitions of MASH required to complete the homogenisation process.

Meteorological variable	Homogenised network size	Number of repetitions of MASH required
AIRTEMP	344	16
DEWPOINT	241	14
AIRMAX	323	9
AIRMIN	323	8
SUNSHINE	182	11
WIND	-	-

3.1 AIRMAX

The final homogenised network of 323 AIRMAX stations is shown in *Fig. 2*, with the stations coloured by station elevation. Stations meeting the minimum record length threshold but not in the final homogenised network (i.e., stations where missing data is not completely filled; a total of 24 stations) are also shown.

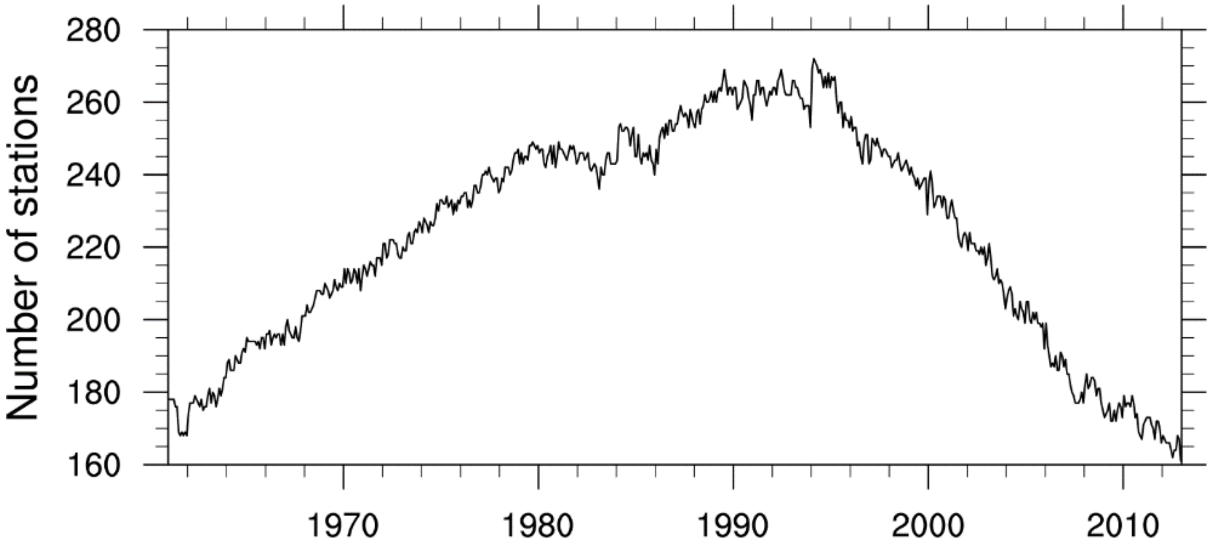


Fig. 2. Network information for AIRMAX. Left: Station network before and after homogenisation, coloured by station elevation; stations not in the final homogenised network are represented by the semi-transparent squares, those in the final network are solid circles. Right: Temporal availability of data across the entire network for stations included in the final homogenised network.

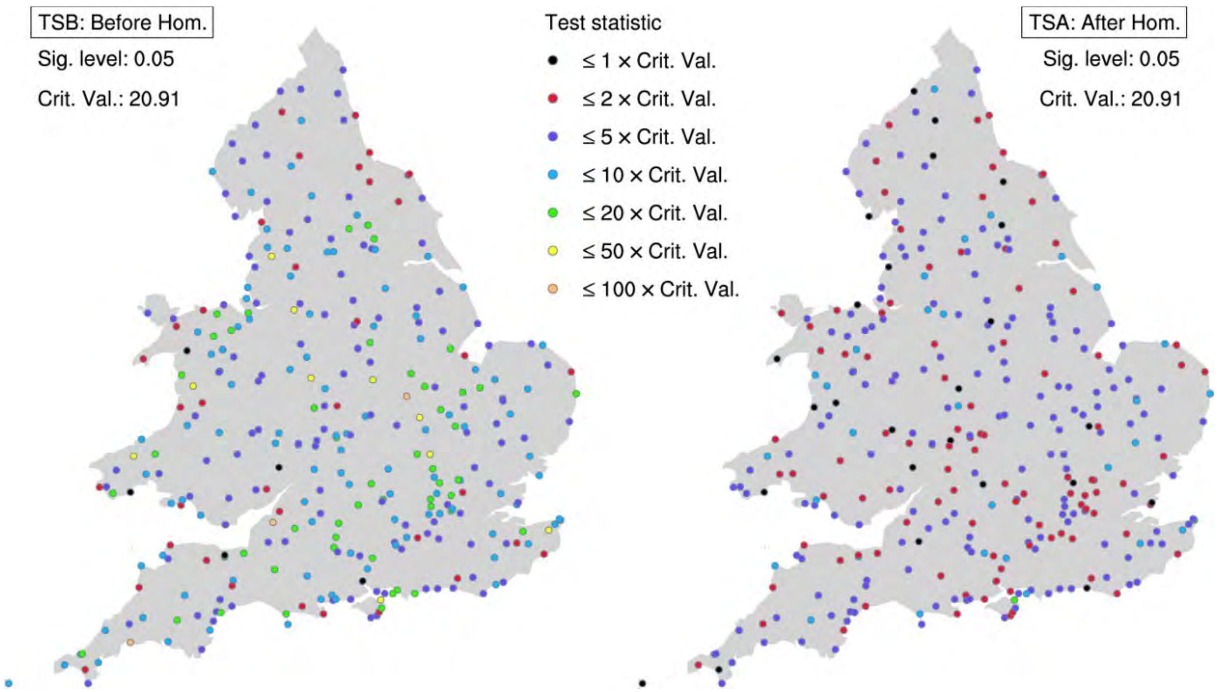


Fig. 6. MASH Test Statistics for AIRMAX for each station, (left) before homogenisation (TSB), and (right) after homogenisation (TSA), coloured in comparison to the critical value.

The majority of higher elevation regions identified in Section 1 are represented in the final homogenised network, except for the North Yorkshire Moors, where both stations at elevations greater than 200 m are removed during the MASH process, and the Lake District, where no stations above 200 m meet the minimum record length threshold. *Fig. 2* also shows the temporal availability of non-missing station data for stations in the final homogenised network. Although the final homogenised network is composed of 323 stations, for any given month during the 1961-2012 period, there is no more than 270 stations with non-missing data, and frequently fewer than this number, particularly near the start and end of the period. This pattern is partly due to the use of a minimum record length criterion (stations that cease observations <20 years after the start of the period, and those that begin observing <20 years before the end, are not included), and partly due to the overall increase in observing sites leading up to the mid-1990s, and the overall reduction thereafter.

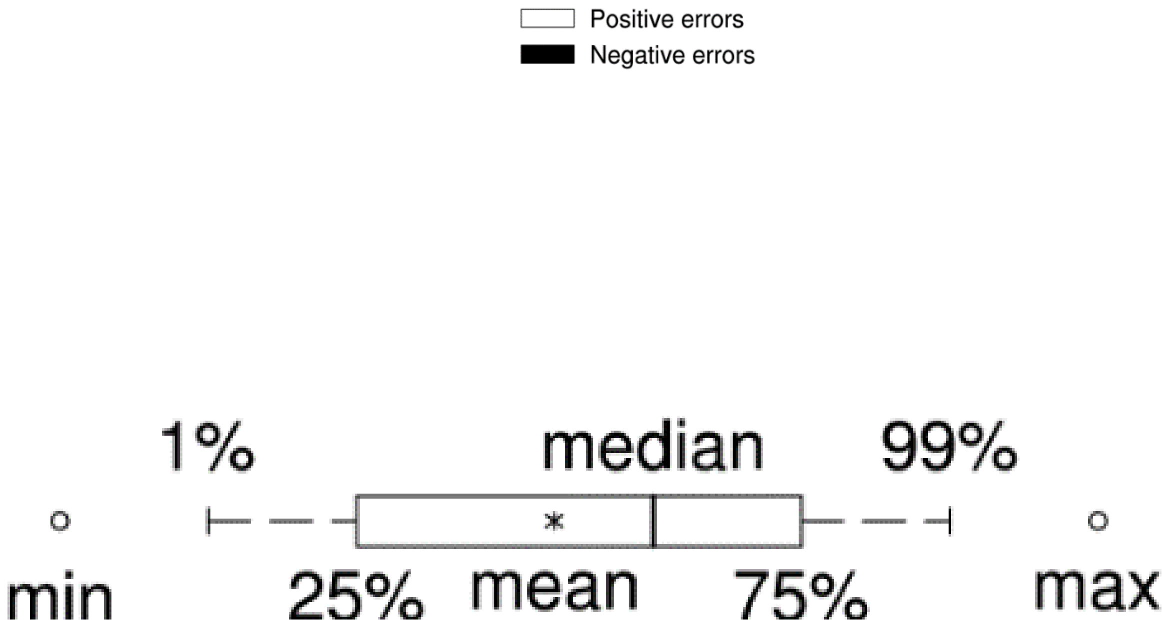


Fig. 4. Seasonal summary of MASH QC positive and negative AIRMAX errors in the homogenised daily station data. Top: Proportion (%) of observations containing errors. Bottom: Distribution of the magnitude (°C) of errors.

Fig. 63 presents the TS values before (TSB) and after (TSA) homogenisation for the final homogenised network for stations for AIRMAX. Taking the network as a whole, homogeneity of the network data series has improved following homogenisation with MASH. Prior to homogenisation, few stations exhibit a TSB value less than twice the critical value; after application of MASH, many more stations fall into this category and only one station exhibits a TSA value greater than 10 times the critical value. Taking stations individually, while in general TSA values are smaller than TSB values, this is not universally true; for a few stations the degree of inhomogeneity in their station data series has increased following the application of MASH.

Evaluation of the results from the MASH QC process reveals some interesting insights into the AIRMAX network data (*Fig. 4*). There exists a seasonal cycle in both the frequency and magnitude of both positive and negative AIRMAX errors in the daily data. Positive AIRMAX errors are more frequent than negative errors throughout the year, but this difference is more pronounced in winter months, when errors are more frequent overall. The magnitude of positive and negative errors is similar in any given month, however errors of both signs are larger in summer months. Similar seasonal cycles are not exhibited by all meteorological parameters homogenised in this study, and in general, MASH handles data for a given month in all years relatively independently from other months, hence it is unlikely the pattern of errors is an artefact of the MASH QC process. This then raises interesting questions of the source of the seasonal pattern of errors in the AIRMAX data. Are there some inherent biases in the instrumentation or observing and recording practices that could lead to more frequent positive errors in winter than summer, but larger errors in summer than winter? For all months, 99% of the errors are smaller than 6°C for both positive and negative AIRMAX errors, with the mean and median errors generally less than 1°C. Therefore, errors in the daily AIRMAX dataset are generally relatively small.

Fig.5 shows the frequency of positive and negative AIRMAX errors for each station in the final homogenised network. As in *Fig. 4*, the frequency of positive errors is higher than that for negative errors across the network as a whole, but this is not necessarily true for individual stations. Most stations have relatively few errors, with only a few stations with a large proportion of observations containing errors. A very small subset of stations exhibit a large number of both positive and negative errors. Such serially offending stations should be investigated further, and possibly removed from any further homogenisation projects to avoid contamination of the homogenised dataset. Additionally, stations with very large errors, of either sign should be investigated to identify the source of the error and potentially remove those offending observations.

3.2 Other issues

The homogenisation of some of the meteorological variables using the available data is less straightforward than the example presented above for AIRMAX. For example, the SUNSHINE dataset exhibits significant gaps in coverage across several regions of the domain when considering the available stations meeting the minimum record length threshold (15 years), both before and after the homogenisation process with MASH (*Figure 6*). In particular, the lack of coverage in inland areas and the south coastal of Wales, southeast England and coastal areas of eastern England is of concern. These gaps may have a significant impact of the efficacy of the calculated PET dataset in these areas. It is hypothesised that these gaps may in part be due to the steady decline in the number of sunshine-observing stations from approximately 1980 onwards. With insufficient active available in the latter portion of the record to infill data missing from stations that have ceased observing, a large number of stations (66) are dropped from the network during the MASH data-infilling process prior to homogenisation.

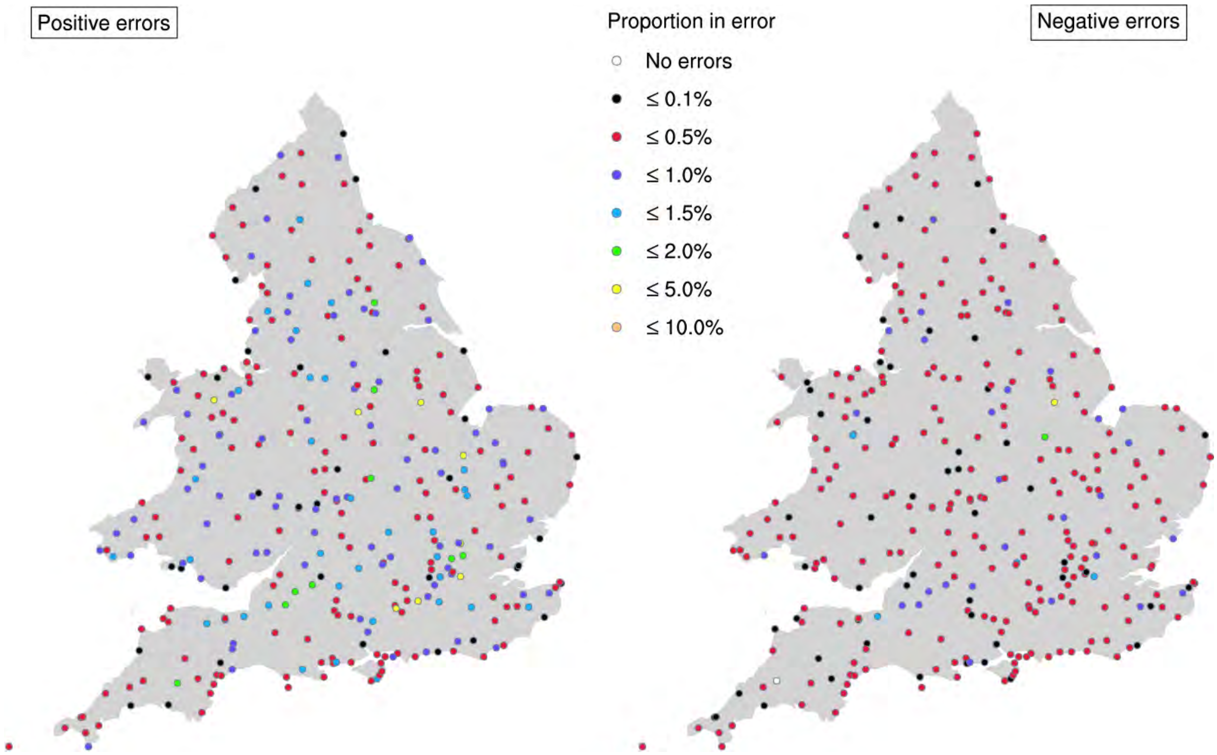


Fig. 5. Frequency of (left) positive and (right) negative errors in daily AIRMAX from the MASH QC process.

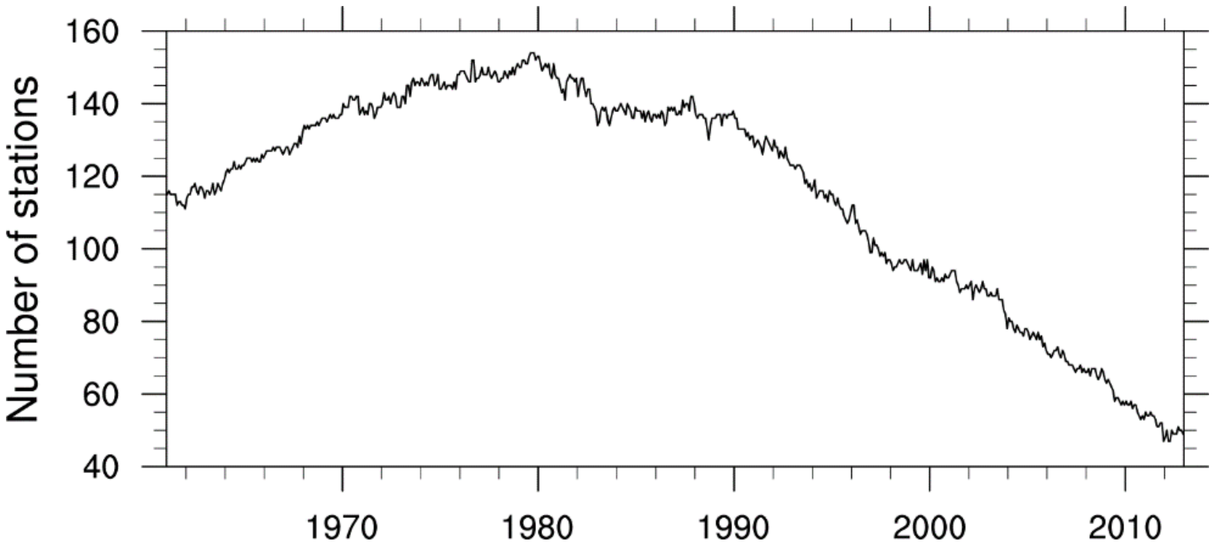


Fig. 6. Network information for SUNSHINE. Left: Station network before and after homogenisation, coloured by station elevation; stations not in the final homogenised network are represented by the semi-transparent squares, those in the final network are solid circles. Right: Temporal availability of data across the entire network for stations included in the final homogenised network.

A similar result is also exhibited by the DEWPOINT dataset (*Fig.7*), however, in this case the absence of the majority of stations during the 1994-1999 period, as noted in Section 0, results in the significant gaps in coverage. Again, inland areas of Wales are particularly impacted in the final homogenised DEWPOINT dataset. For future homogenisation studies using this dataset, this issue could be negated by recalculating the missing dewpoint values from the archived wetbulb temperature data.

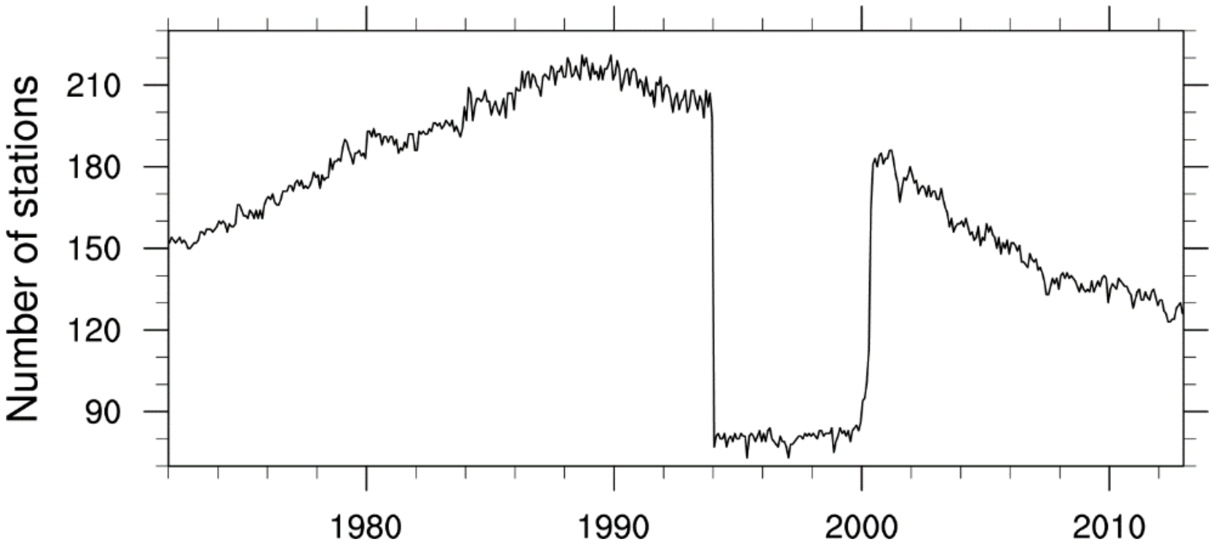


Fig. 7. Network information for DEWPOINT. Left: Station network before and after homogenisation, coloured by station elevation; stations not in the final homogenised network are represented by the semi-transparent squares, those in the final network are solid circles. Right: Temporal availability of data across the entire network for stations included in the final homogenised network.

The DEWPOINT dataset also exhibits some interesting characteristics in the results from the MASH QC process on the homogenised daily data (*Figure 8*). In terms of the frequency of errors, the 1994-1999 period exhibits relatively few observations in error. There are two possibilities for this: i) observations where the dewpoint is directly sensed rather than calculated from the dewpoint are less prone to errors, or ii) as a consequence of the missing data filling process, a larger proportion of observations in this period are derived from the non-missing data during this period, so there is less chance for observations to be determined to be in error. Further investigation is required to determine which of these is correct.

There are years when positive errors are nearly twice as frequent as negative errors, and this occurs in both the pre- and post-MIDAS QC era. Further investigation of the observations in error and their associated MIDAS QC flags is required to determine if the more frequent positive errors in the post-MIDAS QC era are identified by the MIDAS QC process. Preliminary investigations have revealed that the occasions of frequent positive errors appear to cluster together at a given stations, but such clustering of errors occurs at many stations, but not at the same time. Again, this may be due to systematic issues with instrumentation or observing and recording practises, but in-depth investigations would be required to identify the source of such errors.

The magnitude of negative errors can be very large, in the range 40-80°C. Preliminary investigations into these erroneous observations have revealed that the majority of these very large DEWPOINT errors are flagged as erroneous in both the pre- and post-MIDAS QC era. However, some of the moderately large DEWPOINT errors, in the range 10-20°C, are not flagged as erroneous, even in the recent MIDAS QC era. This has significant impacts in the quality of the data archived in MIDAS, and Therefore this information is being fed back to the UKMO to help to improve the MIDAS QC process. The source of such large negative DEWPOINT errors also requires further investigation. There is a slight tendency for the largest errors to occur in winter months (not shown), however the moderate negative errors (10-20°C) are equally common in all months.

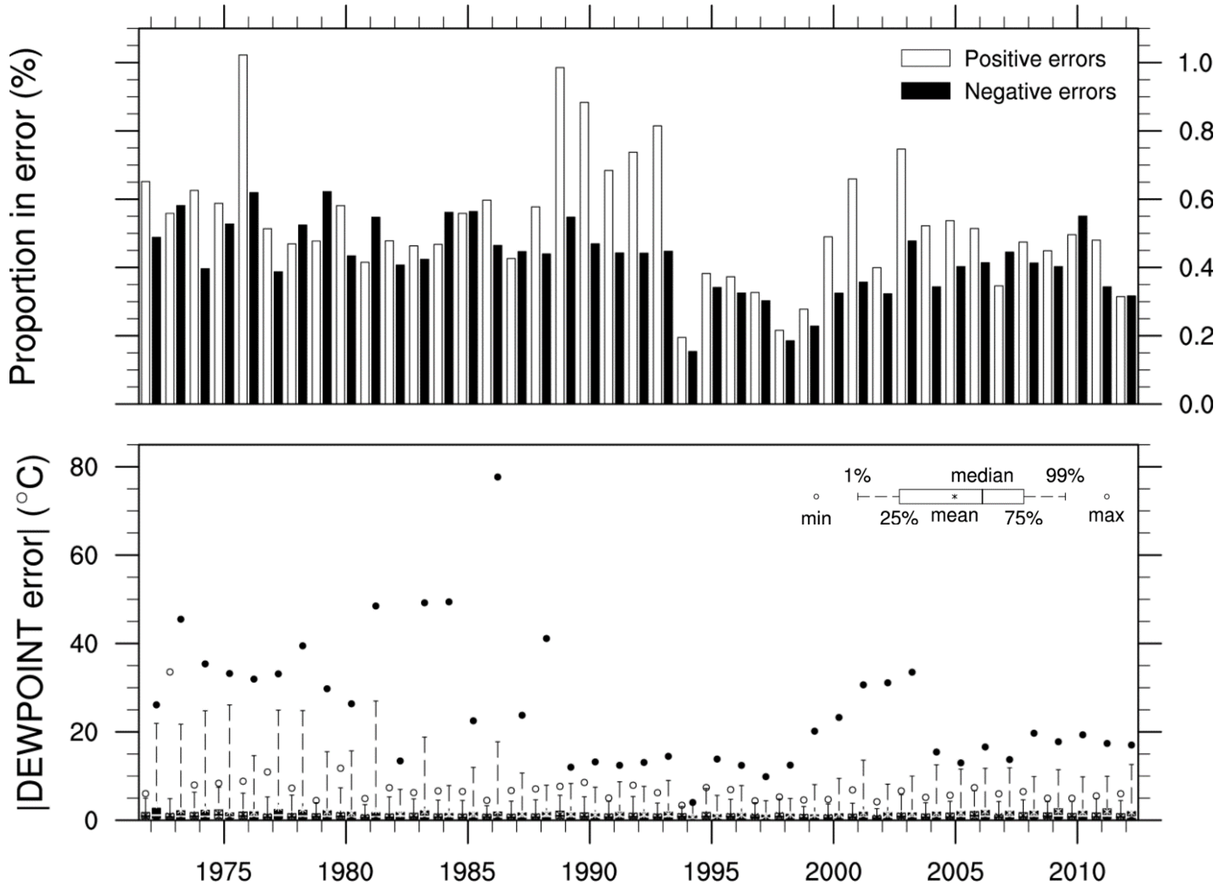


Fig. 8. Annual summary of MASH QC positive and negative DEWPOINT errors in the homogenised daily station data. Top: Proportion (%) of observations containing errors. Bottom: Distribution of the magnitude (°C) of errors.

4. CONCLUSIONS

This paper has identified that homogeneous datasets now exist for five of the six meteorological parameters required for PET calculation in England and Wales. Specific issues exist with some of the parameters due to the availability of stations data with sufficiently long records. For SUNSHINE, there exist significant gaps in coverage, particularly in Wales. These gaps are most likely due to the steady decline in the number of sunshine-observing stations since 1980. In future homogenisation studies covering this domain, directly observed radiation data could be used to increase the coverage of the SUNSHINE dataset by creating a combined radiation dataset. For DEWPOINT, significant gaps in coverage are also present, again, particularly in inland Wales. The limited DEWPOINT data availability during the 1994-1999 period is almost certainly responsible for most of the gaps in coverage. In future homogenisation studies covering this domain, DEWPOINT data for wetbulb observing stations could be recalculated for this period to negate this problem.

Following the MASH homogenisation process, the application of the MASH QC process to the homogenised daily data has revealed some interesting insights into the datasets. In general, the MASH QC process appears to be an effective tool to identify erroneous observations. Preliminary investigations into some of the errors identified by the MASH QC process have verified the observations as suspect and most likely erroneous. The MASH QC process is useful for identifying individual stations with a high proportion of erroneous observations. Future homogenisation studies covering this domain should consider removing such stations prior to homogenising the data network to avoid contamination of the homogenisation process by bad data. Of particular interest from the MASH QC results is the appearance for seasonality in the magnitude and frequency of errors in some of the meteorological parameters; the seasonality is not believed to be an artefact of the MASH QC process. Further investigation is required to determine the source of the identified seasonality in the errors, but it does pose interesting questions about possible biases in instrumentation or observing and recording processes.

Next steps

The current study is now to be extended to include the following updates:

- Extend the station records to the end of 2015 for all parameters.
- Include stations in Scotland close to the border with England to improve the availability of data for homogenisation for stations in northern England.
- Include radiation data to create a combined radiation dataset based on directly observed radiation and that derived from the sunshine dataset.
- Investigate options for extending the temperature and dewpoint datasets back to 1961, such as interpolating the 0900 values from the 0600 and 1200 values.
- Include the additional wind mast height information obtained from digital and paper records at the UKMO.

From the current study, any serially erroneous stations based on the MASH QC process will be removed prior to beginning the MASH process. Additionally, further rounds of homogenisation may be employed after the final homogenised network has been obtained. This may improve the homogenisation performance at stations where homogeneity decreased after the MASH process was completed.

The use of an automatic homogenisation software package in the current study has raised questions about how best to maintain an operational PET dataset based on homogenised station data. The gridded PET dataset that is the ultimate aim of this project will be used to re-calibrate rainfall-runoff and groundwater models for the historical period, and these models may then be used operationally in real time to determine groundwater recharge, river flow and soil moisture deficit to assess water availability If the full homogenised meteorological dataset changes each time new, recent data is added in and the entire dataset is homogenised together, this will result in very costly re-runs of the historical rainfall-runoff and groundwater models. What is the best, and most cost-effective, way to handle updating the system every 5 or 10 years?

Additionally, there are a large number of stations with record lengths not meeting the minimum length threshold. Is there a way these stations could be incorporated into subsequent rounds of homogenisation after the set of stations meeting the threshold have been homogenised?

Acknowledgements

This project has been funded by the Environment Agency. The views in this paper are those of the individual authors and not necessarily those of the organisations they represent.

References

Allen, R.G., L. A. Pereira, D. Raes, M. Smith (1998). Crop Evapotranspiration. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. *FAO Irrig. and Drain. Paper* No. 56. United Nations FAO, Rome, Italy.

Amante, C. and B.W. Eakins, (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M [accessed 29 Jun 2017].

Legg, T. (2014). Comparison of daily sunshine duration recorded by Campbell–Stokes and Kipp and Zonen sensors. *Weather*, 69: 264–267. doi:10.1002/wea.2288.

Sharpe, M. C. (2012), Inhomogeneities in MORECS wind data, Met Office, Exeter, unpublished report for Environment Agency.

Szentimrey, T. (2014). Multiple Analysis of Series for Homogenization (MASH v3.03). Hungarian Meteorological Service, Budapest, Hungary, 70pp.

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

HOMPRA EUROPE – A GRIDDED PRECIPITATION DATA SET FROM EUROPEAN HOMOGENIZED TIME SERIES

Elke Rustemeier[1], Alice Kapala[2], Anja Meyer-Christoffer[1], Peter Finger[1], Udo Schneider[1], Victor Venema[2], Markus Ziese[1], Clemens Simmer[2], Andreas Becker[1]

[1]Deutscher Wetterdienst, Hydrometeorology, Offenbach am Main, Germany
[2]Meteorological Institute, University of Bonn, Bonn, Germany

Abstract

For the understanding and analysis of long-term trends in precipitation, a reliable data base of quality assured and homogenized records is essential. Changes in the environment or location, but also errors during the data recording, lead to breaks and outliers in the data, hampering the representativeness of the record itself and any subsequent analysis. To address these issues we have developed – „HOMPRA Europe", a HOMogenized PRecipitation Analysis of European in situ data as a BIAS-corrected monthly rainfall data set with 1° spatial resolution, and added it to the portfolio of observational data products of the Global Precipitation Climatology Centre (GPCC). The data set consists of 5536 precipitation time series, carefully selected from the GPCC data archive, being the largest archive of quality controlled precipitation data world-wide. For this data collective of time series covering the period 1951-2005 with less than 10% of missing values we have applied a newly developed automated homogenization scheme, and present it in sufficient detail here.

The automated homogenization approach was forced by the constraint to process a high number of series within a reasonable time. However substantial testing and refinements have been necessary to make the algorithm resilient enough to cope with the large variety of rainfall distributions in Europe representing also arid regions where consecutive months without precipitation occur regularily.

Therefore, the detection and correction performance of the homogenization algorithm was tested against artificial data sets with defined breaks. In addition, a sensitivity study was applied by varying the reference stations. If available in digitized form, the station history was also used to search for systematic errors in the break-point detection. Finally, the new HOMPRA Europe product is produced by interpolation of the homogenized series onto a 1° grid using one of the interpolation schemes used operationally at GPCC (*Becker et al.,* 2013).

1. INTRODUCTION

Precipitation measurements are in particular sensitive to changes that influence the wind at the observation site, e.g. relocation or changes in exposure (*Beaulieu*, 2009). Since time-series analyzes must be based on robust time series, it is essential to correct the data for inhomogeneities to make reliable statements.

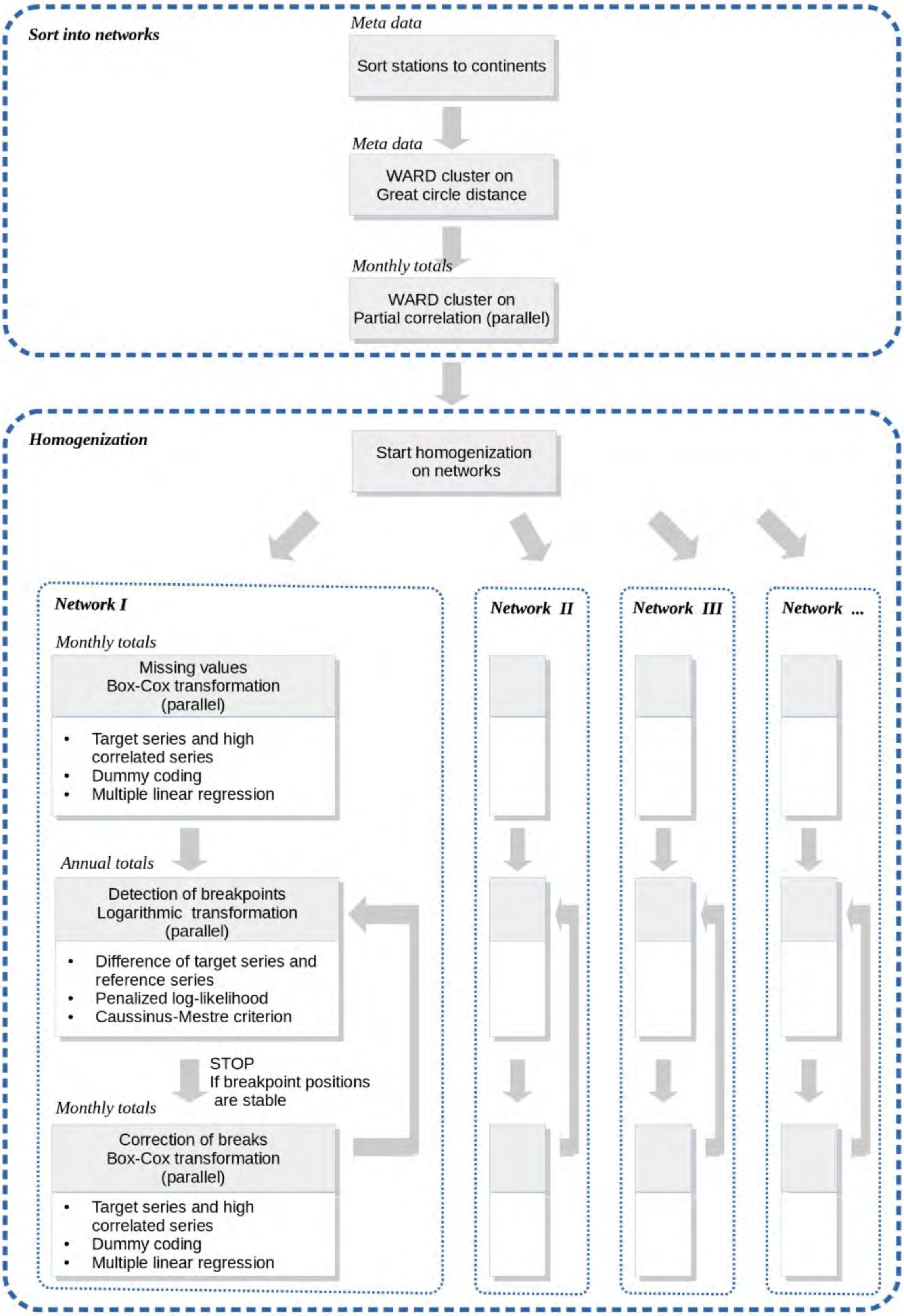


Fig. 1. Flow chart of the homogenization algorithm for the creation of the HOMPRA Europe data set

This paper presents a new gridded and homogenized European precipitation data set, which allows for the analysis of monthly, seasonal and annual trends between 1951-2005. For this purpose, high-quality stations, selected from the GPCC data base, are compared with other series and examined for their unique systematic changes. This is to ensure that only the artificial changes and no natural changes are corrected. These changes are then corrected using the surrounding stations. This method is applied with the hypothesis that natural changes would synchronously occur also in the neighbouring stations, so a unique break can only have an artificial cause. Obviously the method requires an existing correlation between the station of regard and its neighbours, which is more or less valid depending on the distance between the stations and the variability and the intermittency of the precipitation records examined.

From now on, the time-series to be homogenized will be referred to as target series and the homogenization supporting time-series as reference or neighbor series.

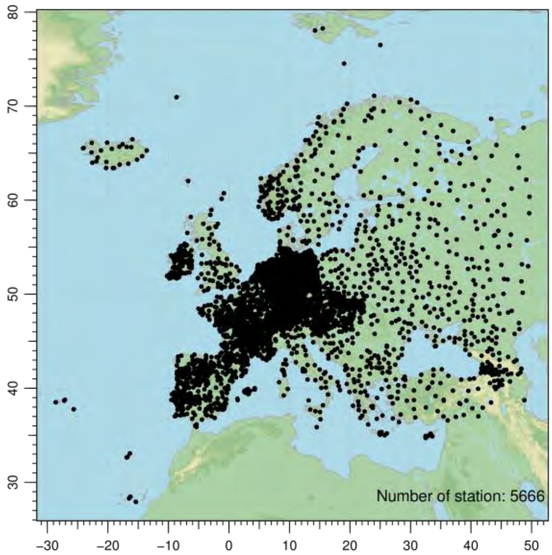
The homogenization algorithm is summarized in *Figure 1* and basically consists of two parts:

- Selection of overlapping station networks in the same precipitation regime, based on partial rank correlation and WARD's method of minimal variance (*Wilks*, 2006). Since the underlying time series should be as homogeneous as possible, the station selection is carried out by deterministic first derivation in order to reduce artificial influences.
- The natural variability and trends were temporally removed by means of highly correlated neighboring time series to detect artificial break-points in the annual totals. This ensures that only artificial changes can be detected. The method is based on the algorithm of *Caussinus and Mestre* (2004). Finally, the detected breaks are corrected monthly by means of a multiple linear regression (*Mestre*, 2004).

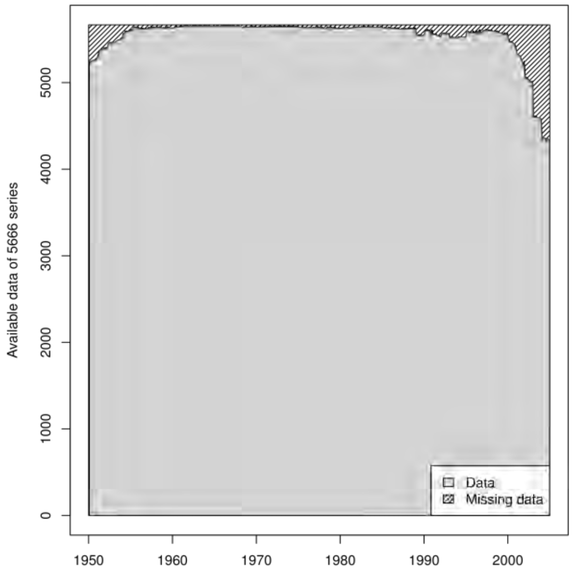
The presentation of the procedure and the results are organized as: First, the data base is presented in Section 2. Subsequently, the fully automated homogenization is described in the Section 3, which consists of the selection of reference time-series, the detection of break-points and the correction of the data. Because of the automation the validation of the results described in Section 4 is particularly important. The final data set is generated by interpolation to a 0.5° regular grid using the GPCC standard methods, which are described in Section 5. Finally, in the last section 6 the effects and changes resulting from the homogenization are discussed.

2. DATA BASE

The data base for homogenization is a subset of the data collective residing in the GPCC archive. The collective is inspected for time series, which already in itself have a high quality. For Europe, 5666 stations can be selected. All these time series are quality-controlled and have passed the GPCC's standard controll procedures i.e. outlier checks, and comparisons with neighboring stations and area values (for details see *Schneider et al.*, 2014). The time-series cover the period 1951-2005 and have only few missing values. This is shown in *Figure 2*. The figure also shows that most of the missing values occure at the beginning and especially at the end of the time-period. The spatial distribution of the stations is very inhomogeneous. The station density over Germany is particularly high, whereas in the south-eastern Azerbaijan only the time-series of city Neftçala fulfills the criteria, so that time series from Iran had to be used as reference for Neftçala.



(a) Map of the stations thatform the base of the HOMPRA Europe product



(b) Overview of completeness and missing values in the data base

Fig. 2. Station data base for the HOMPRA Europe product for the period 1951-2005

The high density of the stations in many areas also makes it possible to vary the reference stations used for the homogenization and thus allow for an estimation of the general influence of the neighbor stations on the homogenization, including the number of stations and correlation with the target station. It also permits to estimate the uncertainty of individual stations. This is described in more detail in Section 3.4.

3. AUTOMATIC HOMOGENIZATION ALGORITHM

The homogenization algorithm, as already shown in the introduction in *Figure 1*, consists of two steps. The first is the subdivision of the stations into station groups or networks, and the second step consists of an iterative detection of break-point positions and correction of these breaks within the time series. The iterative process is necessary, because it can not be assumed that the reference stations, are homogeneous. Therefore, the iterative process of detection and correction is performed on target and reference series, until the algorithm converges and the detected break-point positions no longer change.

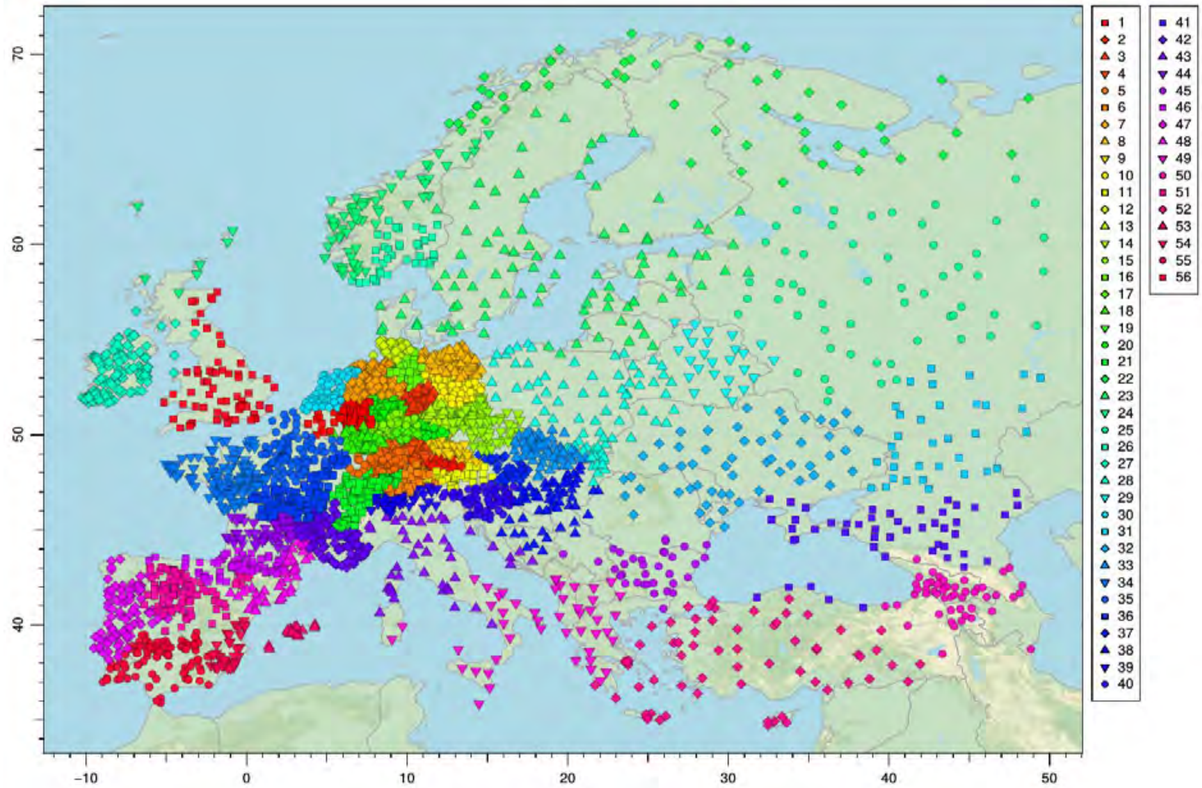


Fig. 3. Station networks of the European stations as used for homogenization.

3.1 Networks of similar time series

To warrant and efficient utilization of computing time and random access memory (RAM), we did not offer all stations as potential neighbor stations, which is not necessary, since higher correlated time series are available. Therefore, the 5666 stations have been divided into groups with an average number of 300 stations. In order to avoid border effects the network topologies have furthermore been chosen to overlap each other.

Since it can be assumed that many of the time series are inhomogeneous, they are used as homogeneous as possible for the selection of the networks. As suggested by *Moberg and Alexandersson* (1997) the consecutive differences, called the “ deterministic first derivation”, are calculated from each time series. As a result, any changes on average show only one high or low outlier, but otherwise the time series is not affected. In addition, not the values but the ranks are considered, since precipitation is not normally distributed.

Since a network selection complicated by an inhomogeneous spatial distribution, a partial rank correlation is calculated instead of the standard correlation, based on partial least squares (PLS) regression (*Geladi and Kowalski*, 1986). Compared to other partial correlation methods the PLS regression has a high mean squared error, but is also very fast, and has a high true discovery rate (*Krämer et al*., 2009). Partial correlation causes the correlation coefficients to become lower when the spatial coverage is dense. The parameters, such as the number of neighbor series included for calculating the correlation coefficients, were optimized for the subsequent cluster algorithm: The WARD Cluster method (*Wilks*, 2006) was used as a clustering algorithm, since it tends to select groups of the same size based on the variance, so in this case the variance of the partial rank correlation.

As shown in *Figure 3*, the resulting 56 networks show very clear spatial structures and very distinct boundaries, although only the measured time series and no metadata including coordinates were included. The algorithm can handle the division between dense and less dense regions simultaneously. There are also very few network borders which coincide with the actual country borders. It is not shown here that the method also works very well with other precipitation distributions on all other continents.

For the actual homogenization, stations were assigned to the respective networks according to the coordinates from the surrounding networks in order to avoid boundary effects. For very dense networks, the time series were thinned since the homogenization parameters were not optimized for such dense networks, so eventually 5536 were homogenized.

3.2 Detection of break-points

The detection of break-points is carried out on the annual totals only, as this substantially reduces the risk of noise related false break-detection, that is likely in general given the very small signal-to-noise ratio for precipitation. Since the detection algorithm assumes a normal distribution, the time series are transformed by applying the logarithm to the data. Normal distribution tests show that the result is not yet optimal, but much improved.

In order to filter the natural changes, the homogenization is applied on the difference between the target station and the composite average series computed from several highly correlated reference series. As a reference time series within the network are chosen series which show a high Spearman-correlation (*Wilks*, 2006) based on the deterministic first derivation. A difference between the target and the neighbor series is indicative for the detection of artificial changes instead of natural ones. The composite average series of the neighbor time series is not weighted so that breaks in one of the neighbor stations can not dominate the result.

The actual detection is applied on these difference series according to the Caussinus-Mestre procedure (*Caussinus and Mestre*, 2004): First, the best break-point position for any number of breaks is determined by minimizing the log-likelihood of the difference series. The process is optimized in time by parallel computing and optimal segmentation by dynamical programming (Picard et al., 2007). Finally the number of the breaks is determined by a penalty term (*Caussinus and Mestre*, 2004).

3.3 Correction of breaks

Since break heights and even directions turned out to be specific for the month, the correction is carried out on a monthly basis. In order to obtain a normal distribution again, the Box-Cox Transformation (*Wilks*, 2006, see Chapter 3.4.1 Power Transformations) displayed in Equation 1 is applied to target and highly correlated neighbor series

$$Y_{new} = \begin{cases} \frac{Y^k - 1}{k}, & k \neq 0.000 \\ ln(Y), & k = 0.000 \end{cases} \tag{1}$$

with Y being the monthly target series and the neighbor series, and Y_{new} being the power transformed monthly series. The parameter k is optimized based on the t-test so that the transformed series Y_{new} is normally distributed.

The neighbor series are selected similiar to the break-point detection. The Box-Cox transformation is chosen, as the logarithm shows inadequate results in dry regions such as

Southern Europe and in very heavy precipitation areas as in Western Norway. Individual months in which there is almost no precipitation are excluded from the correction at single stations.

The correction of the time series is implemented as by *Mestre* (2004) in three steps. First, the data is binary-coded *Bortz* (2004), with parameters for the homogeneous segments obtained from the break-point detection and the time. This is followed by a multiple linear regression over the data. There are two different sets of regression parameters: On the one hand temporal parameters, which describe the temporal variability common between all stations, and on the other hand parameters describing the homogeneous segments, which form a kind of average over the homogeneous segments. The differences between the latter regression coefficients describing the segment, form the break heights and are used for the correction in the next step. Due to the previously applied Box-Cox transformation, the resulting values are normal distributed. Therefore an additive correction can be used instead of a multiplicative correction, in contrast to the common practice for precipitation. This is done monthly by using the segment regression coefficients to adjust the mean values of the segments.

The mean value of the entire time series is used as a fix point, which means that the average annual cycle of the time series does not change. One advantage over the common practice of keeping the last segment constant is that very short segment at the end can not distort the series. One disadvantage is that new observations can not simply be attached. Subsequently, the corrected time series are back-transformed using again the previously calculated k coefficients of the Box-Cox transformation. This interaction of detection and correction is repeated iteratively until the break-point positions converge.

3.4 Uncertainty assessment of the homogenized time series

Since the neighboring stations have been used as a reference, a sensitivity study can be performed to estimate the resulting uncertainty, by varying the neighboring stations. Different criteria were used: the number of neighbor series and the correlation of neighbor series. In this way, for stations in the relatively dense network, it is possible to estimate how reliable the break height is determined and how stable the break-point position is. This also results in a trend uncertainty. Unlike the default run, the reference series are not homogenized with their own optimized reference, but with the series of the target station. In contrast to the default procedures for HOMPRA Europe, series are also homogenized with poorly correlated neighbor series. Each time series is homogenized 50 times with different reference series. If there are enough neighbor stations, they are also classified into categories by means of correlation and number of neighbor stations between 5-10 for a better understanding of the influence of correlation and number of reference series on the result.

In addition, some tests are applied during the homogenization process, in order to find errors that were not detected during quality control. This includes a very high correlation between two time series, which suggests a duplicate entry in the data base, a very high correction factor, which may be a factor 10 errors, or a high number of 0 values compared to neighbor series, which may indicate that missing values were replaced by 0. These uncommon values are checked manually.

Before the subsequent interpolation for the established HOMPRA Europe final product, there is still a manual control. To be able to comprehend the results for each station, the following outputs are automatically created: the CRADDOCK test (*Craddock*, 1979) on original and homogenized data including neighbor series and detected breaks, the annual cycle including reference series, since they serve as a fix-point during the correction, absolute and relative time series before and after homogenization. In this way, series with conspicuous corrections (very high) or with very many breaks or stations without highly correlated reference series can be controlled manually and if required be excluded from the interpolation.

4. VALIDATION OF THE HOMOGENIZATION ALGORITHM

Table 1: False detected shifts [%] with homogenous neighbor series (but expected inhomogenous)

Record length	**Number of neighbors**	**Amplitude [standard deviation] 0-0.25**	**0.25-0.5**	**0.5-1**	**1-2**	**>2**	**Average**
55	**3**	0.04	0.58	5.78	0.28	0.00	1.33
	5	0.04	0.24	2.68	0.06	0.00	0.60
	8	0.00	0.10	0.90	0.00	0.00	0.20
	10	0.00	0.10	0.38	0.00	0.00	0.10
75	**3**	0.00	0.98	5.44	0.28	0.00	1.34
	5	0.02	0.64	2.00	0.04	0.00	0.54
	8	0.00	0.34	0.70	0.00	0.00	0.21
	10	0.00	0.18	0.32	0.00	0.00	0.10
100	**3**	0.00	1.84	4.48	0.36	0.00	1.34
	5	0.00	0.94	1.26	0.00	0.00	0.44
	8	0.00	0.42	0.24	0.00	0.00	0.13
	10	0.00	0.20	0.16	0.00	0.00	0.07

The validation of the algorithm is particularly important for automated homogenization because there is no detailed manual control by the user. The typical validation approach to test the code on independent data, is not possible in this case, since the requirement would be data with known breaks including break height in otherwise homogeneous time series, which does not exist. Therefore, the algorithm is tested on the benchmark data set "COST (European Cooperation in Science and Technology) Action ES0601: Advances in HOmogenisation MEthods of climate series: an integrated approach (HOME)" described by *Venema et al.* (2012) and on artificial data sets.

The artificial data is based on *Beaulieu et al.* (2007) the break-point position and the height are chosen randomly, with the break height being relative to the standard deviation (σ) between -3σ and $3\ \sigma$ with a maximum at $\pm\ 1.5\sigma$. Results for the artificial data are shown in *Table 1*, which shows the false break detections in percent, depending on the length of the time series and the number of neighboring stations. *Table 2* shows how sensitive the algorithm reacts to breaks in the reference series depending on thebreak direction, which is shown for 5 reference time series and varying time series length. For comparison, the length of time series of HOMPRA Europe are 55 years.

It can be seen in both tables that short time series are more difficult to correct than long ones. The performance can be improved, if more reference stations are used, which is especially important for short time series. Breaks which are smaller than one standard deviation, i.e. having a signal-to-noise ratio of less than 1, are particularly difficult. The software seldom tries to detect even smaller breaks. Larger jumps are easier to find and mostly positioned correctly. As in *Beaulieu et al.* (2007) three classifications are used in *Table 2*: correctly identified, when the exact position was detected and the relative difference between estimated magnitude and real amplitude is less than 10% of the real magnitude; well identified, when the estimated break position is less than 2 years from the real position and the absolute error of the magnitude is less or equal of 50% of the real magnitude; and well positioned, wenn the break is less than 2 years from the real position. *Table 2* shows the influence of the break direction. Again the number of reference series is important for the performance of the algorithm.

5. INTERPOLATION

The homogenized stations are reprocessed with the a modified SPHEREMAP algorithm as described in detail in *Becker et al.* (2013. This method uses a weighted average based on distance and angle and is applied to the anomalies on a monthly basis. In addition to Kriging, the modified SPHEREMAP interpolation scheme is one of the operationally used interpolation methods of the GPCC. HOMPRA Europe is available with 0.5°, 1.0° and 2.5° spatial resolution on a regular grid together with the interpolation error by *Yamamoto* (2000).

Table 2. Sensitivity test for break direction with 5 neighbor station and 1 break each [%]

Break direction	Record length	Correct identified	Well identified	Well positioned
+ + + + +	55	38.15	53.94	59.39
	75	42.31	60.26	62.82
	100	41.07	62.50	68.45
+ + + + –	55	40.68	63.24	69.07
	75	48.01	69.48	74.57
	100	48.09	73.22	79.64
+ + + – –	55	51.91	74.64	79.58
	75	54.71	78.54	85.41
	100	57.45	79.55	86.05
+ + – – –	55	54.76	78.27	85.07
	75	58.84	81.21	89.11
	100	61.45	81.81	90.59
+ – – – –	55	52.76	75.32	82.14
	75	58.84	76.86	84.47
	100	61.45	80.71	87.84
– – – – –	55	52.36	77.56	87.82
	75	52.43	83.12	91.56
	100	59.54	81.50	91.33

\+ Neighbor series with the same break direction as in target station

– Neighbor series with opposite break direction as in target station

6. RESULTS

The automation of the homogenization algorithm allows to homogenize a large number of time series and thus to create a European homogenized data set, the HOMPRA Europe data set, based on 5536 stations.

The saisonal trends before and after the homogenization are shown in *Figure 4*. It is noticeable that the distribution of the trends changes, but the maximum remains with the most frequent trends at the same value. In fall and summer, the maximum after homogenization is almost 0±0.25 mm/yr. In spring and winter the maximum is at a slightly positive trend of 0.5±0.25 mm/yr. The time series are corrected towards the direction of a common mean value. This narrows the range of trends, when the Inter Quartile Range of all seasonal trends before is 0.9, it is 0.8 after the homogenization. On average, the trends are corrected downwards by 0.03 mm/yr for each season. The median change is 0 mm/yr per season. Seasonal trend corrections often are of a different sign within one station.

For the uncertainty estimation of the homogenization, the respective target stations were homogenized with different reference stations. Three examples are shown in Figure 5. For each run, 10 reference series with a correlation of at least 0.8 were randomly selected. When varying neighbor series, the results for the individual stations can be very different. The results of Geneve-Cointrin, Twente and Basel/ Binningen are shown together with the trends of the original data as a red dot (•) and the HOMRA Europe results as a blue cross (**x**). In each case 10 neighbor series with a correlation of 0.8 are selected to compare homogenization results.

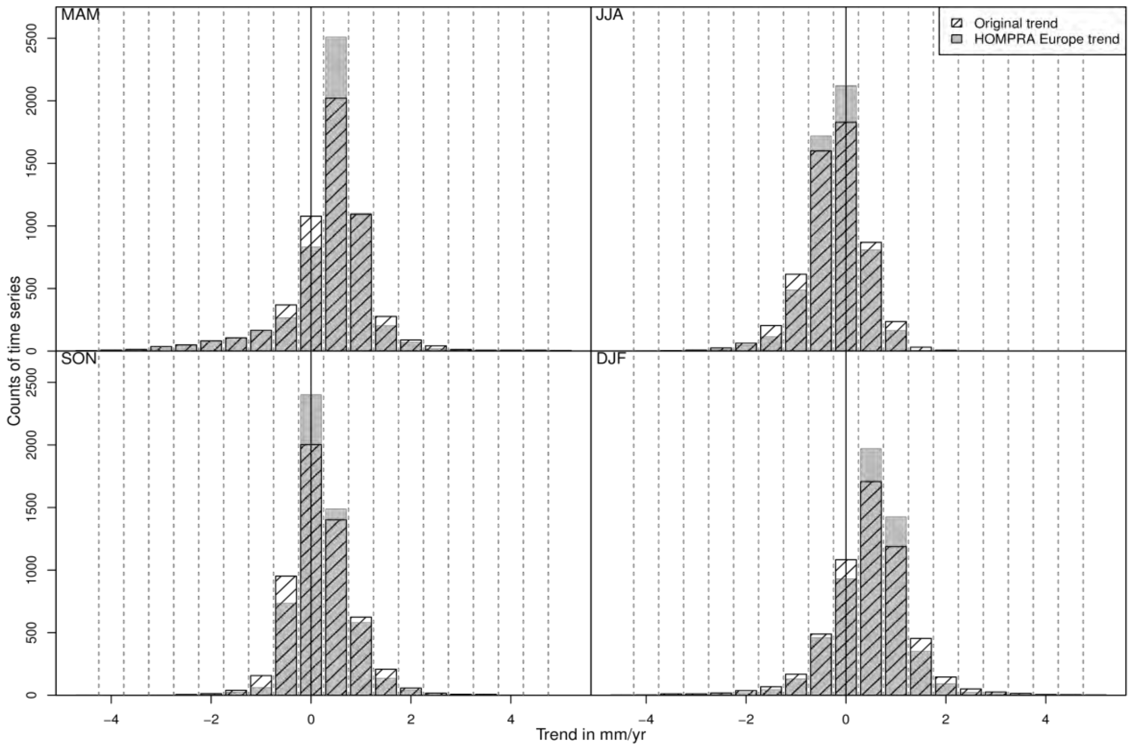
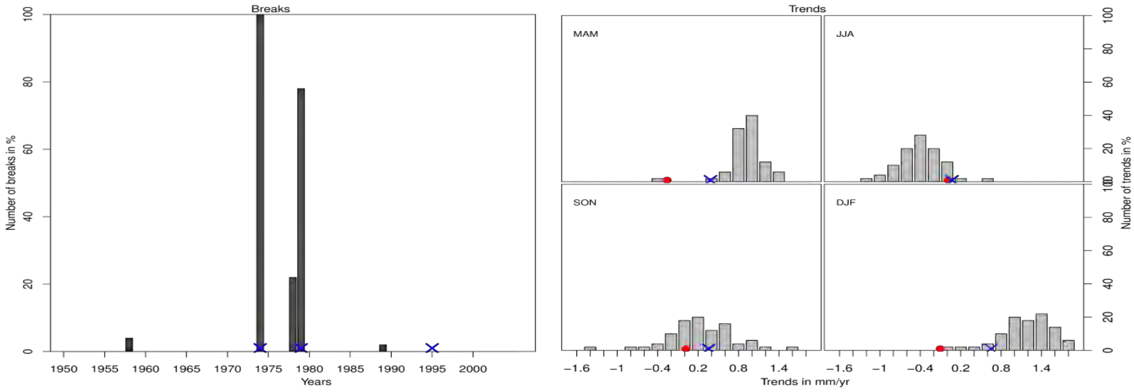
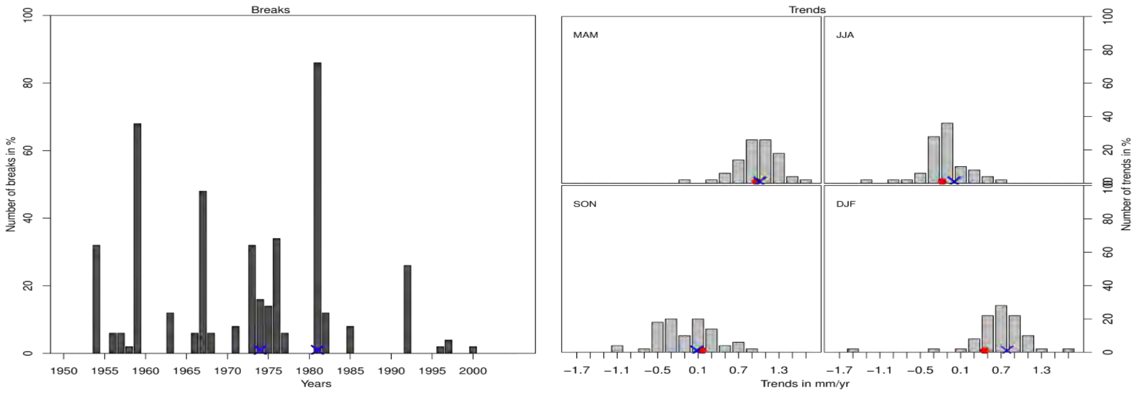


Fig. 4. Frequency distribution of saisonal precipitation trends at individual stations before the homogenization (black shaded) and after the homogenization in (gray).

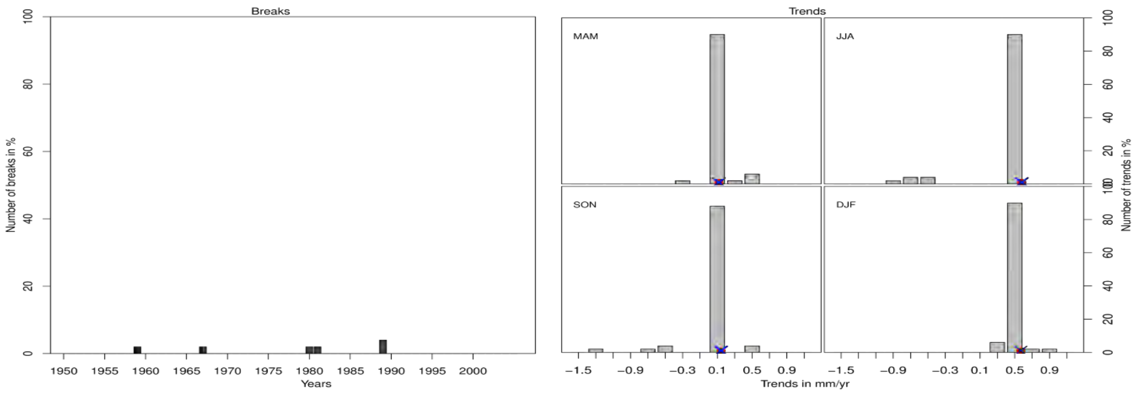
It can be seen that all three stations have very different results. The station Geneve-Cointrin in *Figure 5(a)* shows two trends in the sensitivity study, both in nearly 100% of the cases, that is with all reference series. One break-point is around 1974, the other either around 1978 or more likely around 1979. In HOMPRA Europe time-series a further break-point was discovered in addition to these two breaks, which cannot be confirmed by the sensitivity study. On the right, you can see how the detected breaks affect the trend. In summer the correction is weaker than suggested in the sensitivity study. In spring and winter, there is a clear change to a stronger positive trend in which HOMPRA Europe chooses the weakest confirmed by the sensitivity study. In fall, the sensitivity study shows that the trend results are not as clear as for the other seasons. For the station Twente in *Figure 5b)* the break-point positions are not as defined as for Geneve-Cointrin. Possible break periods are found for the years 1956-1959, 1966-1968, 1973-1977 and 1981-1982. It is therefore very certain that the time series is inhomogeneous, but the break-point positions are very difficult to determine. In this case, verification by metadata would be desirable, but the data is not available. The optimized HOMPRA Europe run only selected two of the break-point periods: 1976 and 1981. These periods, if viewed as a whole, are also most often detected as break-point. Looking at the results of the possible trends, we can see that, compared to the previous station, the trend accuracy remained stable, although the break-point position is more uncertain. However, comparing the results with the original inhomogeneous trend, the reason for the uncertaity is obvious: The changes made to the series are low, and the signal to noise ratio small. Also the HOMPRA Europe trend lies decidedly in the proposed period. Again, the fall trend shows the largest uncertainty. The third station Basel/ Binningen shows a typical example of a homogeneous station. Based on the in-homogeneity of neighboring stations, single break-points are found. Also for the HOMPRA Europe data set this station was classified as homogeneous for the period of 1951-2005.



(a) Geneve-Cointrin (WMO number 6700)



(b) Twente (WMO number 6290)



(c) Basel/ Binningen (WMO number 6601)

Fig. 5. Examples of homogenization with varying neighbor series. For each run, 10 reference series with a correlation of at least 0.8 were randomly selected. On the left the break-point positions and on the right the results of the trends are depicted. Red circles (•) describe the non-homogenized raw data and blue crosses (x) the HOMPRA Europe results.

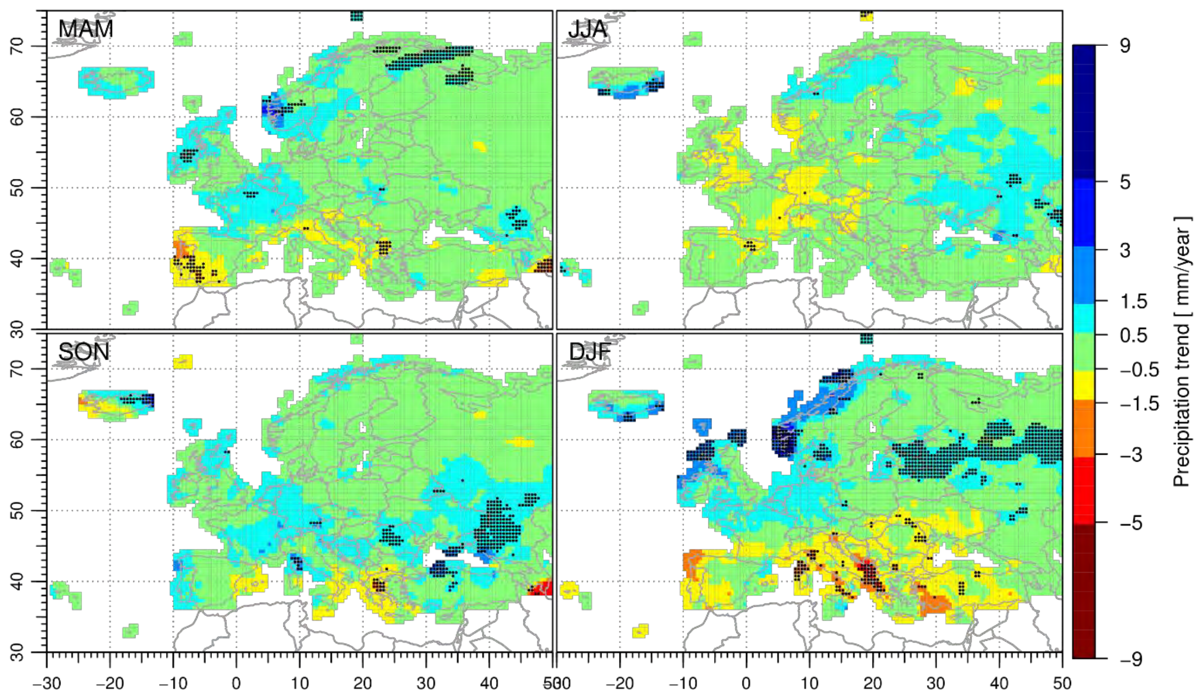


Fig. 6. Seasonal SEN-trends (Sen, 1968) of the final product HOMPRA Europe. Interpolated to a 0.5 degree grid using SPHEREMAP (Becker et al., 2013). Points indicate significant ($p < 0.05$) trends based on the Mann-Kendall test (Yue et al., 2002).

The completed HOMPRA Europe product is available in monthly resolution (10.5676/DWD_GPCC/HOMPRA_EU_M_V1_050). The seasonal SEN-trends (McLeod ,2011) of the final product are shown in *Figure 6*. Significant trends ($p<0.05$) based on the Mann-Kendall test are marked with a point (·). The largest precipitation changes in 1951-2005 are seen in the winter months. The already humid north has a strong positive trend of almost 9 mm/yr, while the south has become significantly drier.

We present the scientific community with HOMPRA Europe a data set that is truly eligible for trend analysis, given its capabilities in data quality and homogeneity. We will try to extend the scope of HOMPRA Europe to a world-wide eddition but the challenges discussed already in terms of the the availability of sufficient stations for detection and in particular correction will be even bigger compared to the data product presented here.

References

Beaulieu, C. Homogénéisation des séries de précipitations: identification des techniques les plus prometteuses et nouveaux développements. Thèse de doctorat, Université du Québec, Institut national de la recherche scientifique, 2009.

Beaulieu, C., T. Ouada, and O. Seidou. Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations= A review of homogenization techniques for climate data and their applicability to precipitation series. *Hydrological sciences journal*, 52(1):18–37, 2007.

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese. A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901–present. Earth System Science Data, 5(1):71–99, 2013. doi: 10.5194/essd-5-71-2013. URL http://www.earth-syst-sci-data.net/5/71/2013/.

Bortz, J. Statistik: Für Human-und Sozialwissenschaftler. Springer, 2004. ISBN 354021271X.

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.

Craddock, J. Methods of comparing annual rainfall records for climatic purposes. *Weather*, 34(9):332—-346, 1979.

Krämer, N., J. Schäfer, and A.-L. Boulesteix. Regularized estimation of large-scale gene association networks using graphical gaussian models. BMC bioinformatics, 10(1):384, 2009.

McLeod. A., Kendall: Kendall rank correlation and Mann-Kendall trend test, 2011. URL https://CRAN.R-project.org/package=Kendall. R package version 2.2.

Mestre, O. Correcting climate series using ANOVA technique. In Proceedings of the Fourth Seminar for Homogenizationand Quality Control in Climatological Databases, Budapest, Hungary, pages 93–96, 2004.

Moberg, A., and H. Alexandersson. Homogenization of Swedish temperature data. Part II: Homogenized gridded air tem-perature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology*,17(1):35–54, 1997.

Picard, F., S. Robin, E. Lebarbier, and J. Daudin. A segmentation/clustering model for the analysis of array CGH data.Biometrics, 63(3):758–766, 2007.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: "GPCC's new land-surface precipitation climatology based on quality-controlled in-situ data and its role in quantifying the global water cycle", *Theor. Appl. Climatology*, 115, 15-40, DOI: 10.1007/s00704-013-0860-x.

Sen, P. K. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324):1379–1389, 1968.

Venema, V. C. K.; O. Mestre, E. Aguilar, I. Auer, J.A. Guijarro, P. Domonkos, G. Vertacnik, T. Szentimrey, P. Stepanek, P. Zahradnicek, et al., Benchmarking homogenization algorithms for monthly data, *Climate of the Past*, 8, 89–115, 2012.

Wilks, D. S. Statistical methods in the atmospheric sciences, volume 100. Academic press, 2006.

Yamamoto, J. K. An alternative measure of the reliability of ordinary kriging estimates, *Mathematical Geology*, 32(4), pp.489-509, 2000.

Yue, S., P. Pilon, and G. Cavadias. Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259(1):254–271, 2002.

HOMOGENIZING GPS INTEGRATED WATER VAPOUR TIME SERIES: METHODOLOGY AND BENCHMARKING THE ALGORITHMS ON SYNTHETIC DATASETS

R. Van Malderen[1], E. Pottiaux[2], A. Klos[3], O. Bock[4], J. Bogusz[3], B. Chimani[5], M. Elias[6], M. Gruszczynska[3], J. Guijarro[7], S. Zengin Kazancı[8] and T. Ning[9]

[1] Royal Meteorological Institute of Belgium, Brussels, roeland.vanmalderen@meteo.be
[2] Royal Observatory of Belgium, Brussels, Belgium
[3] Military University of Technology, Warsaw, Poland
[4] IGN LAREG, University Paris Diderot, Sorbonne Paris, France
[5] Central Institute for Meteorology and Geodynamics, Austria
[6] Research Institute of Geodesy, Topography and Cartography, Czech Republic
[7] AEMET (Spanish Meteorological Agency), Spain
[8] Karadeniz Technical University, Turkey
[9] Lantmäteriet (Swedish Mapping, Cadastre and Land Registration Authority), Sweden

1. MOTIVATION AND INTRODUCTION

Within the COST Action ES1206 "Advanced Global Navigation Satellite Systems tropospheric products for monitoring severe weather events and climate" (GNSS4SWEC), there was a clear interest and need to homogenize Integrated Water Vapour (IWV) datasets retrieved from Global Navigation Satellite System (GNSS) observations, by correcting (artificial) breakpoints due to e.g. instrumental changes. Based on the results of an inquiry, a homogenization activity was started within Working Group 3 ("Use of GNSS tropospheric products for climate monitoring"), targeting the following objectives: (i) select one or two long-term reference datasets, (ii) apply different homogenization algorithms on these reference datasets, and build up a list of commonly identified inhomogeneities based on statistical detection and metadata information, and (iii) come up with an homogenized version of the reference dataset that can be re-used to study climate trends and time variability by the entire community.

As a first reference dataset, we decided to focus on the existing first tropospheric product given by the data reprocessing of the International GNSS Service (IGS) network, named hereafter IGS repro 1. This homogeneous reprocessing (one single strategy) of the data results from a set of 120 GPS (Global Positioning System) stations distributed worldwide providing continuous observations from 1995 until the end of 2010. However, as can be seen in *Figure 1*, the bulk of the sites are located in the northern hemisphere mid-latitudes. Within the IGS network, the metadata of the stations are archived and publicly available in the so-called IGS logfiles. These contain invaluable information about changes in equipment, operating procedures, site conditions, etc. The retrieved Zenith Total Delays (ZTDs) estimated from the GNSS receiver observations at the stations have been screened, and the outliers have been removed as described in *Bock* (2015). To convert those ZTD measurements in IWV, the surface pressure at the station location and a weighted mean temperature are needed. The European Centre for Medium-Range Weather Forecasts (ECMWF) numerical weather prediction model reanalysis ERA-interim (or ERAI, *Dee et al.* 2011) has been used to obtain those auxiliary meteorological parameters (*Bock*, 2016).

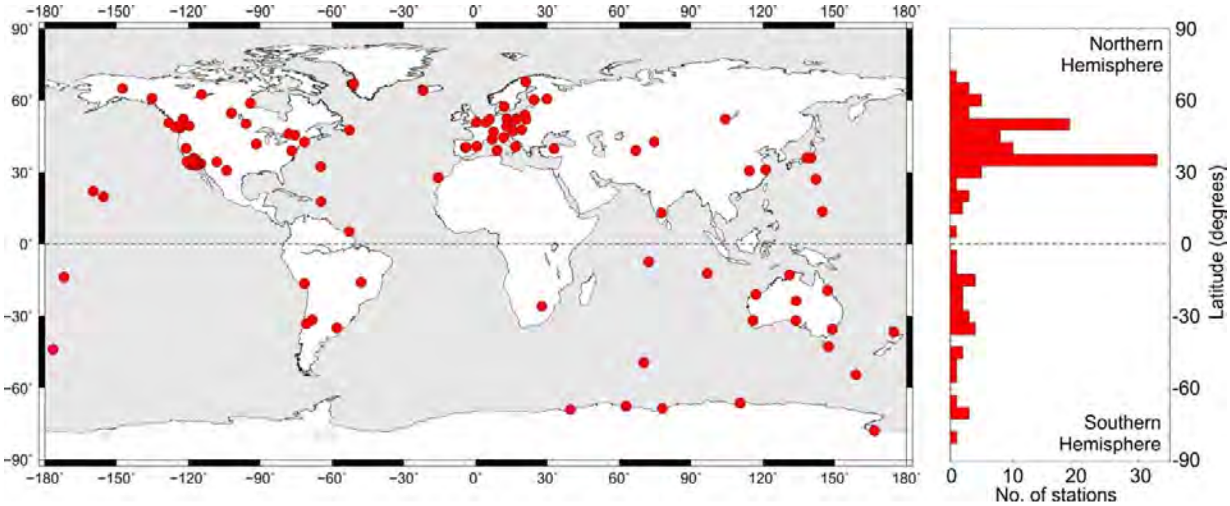


Fig. 1. Distribution of the 120 IGS repro 1 stations with data available from 1995 until the end of 2010.

2. METHODOLOGY

As the distribution of the sites over large areas of the world is rather sparse (see *Figure 1*), the correlations between the IWV time series of our sample sites are rather poor in most areas. As a consequence, the use of neighbouring sites as reference series to remove similar climatic features and to reduce the complexity of the noise characteristics is problematic. Alternatively, various homogenization methods exist that can be used without a reference series (absolute statistical homogenization), but are less reliable (e.g. *Venema et al*. 2012).

Therefore, for a particular GNSS station, we chose to use the ERA-interim IWV time series at this GNSS site location as the reference series for the candidate IGS repro 1 IWV time series. As can be seen in *Figure 2*, for the large majority of the sites, the IGS repro 1 and ERA-interim IWV time series are highly correlated; the lower correlations are ascribed to a bad spatial representation by the model at those sites (e.g. large differences in orography in adjacent pixels). It should however be mentioned that the IGS repro 1 and ERA-interim IWV time series are not completely independent from each other: ERA-interim is used in the ZTD screening process and, as has been noted already above, the surface pressure and weighted mean temperature values, needed for the IGS repro 1 ZTD to IWV conversion, are taken from ERA-interim as well. Another important remark is the fact that ERA-interim might have inhomogeneities of its own, e.g. when new satellite datasets are introduced in the data assimilation system (see e.g. *Schröder et al*. 2016).

Most of the inhomogeneities in the GNSS-derived IWV time series due to antenna or radome changes and changes in the observation statistics (= events) are characterized by jumps in the IWV time series (*Vey et al.*, 2009). Therefore, for each site, we calculate differences time series between the IGS repro 1 and ERA-interim IWV datasets, and we will look for the epochs of those events causing offsets in the difference time series. This approach has already been applied on a similar dataset in *Ning et al.* (2016).

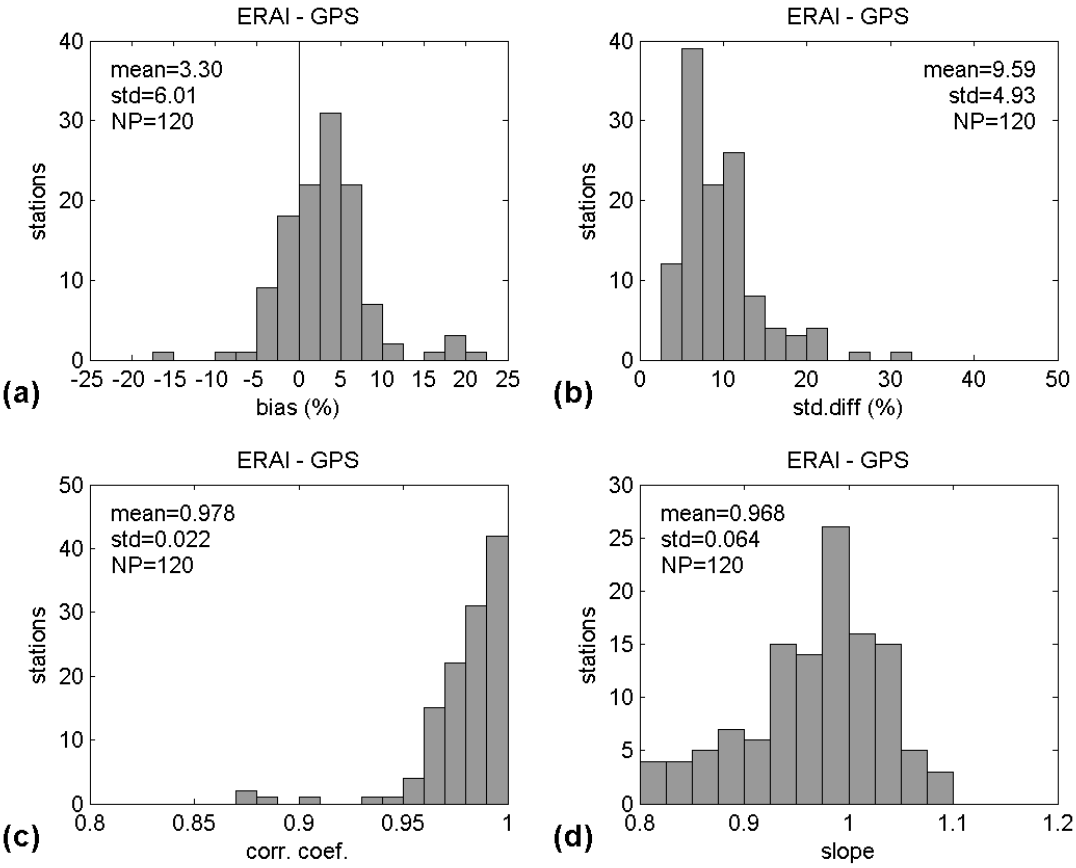


Fig. 2. Histograms of the relative biases (a), relative standard deviations (b), correlation coefficients (c), and linear correlation slope coefficients (d) between the IGS repro 1 and ERA-interim IWV time series for our sample of 120 GNSS stations.

3. SYNTHETIC DATASET GENERATION

We tested different homogenization algorithms on the ERAI-IGS repro 1 differences, and compared their lists of identified epochs of offsets with a list of manually detected breakpoints from the metadata information. At some sites, breakpoints were detected in the metadata and by visual inspection, but not by any of the algorithms. In other cases, breakpoints were detected by a number of (or all) statistical tools, but no metadata information was available for the considered epoch. Therefore, we decided to first generate synthetic time series, with known inserted offsets, on which the different homogenization tools could be blindly applied and assessed. Additionally, we undertook a sensitivity analysis of the performance of the homogenization algorithms on varying characteristics of the synthetic time series.

It should be noted here that we generated synthetic time series of IWV differences directly, based on the characteristics of the real IGS repro 1 and ERA-interim IWV differences. By considering the differences, seasonal variability will be removed and the complexity of the noise will be reduced, making the generation of synthetic time series an easier task. First, we characterized the properties of the offsets (typical number per site and amplitudes) in the real IWV differences, based on the manual detection of 1029 events of instrumental changes, reported in the metadata files of the stations. Of those 1029 events, about 164 epochs were confirmed by visual inspection, and 57 new epochs were added. We derived the amplitudes of the offsets arising at those epochs and these are used for a first-order correction of the real IWV differences at those 221 identified epochs. Subsequently, we analysed the significant frequencies, the noise model, the presence of a linear trend and gaps in those corrected IWV differences with a Maximum Likelihood Estimation (MLE) in the Hector Software (*Bos et al.* 2013). As it is illustrated for the KOSG station in *Figure 3*, we found that the most adapted noise model is given by the combination of white noise (WN) plus autoregressive noise of the first order (AR(1)), characterised by the amplitudes of white noise (with median value 0.35 mm) and autoregressive noise (median value 0.81 mm), the fraction and coefficient of AR(1), with respective median values 0.71 and 0.50. Another important finding is the presence of trends (of the order of ±0.05 kg/m^2 /yr) in the IWV differences series.

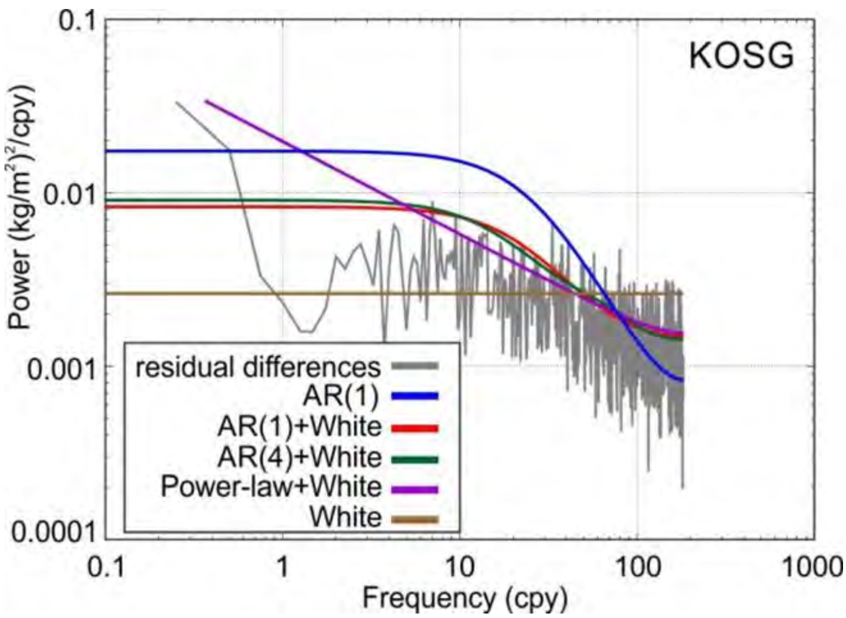


Fig. 3. Power spectrum of the IGS repro 1 and ERA-interim IWV residual differences at the site KOSG (Kootwijk, the Netherlands, 52.18°N, 5.81°E). The colour lines denote different noise models that we tested to decide on the noise model and that best characterize the IWV residuals.

So, based on the characteristics of the IWV differences series at each site separately, we generated for every site a synthetic time series of daily values that includes a number of offsets in the mean. As a matter of fact, to test the sensitivity of the performance of the homogenization tools on the complexity of the time series, 3 datasets of 120 synthetic daily IWV differences time series have been created, with increasing complexity:

- “easy” dataset: includes seasonal signals (annual, semi-annual, ter- and quarter-annual, if present for a particular station) + offsets + white noise (WN)
- “less-complicated” dataset: same as “easy” + autoregressive process of the first order (noise model = AR(1)+WN)
- “fully-complicated” dataset: same as “less-complicated” + trend + gaps (up to 20% of missing data). This dataset is closest to the real IWV differences.

These sets of synthetic time series were made available to the community for a blind testing of homogenization algorithms in use. The inserted offsets of the easy dataset are available to be revealed, if asked for by a participant, for fine-tuning of the algorithm on the use of IWV differences.

4. INVOLVED HOMOGENIZATION ALGORITHMS

In this section, we give a small summary of the homogenization algorithms that participated so far in the blind homogenization of at least one of the variants of the synthetic time series. Those homogenization tools have been applied on daily and/or monthly values of the synthetic datasets.

4.1 Two-sample *t* test (operator: M. Elias)

The procedure applied for the purpose of breakpoint detection is based on hypothesis testing. In this study we used a test statistic that is of so-called "maximum type" (see *Jaruskova*, 1997). Within the field of mathematical statistics the problem can be solved by testing the null hypothesis that claims that there is no change in the distribution of the series, against the alternative hypothesis that claims that the distribution of the series changed at the time k. The null hypothesis is then rejected if at least one of the estimated statistics is larger than the corresponding critical value. Approximate critical values are obtained by the asymptotic distribution (see *Yao and Davis*, 1986). Two ways of time series proceedings and method application were discussed; (i) the proposed method was applied to the uncorrected original series of IGS repro 1 and ERA-interim IWV differences and (ii) the method was applied to corrected difference series when the seasonality was removed and also the gaps were filled in the series before the breakpoint detection, for instance. The method of breakpoint detection is applicable on both monthly and daily time series. A confidence interval for the detected breakpoint is also possible to estimate.

4.2 PMTred (operator: T. Ning)

The rationale of this adapted *t* test is based on *Wang et al.* (2007), which describes this penalized maximal *t* test (PMT) to empirically construct a penalty function that evens out the U-shaped false-alarm distribution over the relative position in the time series. Another modification, named the PMTred test, accounts for the first-order autoregressive noise and it was this test that was used for the homogenization. The critical values (CVs) of the PMTred test were obtained by Monte Carlo simulations running for 1 000 000 times as a function of the sample length N (monthly data, might have to be redone for daily data). In addition, the CVs were calculated for the lag-1 autocorrelation from 0 to 0.95 with an interval of 0.05 and for the confidence levels of 90%, 95%, 99%, and 99.9% (see *Ning et al.,* 2016). This test runs on monthly and daily values, but the critical values are calculated based on monthly data. The detection of multiple breakpoints is achieved by applying the test to the remaining segments.

4.3 HOMOP (operator: B. Chimani)

The homogenization code HOMOP is a combination of PRODIGE (for detection, *Caussinus and Mestre*, 2004), SPLIDHOM (adjustment, *Mestre et al*, 2011), an adapted interpolation (*Vincent et al*., 2002), and improved by some additional plots for facilitating the decision of the homogenisation and extended with some uncertainty information by using different reference stations as well as bootstrapping methods (HOMOP, *Gruber et al* 2009). The approach is neighbour-based, and in the particular case of our synthetic datasets, a lower limit of 0.6 for the correlation coefficients was imposed for selecting potential reference stations. Break detection is done at annual or seasonal base.

4.4 CLIMATOL (operator: J. Guijarro)

Another neighbour-based homogenization algorithm is CLIMATOL, which performs a form of orthogonal regression known as Reduced Major Axis (RMA, *Leduc*, 1987) between the standardized anomalies $(x-\mu x)/\sigma x$ and $(y- \mu y)/\sigma y$ of the two distributions. Orthogonal regression is adjusted by minimizing the perpendicular distance of the scatter points to the regression line, instead of minimizing the vertical distance to that line as in Ordinary Least Squares regression (OLS). In the case of our synthetic datasets, it was imposed that the only reference time series at the site is the ERA-interim time series. The Standard Normal Homogeneity Test (SNHT, *Alexandersson*, 1986) is applied to find shifts in the mean of the anomaly series in two stages. The code incorporates a filling in of missing data and outlier removal. The adjustment of the identified offsets can be done with a varying amplitude: by including e.g. σx in the standardization, you might include seasonality in the amplitudes. As in the other algorithms described so far, the detection of multiple breakpoints is done by applying the test to the remaining segments. CLIMATOL can be applied to any time scale data, but it is advised to detect the breakpoints at the monthly scale, and then use the break dates to adjust the daily series. This algorithm does not provide the amplitudes of breaks, as they are time varying. We might obtain the amplitudes by differencing the non-homogenized series with the homogeneous series.

4.5 Non-parametric tests (operator: R. Van Malderen)

In this case, the used statistical tests are non-parametric distributional tests that utilize the ranks of the time series to find breakpoints (or more general to test the equality of the medians of two distributions). Because such tests are based on ranks, there are not adversely affected by outliers and can be used when the time series has gaps. On the other hand, the significance of the test statistic cannot be evaluated confidently within 10 points of the ends of the time series and those tests show an increased sensitivity to breakpoints in the middle of the time series, when a clear trend is present (*Lanzante*, 1996). We used two of such non-parametric tests: the Mann-Whitney (-Wilcoxon) test and the Pettitt (-Mann-Whitney) test, nicely described in *Lanzante* (1996). As an additional reference, the CUSUM test, based on the sum of the deviations from the mean, is also used. We developed an iterative procedure to detect multiple breakpoints: if 2 out of those 3 tests identify a statistical significant breakpoint, the time series is corrected (by adjustment of the oldest segment with the detected amplitude of the offset) and the 3 tests are applied again on the complete corrected time series. These tests have been applied on both the monthly and daily values.

4.6 Pettitt test (operator: S. Zengin Kazancı)

The Pettitt test (*Petitt*, 1979) has been applied by another operator on the ranks of the daily values, together with the von Neumann ratio (*von Neumann*, 1941) to determine if there is a breakpoint in the time series. If the series is homogeneous, the von Neumann ratio is equal to 2, for lower values of this ratio the series has a breakpoint (*Wijngaard et al*., 2003). The Pettitt test statistic is related to the Mann-Whitney statistic (see above).

5. ASSESSMENT OF THE PERFORMANCE OF THE TOOLS ON THE SYNTHETIC DATASETS

In this section, we will assess the performance of the different homogenization tools on the synthetic datasets on two different aspects: (i) the identification of the epochs of the inserted breakpoints (+ sensitivity analysis) in the synthetic datasets, and (ii) the estimation of the trends that were or were not imposed to the 3 sets of synthetic IWV differences.

5.1 Identification of the breakpoints

To assess whether or not the breakpoint given by a statistical detection tool coincides with the inserted, known, epoch of the break depends on the choice of the time window. Some homogenization algorithms give a confidence interval for the detected breakpoints, but other tools do not. To treat those different methods in a consistent manner, a proper, fixed time window for successful detection has to set. Therefore, we calculated for every homogenization tool the mean time difference between estimated and inserted epochs of offsets (with e.g. an upper limit of half a year). The results for the fully complicated dataset is given in *Figure 4*. From this figure, it can be derived that the here adopted time window of 2 months (62 days) is a good compromise. It should also be noted that the means of the time differences based on daily values (7 left bars) are in line with those calculated from monthly values (6 right bars), and that the mean epoch differences obtained for the easy and less-complicated datasets also have comparable values.

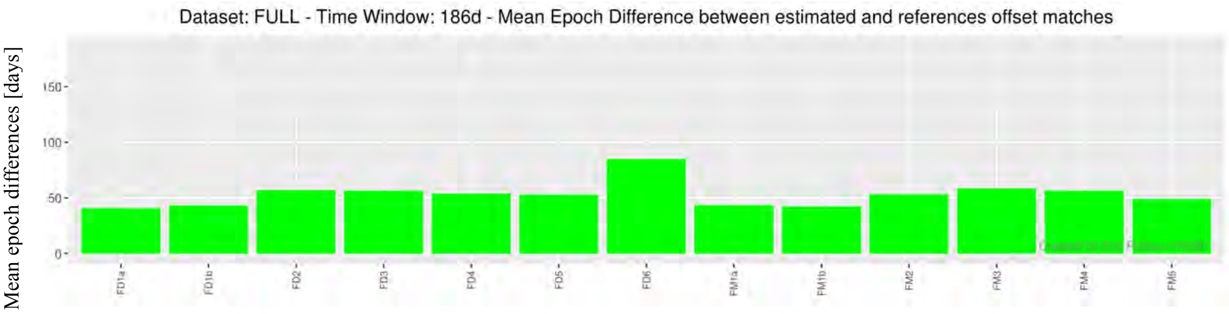
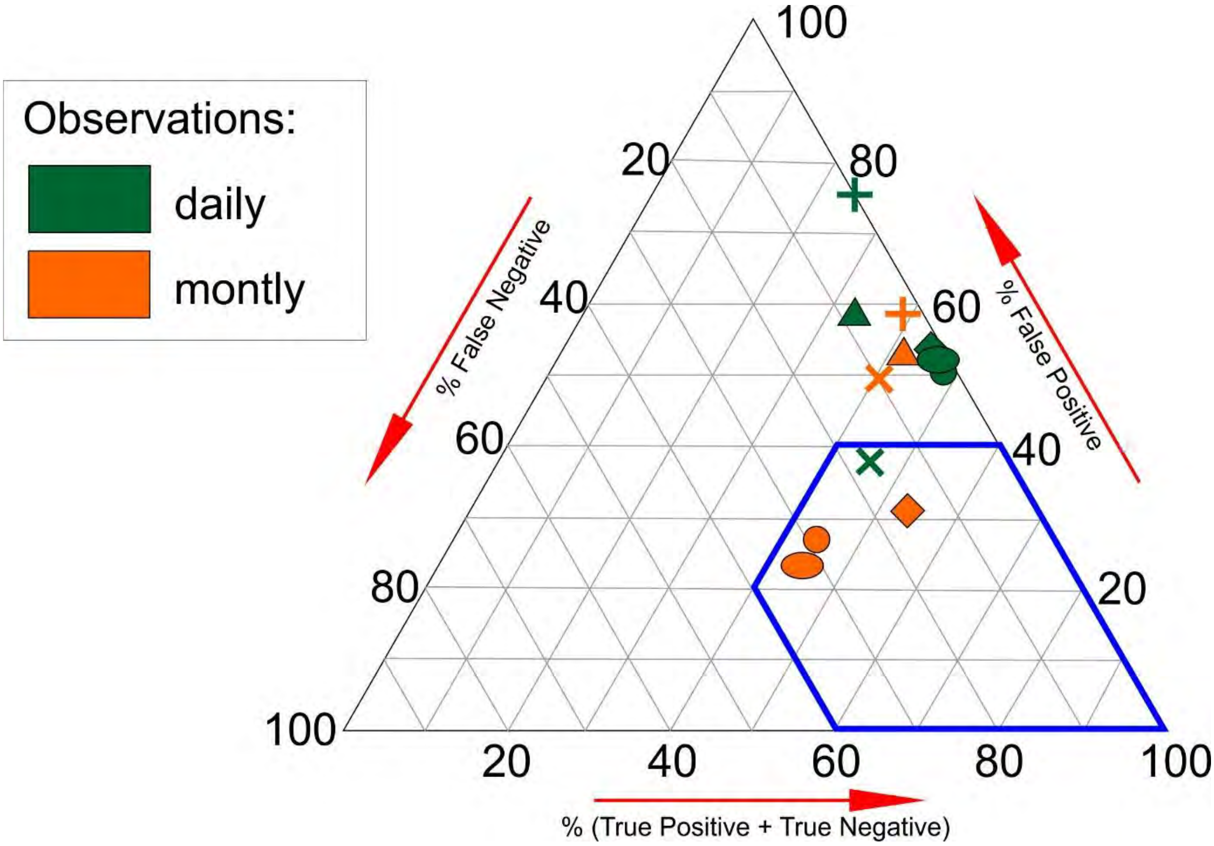


Fig. 4. Means of the time differences between the estimated and inserted epochs of offsets for the fully complicated synthetic dataset, with an upper limit of 186 days (half a year) for the time window. Every bar represents another breakpoint detection solution, with "D" or "M" for application on daily or monthly values respectively, and the numbers referring to the different homogenization tools described in section 4.

With the time window set, we calculate for every breakpoint detection tool the statistical scores: the true positives (TP, "hits"), true negatives (TN: no breaks inserted, no break found), false positives (FP, "false alarms"), and false negatives (FN, "misses"). More details on how to calculate these scores can be found in e.g. *Venema et al.* (2012). To visualize the performance of the different tools in terms of those different statistical scores, we adapted the ternary graph representation from *Gazeaux* et al. (2013), shown in *Figure 5*, again for the fully complicated dataset. It depicts the ratios of the statistical detection scores (TP+TN, FP, and FN) by their position in an equilateral triangle, highlighting the trade-off between those. A perfect solution would appear on the bottom right corner of the triangle (see dashed lines in the figure). From a glance on this figure, it can be directly noted that the involved homogenization tools do not perform very well for the fully complicated dataset: especially the number of false positives are too high. Indeed, when we calculate the probabilities of true detection (POD = TP/(TP+FN)) and the probabilities of false detection (POFD = FP/(FP+TN)), we found high values for the probabilities of false detection (POFD > 0.75 for all but 2 variants of the same method). Fortunately, the probabilities of true detection are also high (POD > 0.75 for two third of the methods). Some methods nearly detect all the inserted breakpoints, but at the cost of a high number of false alarms, while other methods are more conservative in detecting breakpoints, resulting in low scores for both the POD and POFD. It is therefore not surprising that the Pierce Skill Scores, defined as POD – POFD, are very low (around zero and even negative) for the homogenization tools applied on the fully complicated dataset. For comparison: the Pierce Skill Scores lie in the range of 0.10 to 0.63 and 0.04 to 0.22 for the homogenization algorithms that took part in the benchmark for monthly temperature, respectively precipitation time series described in *Venema et al.* (2012).



	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6	Method 7
Symbol	● ⬬	▲	+	×	◆	▼	-
Operator	M. Elias	R. Van Malderen	R. Van Malderen	J. Guijarro	T. Ning	S. Zengin	B.Chimani
Method / SW	2-sample t-test	2 of 3	PMW	CLIMATOL	PMTred	Pettitt	HOMOP
Daily/Monthly	D+M	D+M	D+M	D+M	D+M	D	X
Easy/Less/Full	E+L+F	E+L+F	E+L+F	L+F	E+L+F	E+L+F	E+F

Fig. 5. Ternary graph representing the ratio between three performance measures of the breakpoint detection solutions (TP+TN, FP, and FN). The performance increases with decreasing numbers of false positives and false negatives and increasing numbers of true positives and negatives, so that a perfect solution is located in the lower right corner, marked by the blue area. The different solutions are marked with the symbols and colours outlined in the legend and in the table.

So far, we only discussed the results on the breakpoint detection on the fully complicated dataset. It should however be noted that a good performance of the tools is achieved for the majority of the participating methods on the easy and less complicated datasets, especially due to a lower amount of false positives. As a consequence, the Pierce Skill Scores are now positive for almost all homogenization tools. So, we can conclude that the performance decreases for almost all the tools when adding gaps and a trend in the benchmark time series; adding autoregressive noise of the first order has less impact.

Some of the homogenization algorithms also provided the (constant) amplitudes of the detected offsets. These were compared with the amplitudes of the offsets that were put in the synthetic time series. The result, again for the fully complicated dataset, are shown in *Figure 6*. From this figure, it could be seen that some methods tend to underestimate the number of offsets with small amplitudes relatively (e.g. ME1 and ME2), while other methods on the contrary overestimate the amount of those offsets (e.g. RVM 2of3 D), but on the other hand underestimate the number of offsets with large amplitudes. Clearly, the different methods have a different sensitivity to the amplitudes of the offsets, and some fine-tuning on the statistical thresholds might be advised for some methods. For the other variants of the synthetic datasets, the amplitude distribution of the detected offsets more closely follows the amplitude distribution of the true inserted offsets.

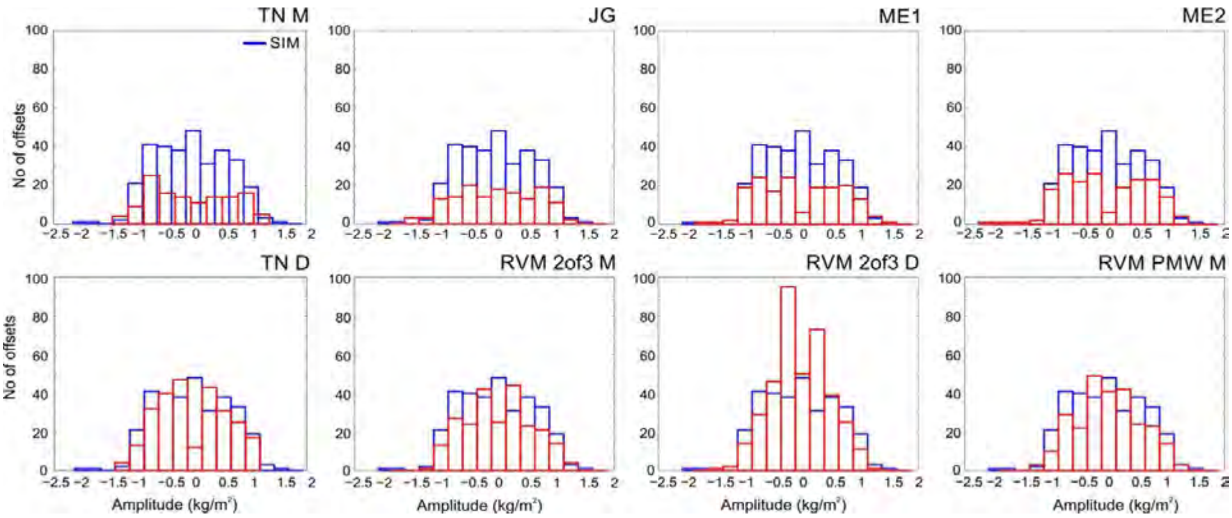


Fig. 6. The histograms of the amplitudes of the detected offsets by the different methods (in red in all figures), compared to the amplitude distribution of the inserted offsets in the fully complicated synthetic dataset of IWV differences (in blue).

5.2 Trend estimation for the homogenized datasets

Only in the fully complicated dataset, a trend was inserted in the IWV differences series, and the homogenized time series by the different time series should hence reveal the same trend. However, as illustrated in *Figure 7* for the easy variant of the generated synthetic IWV differences series for the station CRO1 (Virgin Islands, USA), trends as large as 0.1 kg/m² (or mm) per year arise after correcting for the detected offset by some methods. We stress here that a consistent approach was followed to correct the time series for all detection methods: based on the epoch of the detected offsets, the corresponding amplitude was calculated and corrected for by Maximum Likelihood Estimation (MLE) with the Hector software (*Bos et al.,* 2013). From *Figure 7*, it can be also seen that the calculated trend with this method (denoted with "Hector") deviates around 0.05 kg/m² per year from the true zero trend, due to the uncertainty of the method itself.

The distribution of the differences of trends calculated from the true epochs of the offsets and from the epochs given by different detection methods is given in *Figure 8* for the fully complicated synthetic datasets. The figure shows that most trends differ within ±0.05 kg/m² per year, although those distributions vary between the different detection methods and when applied on either daily (marked with "D") or monthly (marked with "M") values. As one of the mail goals of our homogenization activity is the provision of a homogenized dataset of GNSS IWV time series for use in trend analysis, special care should be taken not to introduce spurious trends in the time series after correction. In this sense, the impact of homogenization on the estimated trend uncertainties should also be further elaborated. The effect of increasing complexity in the synthetic dataset generation (from easy to less and fully complicated) on the estimated trends and their uncertainties is another issue on which additional research is needed.

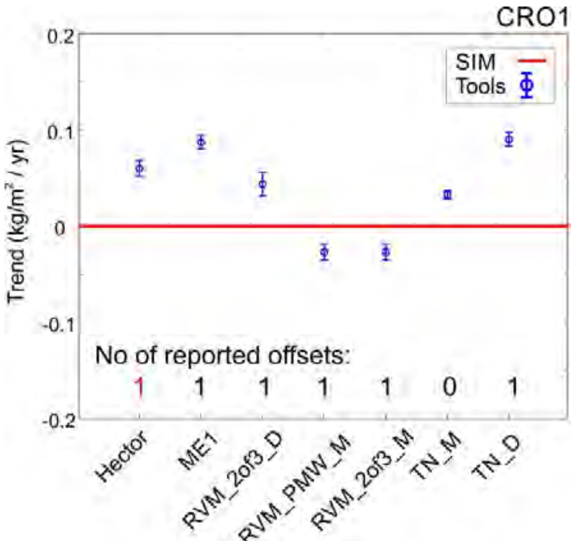


Fig. 7. Trends (in kg/m² /yr) calculated with MLE with the Hector software from the list of identified epochs by the different breakpoint detection tools and from the true epoch of the inserted offset (at the left, marked with "Hector") for the easy variant of synthetic IWV differences for the station CRO1 (Virgin Islands, USA, 17.76°N, 64.58°W).

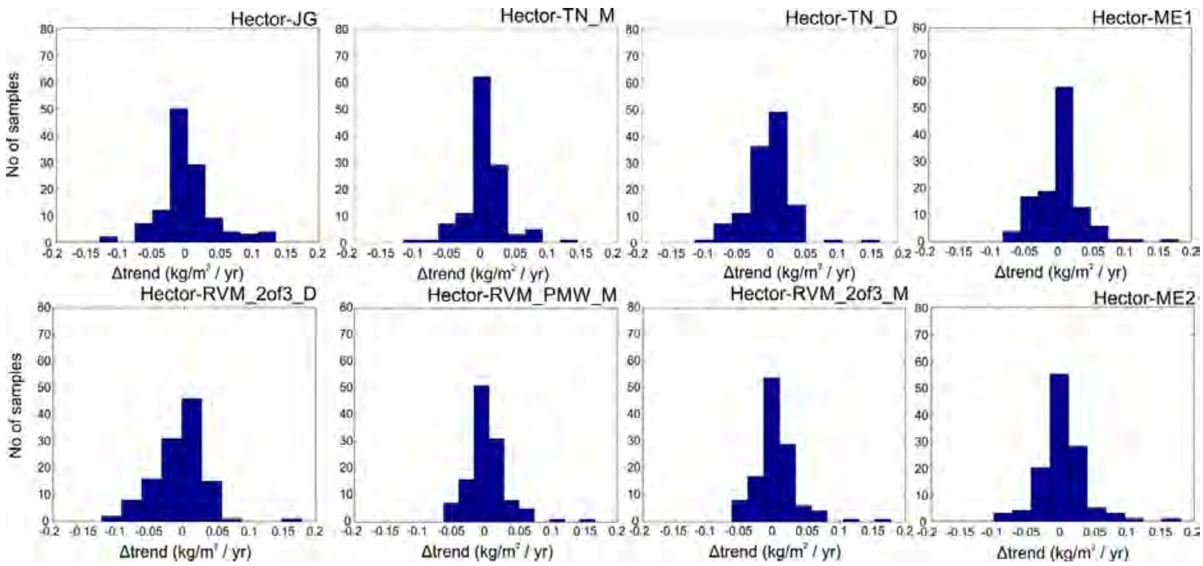


Fig. 8. The histograms of the trend differences calculated for all sites from the fully complicated synthetic time series homogenized either from the true epochs of the offsets or from the epochs given by different detection methods by MLE with the Hector software. Each panel represents another breakpoint detection method.

6. CONCLUSIONS AND OUTLOOK

In this contribution, we described the current activity in homogenizing a world-wide dataset of Integrated Water Vapour (IWV) measurements retrieved from observations made at ground-based GNSS stations. As the distances between those 120 stations are large and correlations are generally low (lower than 0.6 for distances larger than 1°), we used the ERA-interim reanalysis IWV fields at those station locations as the reference time series for relative statistical homogenization. Based on the characteristics of the IWV differences series between the GNSS dataset and ERA-interim and the properties of manually checked instrumental change events reported in the metadata of the GNSS sites, we generated three variants of 120 synthetic IWV difference time series with increasing complexity: we first added autoregressive noise of the first order and subsequently trends and gaps. Those synthetic time series enable us to test the performances of six participating breakpoint detection algorithms and their sensitivity to this increasing dataset complexity.

We found that the performances of those algorithms in identifying the epochs of the inserted offsets especially decreases when adding trends and gaps to the synthetic datasets, due to a larger number of false alarms. On the other hand, the hit rates of most tools are rather good, even when applied on daily values instead of on monthly values. Different tools show a different sensitivity for detecting different ranges of amplitudes of offsets, especially for the most complex (fully complicated) synthetic time series: some tools overestimate (underestimate) the number of small-amplitude (large-amplitude) offsets, while the opposite is true for other breakpoint detection algorithms. After eliminating differences due to different calculation methodologies, we found trend differences mostly within ±0.05 kg/m² per year between the inserted trends and trends calculated from the different homogenization solutions.

Owing to the fact that metadata on instrumental changes are available for the GNSS stations, we primarily focused on the identification of the epochs of offsets until now. At the end, we would like to combine the outcome of statistical breakpoint detection with these metadata. However, we will also assess the performances of the different tools by comparing the final solutions for the time series given by different tools with the original time series (e.g. calculating Centered Root Mean Square errors as in *Venema et al.,* (2012), calculating trends directly from the final solutions, etc.).

Of course, we highly welcome contributions from other groups running homogenization tools, and in the future, our benchmark will already be extended with few more contributions. After providing solutions for the synthetic time series, the participants will get the opportunity to fine-tune their methods on the specifications of the datasets with the help of the knowledge of the true inserted offsets and their amplitudes. Thereafter, a second round of blind homogenization on a newly generated synthetic dataset of IWV values (probably with simulated metadata information) will be held. Based on the performance of the statistical homogenization tools on these synthetic datasets, we will develop a methodology for combining the results of good performing homogenization tools with metadata information. This methodology and those tools will then be applied on the IGS repro 1 dataset of retrieved GPS IWV time series, resulting in a homogenized dataset, which will be validated by other sources of IWV time series and finally made available to the community for assessing the time variability of IWV and for validation of climate model IWV outputs.

Acknowledgements

We would like to thank the COST Action ES1206 GNSS4SWEC for financial support for the two dedicated workshops on homogenization, held in Brussels, Belgium (26-27 April 2016) and in Warsaw, Poland (23-25 January 2017). R. Van Malderen (RMI) and E. Pottiaux (ROB) acknowledge the support from the Solar-Terrestrial Centre of Excellence (STCE).

References

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

Bock, O., “ZTD assessment and screening”, COST ES1206 GNSS4SWEC Workshop, Thessaloniki, Greece, 11-13 May 2015

Bock, O., “Screening and validation of new reprocessed GNSS IWV data in Arctic region”, COST ES1206 GNSS4SWEC Workshop, Potsdam, Germany, 31 August - 2 September 2016

Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., and Bastos, L., 2013, “Fast Error Analysis of Continuous GNSS Observations with Missing Data”, *Journal of Geodesy* 87(4): 351–360, doi:10.1007/s00190-012-0605-0.

Caussinus, H.and Mestre, O., 2004: Detection and correction of artificial shifts in climate series., *Journal of the Royal Statistical Society*, Series C, 53, 405-425

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc*., 137, 553–597, doi: 10.1002/qj.828.

Gazeaux, J., . (2013), Detecting offsets in GPS time series: first results from the Detection of Offsets in GPS Experiment, *J. Geophys. Res. Solid Earth*, 118, 2397–2407, doi:10.1002/jgrb.50152

Gruber C., Auer I., Böhm R., 2009: Endberichte HOM-OP Austria Aufbau und Installierung eines Tools zur operationellen Homogenisierung von Klimadaten mit 3 Annexen.

Jaruskova D., 1997: Some Problems with Application of Change-Point Detection Methods to Environmental Data, *Environmetrics*, Vol. 8, No. 5, Pp. 469-483.

Lanzante, J., 1996, “Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol*., 16, 1197-1226.

Leduc, D.J. 1987. A comparative analysis of the reduced major axis technique of fitting lines to bivariate data. *Can. J. For. Res*. 17: 654–659.

Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S, 2011: SPLIDHOM: A Method for Homogenization of Daily Temperature Observations, *J. Appl. Meteor. Climatol*, 50, 2343-2358.

Ning, T., J. Wickert, Z. Deng, S. Heise, G. Dick, S. Vey, and T. Schöne, 2016: Homogenized Time Series of the Atmospheric Water Vapor Content Obtained from the GNSS Reprocessed Data. *J. Climate*, 29, 2443–2456, https://doi.org/10.1175/JCLI-D-15-0158.1

Pettitt, A.N. (1979): A Nonparametric Approach to the Change-Point Problem, *Applied Statistics*, 28, 126-135, http://dx.doi.org/10.2307/2346729.

Schröder, M., M. Lockhoff, J.M. Forsythe, H.Q. Cronk, T.H. Vonder Haar, and R. Bennartz, 2016: The GEWEX Water Vapor Assessment: Results from Intercomparison, Trend, and Homogeneity Analysis of Total Column Water Vapor. *J. Appl. Meteor. Climatol*., 55, 1633–1649, https://doi.org/10.1175/JAMC-D-15-0304.1

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

Vey, S., R. Dietrich, M. Fritsche, A. Rülke, P. Steigenberger, and M. Rothacher (2009), On the homogeneity and interpretation of precipitable water time series derived from global GPS observations, *J. Geophys. Res.*, 114, D10101, doi:10.1029/2008JD010415.

Vincent, L.A., X. Zhang, B.R. Bonsal, and W.D. Hogg, 2002: Homogenization of Daily Temperatures over Canada. *J. Climate*, 15, 1322–1334, https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2

Von Neumann J., 1941: Distribution of the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics* 13: 367–395.

Wang, X. L., Q. H. Wen, and Y. Wu, 2007: Penalized maximal t test for detecting undocumented mean change in climate data series. *J. Appl. Meteor. Climatol.*, 46, 916–931, doi:10.1175/JAM2504.1.

Wijngaard, J. B., Klein Tank, A. M. G. and Können, G. P. (2003), Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.*, 23: 679–692. doi:10.1002/joc.906

Yao, Y.-C., Davis R. A., 1986: The asymptotic behaviour of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates, Sankhya, 48, 339-353.

NEW DEVELOPMENTS OF INTERPOLATION METHOD MISH: MODELLING OF INTERPOLATION ERROR RMSE, AUTOMATED REAL TIME QUALITY CONTROL

Tamás Szentimrey

Hungarian Meteorological Service
szentimrey.t@met.hu

Abstract

The main difference between MISH and the geostatistical interpolation methods can be found in the amount of information used for modelling the necessary statistical parameters. In general at the geostatistical methods built in GIS the sample for modelling is only the predictors, which is a single realization in time. At MISH method we use the spatiotemporal data for modelling since the long data series form a sample in time and space as well. The long data series is such a specialty of the meteorology that makes possible to model efficiently the statistical parameters in question.

The ongoing developments of the planned new version MISHv2.01 are connected with modelling of climate statistical parameters and the interpolation error moreover a real time data quality procedure will be also built in the system.

At MISH method modelling of the climate statistical parameters is a cornerstone and the interpolation system is based on this one. The earlier modelling system was elaborated for the monthly and daily expected values and the spatial correlations. These are the basic statistical parameters of the interpolation procedures. At the new version the monthly and daily standard deviations and the daily temporal correlations also can be modelled. Consequently the modelling subsystem of MISH will be completed for all the first two spatiotemporal moments. If the joint spatiotemporal probability distribution of the given variable is normal then the above spatiotemporal moments determined uniquely this distribution that is the mathematical model of the climate.

The next development is modelling of the interpolation error RMSE (Root Mean Square Error) in order to characterize quantitatively the uncertainties of the interpolation. This procedure is based on the earlier representativity modelling and the present standard deviation and temporal autocorrelation modelling together.

The last novelty of the new MISH version is an automated real time Quality Control (QC) procedure for observed daily and monthly data. According to the test scheme, the observed values are compared to certain interpolated values using modelled optimal parameters and the modelled interpolation error. During the procedure multiple spatial comparison is tested similarly to the QC procedure built in our MASH (Multiple Analysis of Series for Homogenization; Szentimrey) method. The main difference between the MASH and MISH QC procedures is, while at MASH it is developed for time series and the statistical parameters are estimated from the series in classic statistical way, at MISH it is a real time test and modelled statistical parameters are used.

1. INTRODUCTION

In our conception the meteorological questions and topics cannot be treated separately. Therefore we present a block diagram (*Figure 1*) to illustrate the possible connection between various important meteorological topics. The software MASH (Multiple Analysis of Series for Homogenization; *Szentimrey*, 1999, 2014) and MISH (Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey* and *Bihari*, 2014) were developed by us. These software were applied also in CARPATCLIM project (http://www.carpatclim-eu.org).

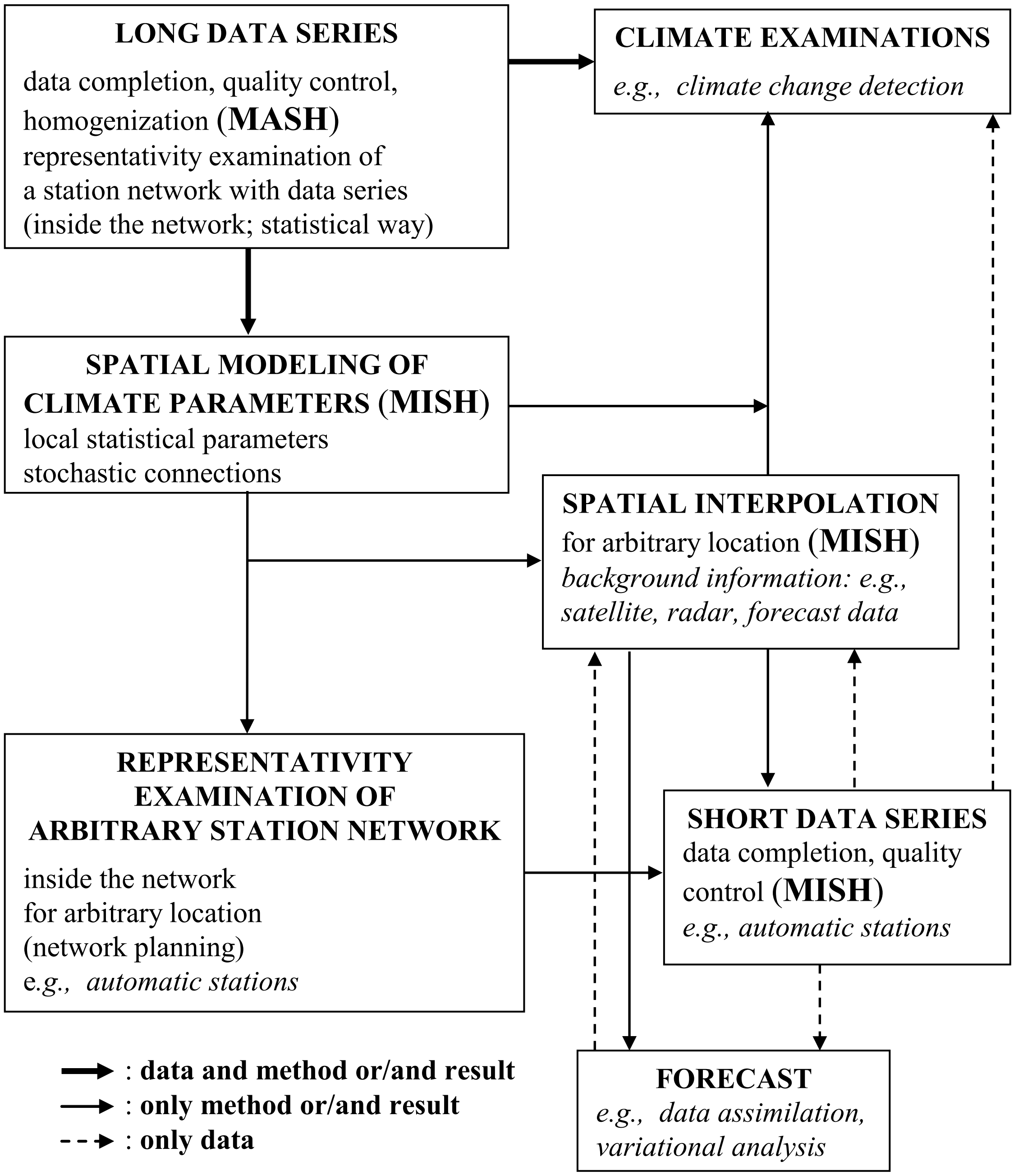


Fig. 1. Block diagram for the possible connection between various basic meteorological topics and systems

2. MATHEMATICAL OVERVIEW OF SPATIAL INTERPOLATION PROBLEM IN METEOROLOGY

According to the interpolation problem the unknown predictand $Z(\mathbf{s}_0,t)$ is estimated by use of the known predictors $Z(\mathbf{s}_i,t)$ $(i=1,...,M)$ where the location vectors $\mathbf{s}$ are the elements of the given space domain D and t is the time.

2.1 Additive model of spatial interpolation

The type of the adequate interpolation formula depends on the probability distribution of the meteorological variable. Assuming normal distribution (e.g. temperature) the additive (linear) formula is adequate.

2.1.1 Statistical parameters

In general the interpolation formulas have some unknown interpolation parameters which are known functions of certain statistical parameters. At the additive interpolation formulas the basic statistical parameters can be divided into two groups such as the local and the stochastic parameters. The local parameters are the expected values $\mathrm{E}(Z(\mathbf{s}_i,t))$ $(i=0,...,M)$. The stochastic parameters are the covariances belonging to the predictand and predictors such as,

$\mathbf{c}$: predictand-predictors covariance vector,

$\mathbf{C}$: predictors-predictors covariance matrix.

This covariance system is equivalent with the standard deviations $D(\mathbf{s}_i)=\mathrm{D}(Z(\mathbf{s}_i,t))$ $(i=0,...,M)$ and the correlation system as,

$\mathbf{r}$: predictand-predictors correlation vector,

$\mathbf{R}$: predictors-predictors correlation matrix.

2.1.2 Linear meteorological model for expected values

At the statistical modeling of the meteorological elements we have to assume that the expected values of the variables are changing in space and in time alike. The spatial change means that the climate is different in the regions. The temporal change is the result of the possible global climate change. Consequently in case of linear modeling of expected values we assume that

$$\mathrm{E}(Z(\mathbf{s}_i,t))=\mu(t)+E(\mathbf{s}_i) \quad (i=0,..,M) \tag{1}$$

where $\mu(t)$ is the temporal trend or the climate change signal and $E(\mathbf{s})$ is the spatial trend.

2.1.3 Additive (Linear) Interpolation Formula

Assuming the linear model (1) the appropriate additive meteorological interpolation formula is as follows,

$$\hat{Z}(\mathbf{s}_0,t)=\lambda_0+\sum_{i=1}^{M}\lambda_i\cdot Z(\mathbf{s}_i,t)$$

where $\sum_{i=1}^{M}\lambda_i=1$ because of unknown $\mu(t)$.

The quality of interpolation can be characterized by the root-mean-square error,

$$RMSE(\mathbf{s}_0)=\sqrt{\mathrm{E}\left(\left(Z(\mathbf{s}_0,t)-\hat{Z}(\mathbf{s}_0,t)\right)^2\right)},$$

and by the representativity value: $REP(\mathbf{s}_0)=1-\dfrac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$.

The optimal interpolation parameters λ_0, λ_i $(i=1,...,M)$ minimize the root-mean-square error and these are known functions of the statistical parameters!

The optimal constant term is: $\lambda_0=\sum_{i=1}^{M}\lambda_i\left(E(\mathbf{s}_0)-E(\mathbf{s}_i)\right)$

The vector of optimal weighting factors $\boldsymbol{\lambda}^{\mathrm{T}}=[\lambda_1,..,\lambda_M]$ written in covariance form,

$$\boldsymbol{\lambda}^{\mathrm{T}}=\left(\mathbf{c}^{\mathrm{T}}+\mathbf{1}^{\mathrm{T}}\frac{\left(1-\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}\right)\mathbf{C}^{-1},$$

and it is known function of the parameters: $D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $(i=1,...,M)$, $\mathbf{r}$, $\mathbf{R}$.

Consequently the unknown statistical parameters are the spatial trend differences $E(\mathbf{s}_0)-E(\mathbf{s}_i)(i=1,...,M)$, the standard deviation ratios $D(\mathbf{s}_0)/D(\mathbf{s}_i)$ $(i=1,...,M)$ and the correlation system $\mathbf{r}$, $\mathbf{R}$. In essence these parameters are climate parameters which in fact means we could interpolate optimally if we knew the climate.

2.1.4 Possibility for modeling of unknown statistical parameters in Meteorology

The special possibility in meteorology is to use the long meteorological data series for modeling of the climate statistical parameters in question. The data series make possible to know the climate in accordance with the fundament of statistical climatology!

The main difference between geostatistics and meteorology can be found in the amount of information being usable for modeling the statistical parameters. In geostatistics the usable information or the sample for modeling is only the actual predictors $Z(\mathbf{s}_i,t)$ $(i=1,...,M)$ which belong to a fixed instant of time, that is a single realization in time. While in meteorology we have spatiotemporal data, namely the long data series which form a sample in time and space as well and make possible to model the climate statistical parameters in question. If the meteorological stations $\mathbf{S}_k$ $(k=1,..,K)$ $(\mathbf{S}\in D)$ have long data series then the spatial trend differences $E(\mathbf{S}_k)-E(\mathbf{S}_l)$ $(k,l=1,...,K)$ as well as the covariances $\mathrm{cov}(Z(\mathbf{S}_k),Z(\mathbf{S}_l))$ $(k,l=1,...,K)$ can be estimated statistically. Consequently these parameters are essentially known and provide much more information for modeling than the predictors $Z(\mathbf{s}_i,t)$ $(i=1,...,M)$ only. However nowadays unfortunately the geostatistical interpolation methods built in GIS software are applied in meteorology mostly.

2.1.5 Relation of daily and monthly data interpolation

Theorem

Let us assume for the daily values within a month the following properties.

i, Expected values and standard deviations:

$$E_t(\mathbf{s}_0)-E_t(\mathbf{s}_i)=e_{0i},\quad D_t(\mathbf{s}_0)/D_t(\mathbf{s}_i)=d_{0i}\quad (i=1,...,M;\,t=1,..,30)$$

ii, Correlations:

$$\operatorname{corr}\left(Z_{t_1}(\mathbf{s}_{i_1}),Z_{t_2}(\mathbf{s}_{i_2})\right)=r^S_{i_1i_2}\cdot r^T_{t_1t_2}\quad (i_1,i_2=1,...,M;\ t_1,t_2=1,..,30)$$

$r^S_{i_1i_2}$: correlation structure in space, $r^T_{t_1t_2}$: correlation structure in time.

Then the optimum interpolation parameters for the daily values and monthly mean are identical: $\lambda_{i,t}=\lambda_{i,month}\ \ (i=0,...,M;\,t=1,..,30)$.

Moreover the representativity values for the daily values and monthly mean are also identical: $REP_t(\mathbf{s}_0)=REP_{month}(\mathbf{s}_0)\ \ (t=1,..,30)$ in the case of optimum interpolation parameters.

2.1.6 Some properties of the optimal interpolation error RMSE

The uncertainties of interpolation can be characterized quantitatively by RMSE that can be expressed as,

$$RMSE(\mathbf{s}_0)=\sqrt{\left(D^2(\mathbf{s}_0)-\mathbf{c}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)+\left(1-\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{c}\right)^2\cdot\frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}}}$$

in the case of optimum interpolation parameters. It can be proved that then the representativity value $REP(\mathbf{s}_0)=1-\dfrac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$ depends on the following parameters only, $D(\mathbf{s}_0)/D(\mathbf{s}_i)\ \ (i=1,...,M)$, $\mathbf{r}$, $\mathbf{R}$.

Moreover if $D(\mathbf{s}_0)/D(\mathbf{s}_i)=1\ \ (i=1,...,M)$ then,

$$REP(\mathbf{s}_0)=1-\sqrt{\left(1-\mathbf{r}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}\right)+\left(1-\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{r}\right)^2\cdot\frac{1}{\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{1}}}$$

3. DEVELOPMENTS IN MISH SYSTEM

Our method MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) for the spatial interpolation of surface meteorological elements was developed (*Szentimrey* and *Bihari*, 2007, 2014) according to the mathematical background that is outlined in Section 2. This is a meteorological system not only in respect of the aim but in respect of the tools as well. It means that using all the valuable meteorological information – e.g. climate and possible background information – is required.

3.1 The basic modelling system in MISH

3.1.1 Modelling of monthly statistical parameters in MISH

i, Spatial expected values (spatial trend) $E(\mathbf{s})$

ii, Spatial standard deviations $D(\mathbf{s})$

iii, Spatial correlations $r(\mathbf{s}_1, \mathbf{s}_2)$

Remark: Support program ANOVA (Analysis Of Variance) for modelling part, in order to evaluate the modelling results (see 3.1.3).

3.1.2 Interpolation applications for monthly and daily data

$$\hat{Z}(\mathbf{s}_0) = \lambda_0 + \sum_{i=1}^{M} \lambda_i \cdot Z(\mathbf{s}_i) \ , \quad REP(\mathbf{s}_0) = 1 - \frac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$$

The optimum interpolation parameters $\lambda_i \ (i = 0,...,M)$ and representativity value $REP(\mathbf{s}_0)$ can be calculated from the above modelled monthly parameters.

3.1.3 ANOVA (Analysis Of Variance) examination

Notations

$Z(\mathbf{s}_i, t) \ (i = 1,..,M; \ t = 1,...,n)$ – monthly station data series $(\mathbf{s}_i : \text{location}; t : \text{time})$

$\hat{E}(\mathbf{s}_i) \ (i = 1,..,M)$ – sample mean at station $\mathbf{s}_i$

$\hat{D}^2(\mathbf{s}_i) \ (i = 1,..,M)$ – sample variance at station $\mathbf{s}_i$

$\hat{E}$ – total sample mean

$\hat{V}$ – total sample variance

Partitioning of Total Variance of station data series

$$\hat{V} = \frac{1}{M \cdot n} \sum_{i=1}^{M} \sum_{t=1}^{n} \left(Z(\mathbf{s}_i, t) - \hat{E} \right)^2 = \frac{1}{M} \sum_{i=1}^{M} \left(\hat{E}(\mathbf{s}_i) - \hat{E} \right)^2 + \frac{1}{M} \sum_{i=1}^{M} \hat{D}^2(\mathbf{s}_i) = \hat{S}^2_{space} + \hat{D}^2_{time}$$

$\hat{S}^2_{space}$ is the variance of spatial trend, $\hat{D}^2_{time}$ is the mean temporal variance.

CARPATCLIM ANOVA results for mean maximum temperature per months in Hungary

	1	2	3	4	5	6	7	8	9	10	11	12
Tx												
D_t:	2.67	3.24	2.69	1.87	1.96	1.64	1.71	1.98	1.96	1.83	2.43	2.11
S_s:	1.00	1.23	1.33	1.21	1.31	1.34	1.37	1.39	1.43	1.34	1.21	1.02
Tn												
D_t:	2.76	2.88	1.86	1.35	1.20	1.12	1.21	1.18	1.29	1.67	1.97	2.12
S_s:	0.85	0.85	0.83	0.88	0.91	0.87	0.90	0.88	0.82	0.77	0.70	0.80
R												
D_t:	22.5	22.9	21.3	25.6	36.2	39.0	39.3	40.7	36.5	35.7	33.3	27.9
S_s:	7.1	5.9	6.9	7.8	7.8	8.9	9.3	10.5	10.2	8.3	10.8	8.6

3.2 Some ongoing developments in MISH

3.2.1 New modelling parts in MISH

iv, Modelling of temporal daily autocorrelations $\rho(\mathbf{s})$ per months.

v, Modelling of daily standard deviations $D_{daily}(\mathbf{s})$ per months.

This development is based on the modelled autocorrelation $\rho(\mathbf{s})$. Let us assume the daily data of a given month constitute an AR(1) process with common standard deviation $D_{daily}(\mathbf{s})$ and temporal autocorrelation $\rho(\mathbf{s})$. Then $D_{daily}(\mathbf{s})$ can be estimated by using the monthly standard deviation $D_{month}(\mathbf{s})$:

$$D_{daily}(\mathbf{s}) \approx \sqrt{30 \cdot \frac{1-\rho}{1+\rho}} \cdot D_{month}(\mathbf{s})$$

Consequently the first two spatiotemporal moments can be modelled for daily and monthly data by the MISH procedure!

3.2.2 Modelling of Present Climate by MISH

Example

Mean temperature data in September for 10 arbitrary locations somewhere in Hungary.

Input: the location coordinates only without any temperature data.

Output: modelled climate statistical parameters

Location indices:

	1	2	3	4	5	6	7	8	9	10
Monthly Expected Values:	14.59	14.99	14.95	15.06	15.16	15.16	15.13	15.08	15.01	15.05
Daily Expected Values:	14.59	14.99	14.95	15.06	15.16	15.16	15.13	15.08	15.01	15.05
Monthly Standard Deviations:	1.34	1.62	1.68	1.67	1.68	1.66	1.72	1.66	1.61	1.64
Daily Standard Deviations:	2.84	3.44	3.47	3.46	3.47	3.60	3.73	3.58	3.48	3.46
Temporal Daily Autocorrelations:	0.74	0.74	0.75	0.75	0.75	0.73	0.73	0.73	0.73	0.74

Matrix of Spatial Autocorrelations:

1.00	0.99	0.99	0.98	0.97	0.96	0.97	0.97	0.98	0.98
0.99	1.00	0.99	0.99	0.98	0.95	0.96	0.96	0.97	0.98
0.99	0.99	1.00	0.99	0.99	0.94	0.95	0.95	0.96	0.97
0.98	0.99	0.99	1.00	0.99	0.91	0.93	0.93	0.95	0.96
0.97	0.98	0.99	0.99	1.00	0.90	0.91	0.91	0.93	0.94
0.96	0.95	0.94	0.91	0.90	1.00	0.99	0.99	0.98	0.98
0.97	0.96	0.95	0.93	0.91	0.99	1.00	0.99	0.99	0.98
0.97	0.96	0.95	0.93	0.91	0.99	0.99	1.00	0.99	0.99
0.98	0.97	0.96	0.95	0.93	0.98	0.99	0.99	1.00	0.99
0.98	0.98	0.97	0.96	0.94	0.98	0.98	0.99	0.99	1.00

3.2.3 Modelling of interpolation error RMSE

The modelled representativity value is the same for monthly and daily data:

$$REP(\mathbf{s}_0) = 1 - \frac{RMSE(\mathbf{s}_0)}{D(\mathbf{s}_0)}$$

Consequently the modelled RMSE can be calculated as,

i, for monthly data: $RMSE_{month}(\mathbf{s}_0) = D_{month}(\mathbf{s}_0)\cdot(1 - REP(\mathbf{s}_0))$

ii, for daily data: $RMSE_{daily}(\mathbf{s}_0) = D_{daily}(\mathbf{s}_0)\cdot(1 - REP(\mathbf{s}_0))$

```
Example MISH output: Predictand location: 17.60  47.40

Predictor Indexes :   12   13   10   30   16   11   14
 Weighting Factors: 0.20 0.07 0.20 0.18 0.11 0.13 0.11

Interpolation without Background Information:
 Predictand Value:     14.77
 Representativity:     0.814    RMSE:    0.57

Interpolation with Background Information:
 Predictand Value:     14.74
 Representativity:     0.822    RMSE:    0.54
```

3.2.4 Automated real time Quality Control for daily and monthly data

Test schema of QC procedure at additive, normal model is: $\dfrac{Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)}{RMSE(\mathbf{s}_0)} \in N(0,1)$,

where $Z(\mathbf{s}_0)$ is the predictand to be controlled, $\hat{Z}(\mathbf{s}_0)$ is the interpolated value using modelled optimal parameters and $RMSE(\mathbf{s}_0)$ is the modelled interpolation error. During the procedure multiple spatial comparison is tested similarly to the QC procedure built in our MASH method for station data series.

3.3 Multiplicative model of spatial interpolation

In this paper only the linear or additive model was described in detail which is appropriate in case of normal probability distribution. However perhaps it is worthwhile to remark that for case of a quasi lognormal distribution (e.g. precipitation sum) we deduced a mixed additive multiplicative formula which is used also in our MISH system and it can be written in the following form,

$$\hat{Z}(\mathbf{s}_0,t)=\vartheta\cdot\left(\prod_{q_i\cdot Z(\mathbf{s}_i,t)\geq\vartheta}\left(\frac{q_i\cdot Z(\mathbf{s}_i,t)}{\vartheta}\right)^{\lambda_i}\right)\cdot\left(\sum_{q_i\cdot Z(\mathbf{s}_i,t)\geq\vartheta}\lambda_i+\sum_{q_i\cdot Z(\mathbf{s}_i,t)<\vartheta}\lambda_i\cdot\left(\frac{q_i\cdot Z(\mathbf{s}_i,t)}{\vartheta}\right)\right)$$

where the interpolation parameters are $\vartheta>0$, $q_i>0$, $\lambda_i\geq 0$ $(i=1,\dots,M)$ and $\sum_{i=1}^{M}\lambda_i=1$.

4. SOFTWARE MISH

4.1 Version MISHv1.03

The last software version MISHv1.03 consists of two units that are the modeling and the interpolation systems. The interpolation system can be operated on the results of the modeling system. We summarize briefly the most important facts about these two units of the developed software.

Modeling subsystem for climate statistical (local and stochastic) parameters:

- Based on long homogenized data series and supplementary deterministic model variables. The model variables may be such as height, topography, distance from the sea etc.. Neighbourhood modeling, correlation model for each grid point.
- Benchmark study, cross-validation test for interpolation error or representativity.
- Modeling procedure must be executed only once before the interpolation applications!

Interpolation subsystem:

- Additive (e.g. temperature) or multiplicative (e.g. precipitation) model and interpolation formula can be used depending on the climate elements.
- Daily, monthly values and many years' means can be interpolated.
- Few predictors are also sufficient for the interpolation and no problem if the greater part of daily precipitation predictors is equal to 0.
- The representativity is modeled too.
- Capability for application of supplementary background information (stochastic variables) e.g. satellite, radar, forecast data.
- Data series completion that is missing value interpolation, completion for monthly or daily station data series.
- Interpolation, gridding of monthly or daily station data series for given predictand locations. In case of gridding the predictand locations are the nodes of a relatively dense grid.

Our MISH-MASH software can be downloaded from:

http://www.met.hu/en/omsz/rendezvenyek/homogenizationand_interpolation/software/

4.2 The planned developments for the new version MISHv2.01

Modeling subsystem for climate statistical (local and stochastic) parameters:

– Modeling of all the first two spatiotemporal moments for daily and monthly data.

Interpolation subsystem:

– The expected interpolation error RMSE is modelled too.

– Real time Quality Control for daily and monthly data.

References

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., Bihari, Z., 2007: Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis), Proceedings of the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2004, COST Action 719, COST Office, 2007, pp. 17-27

Szentimrey, T, Bihari, Z., Lakatos, M., Szalai,S., 2011: Mathematical, methodological questions concerning the spatial interpolation of climate elements. Proceedings of the Second Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2009, *Időjárás* 115, 1-2, 1-11

Szentimrey, T. 2013: Theoretical questions of daily data homogenization, *Időjárás* Vol. 117. No. 1, January-March 2013. pp. 113-122.

Szentimrey,T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60.

Szentimrey,T.,2014: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p. 69.

Szentimrey T., Bihari Z., Lakatos M., 2015: Mathematical questions of spatial interpolation of climate variables, Proceedings of the 8th Seminar for Homogenization and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation Techniques in Climatology and Meteorology, Budapest, Hungary, 2014, WCDMP-No. 84, pp. 107-114

COMPARISON OF DIFFERENT INTERPOLATION METHODS FOR THE GENERATION OF A CLIMATOLOGY OF MAXIMUM AND MINIMUM MONTHLY TEMPERATURES IN SPANISH MAINLAND

Dhais Peña-Angulo [1-2], Celia Salinas Solé [1-2], Azucena Jiménez Castañeda [1], Marcos Rodrigues [1-2], Michele Brunetti [3], Santiago Beguería[4], Sergio Vicente[5], José Carlos González-Hidalgo [1-2]

1. Department of Geography, Zaragoza University, Zaragoza, Spain
2. Institute University of Research in Sciences Environmental (IUCA), University of Zaragoza, Zaragoza, Spain
3. Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy
4. Spanish National Research Council (CSIC), Campus de Aula Dei, Zaragoza, Spain
5. Instituto Pirenáico de Ecología, Spanish National Research Council (IPE-CSIC), Campus de Aula Dei, Zaragoza, Spain

dhaispa@gmail.com, cs@unizar.es, geoazu.flysch@gmail.com, rmarcos@unizar.es, m.brunetti@isac.cnr.it, jcgh@unizar.es, santiago.begueria@csic.es, svicen@ipe.csic.es

Abstract

In the present study we present an exhaustive comparison of the three interpolation methods that recently have been used to develop a high resolution climatology of monthly mean values of temperatures (maximum and minimum temperature) in Spanish mainland. The methods were different in characteristics and variables used (Locally Weighted Linear Regression (LWLR), Regression-Kriging (RK) and Regression-Kriging with Stepwise selection (RKS)), and all of them were computed using the information from archives of the Spanish State Meteorological Agency (AEMet).

Initially, the best interpolation method should be able to reduce the bias in the estimation, having overall low absolute error and small sensitivity to the spatial structure of the input observatories and the altitudinal gradient. Thus, we have studied the estimation errors of the three interpolation methods, paying special attention to the spatial distribution of the error, its structure, and the relationship between error/altitude. Compared to the remaining tested methods, LWLR provides the best results both in terms of the obtained spatial (altitudinal ranges) and temporal variability (monthly variation). The error in the estimation in LWRL is minimum, showing no clear relationship either with elevation or error bias.

Keywords: interpolation, error, spatial distribution, altitude

1. INTRODUCTION

Climatology maps are a key resource in several fields such as agriculture, hydrology, ecology, natural resource management (*Daly et al.*, 2008), and also in climate science itself. For instance, in order to investigate the spatial-temporal behavior of climate signal a trustily high quality climatology is required (*Hofstra et al.* 2008). Climate maps have been traditionally developed applying spatial interpolation techniques to records from weather stations, and quality and reliability of any climatology strongly depends on data availability (i.e. number of stations and records), their spatial distribution (location and spatial pattern of the weather stations) and the interpolation method.

The punctual nature of the climatic data makes the information not be continuous in space which is often not sufficient to properly address climate patterns or integrate climate information in other fields (*Jones and Hulme* 1996, *Dai et al.* 1997, *New et al.* 2000). The spatial distribution of the observatories tends to be biased towards higher density in low altitude and agricultural areas (i.e. populated areas), whereas in mountain areas the density is considerably poorer. To overcome this limitation different types of interpolation have been explored, reaching different conclusions on what method performs best, as a function of the area where they are applied on (*Vicente-Serrano et al.* 2003).

Previous works by *Gonzalez-Hidalgo et al.* (2015a) and *Peña-Angulo* (2016a) developed a new climatology of maximum (Tmax) and minimum (Tmin) temperature in Spain (1950-2010) using the MOTEDAS dataset (*Gonzalez-Hidalgo et al.* 2015b). *Peña-Angulo et al.* (2016a) improved temperature modelling by exploring three interpolation methods: (1) Locally Weighted Linear Regression (LWLR), (2) Regression-Kriging (RK) and (3) Regression-Kriging with Stepwise selection (RKS). These methods were preliminarily compared using several global validation procedures based on the Mean Bias Error (BIAS), the Root Mean Squared Error (RMSE) and the Index of Agreement (*Willmott*, 1982). The performances of the three interpolation models were evaluated by using a complete leave-one-out cross-validation procedure by which each monthly value from each station was excluded from the dataset and reconstructed by the three models, using all the other stations; finally, the estimated value was compared with the observed value. This procedure ensured a higher level of accuracy with respect to the classic approach of leaving a fixed percentage of original data for the validation procedure. According to *Peña-Angulo et al.* (2016a), the LWLR approach over-performed RK and RKS. Nevertheless, the most critical areas in terms of prediction error are mountain ranges in summer for Tmax, while Tmin seems to be more sensitive to station density. These findings suggest differences in the performance of the model both on a temporal and spatial basis which, in turn, depends on the temperature metric.

In the present study, we will show the study errors of estimation of the three interpolation methods to provide a more comprehensive comparison, paying special attention to the spatial distribution of the error, its structure, and the relationship between error and altitude. The core of our proposal revolves around the fact that poorer performances are expected at higher the altitude, and that differences in the error may be found between Tmax and Tmin. In addition, prediction error might vary depending on error bias, i.e. whether the prediction over or underestimates the actual record. The three proposed methods are compared taking into account several altitude ranges, also according to prediction bias. Finally, the spatial structure of the error is explored by means of spatial autocorrelation methods and cluster and outlier analysis. In this way, we would expect to be able to identify significant spatial clusters of error in the prediction.

2. DATASET AND METHODS

2.1. Study area and Data

The climate in Spanish mainland is influenced by its position in the area of subtropical transition in the European western facade (*Lionello*, 2012) (see *Figure 1 a and b*), where the temperature distribution exhibits variations in the north–south directions, inland-coastland and altitude effects caused by: (1) the north–south distance (circa 1000 km) (*Peña-Angulo*, 2016b); (2) Its position between two contrasting characteristics masses of water (Atlantic Ocean and Mediterranean Sea); (3) distribution its main mountain systems, with east–west orientations, dividing the space into three major climatic areas: i) the north coast; ii) the mid-west regions spanning down to the south coast; and iii) the Mediterranean coast (*Font Tullot*, 1983; *Martín and Olcina*, 2001). The large interior space is divided into two large units (the North and South Plateaus) with high elevations (values above 600 m over the sea level in the Northern Plateau and around 400 m over the sea level in the Southern Plateau) adding a new source of variability (*Peña-Angulo*, 2016b).

These conditions create an ideal situation to analyze and evaluate the performance of different interpolation methods, in which local multivariate regression models could be more suitable than global methods to estimate the spatial gradients of temperatures.

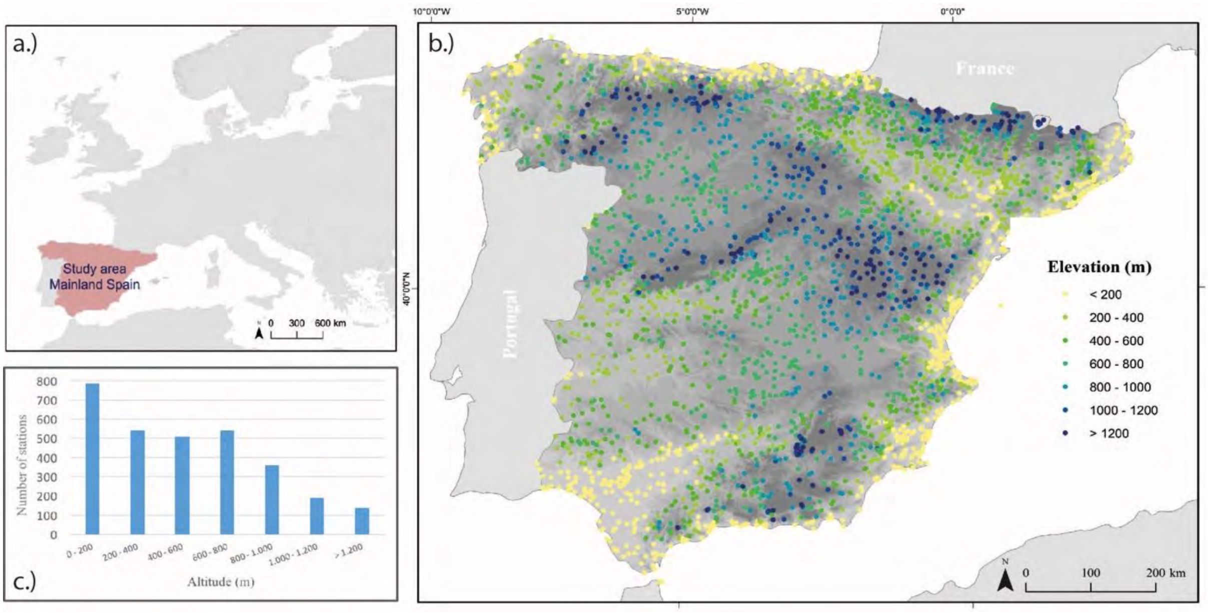


Fig. 1. a.) Study area; b.) Spatial distribution of the MOTEDAS database stations by altitudinal intervals in the Spanish mainland. c.) Number of stations by altitudinal intervals

In the research we have used the most recently updated database of monthly temperatures, the MOTEDAS dataset. It consist of a high-resolution grid (10 x 10 km) from 1951-2010. The development of this database is described in *Gonzalez-Hidalgo et al.* (2015a) and in *Peña-Angulo et al.* (2016a), and basically was developed after exhaustive quality control and reconstruction of series prior to gridded (see *Figure 1b*). As was expected, the spatial and altitudinal distribution of stations was slightly slanted towards low lands (*Figure 1 b and c*).

2.2. Methods

We compared the following interpolating methods: (1) Locally Weighted Linear Regression (LWLR), (2) Regression-Kriging (RK) and (3) Regression-Kriging with Stepwise selection (RKS) of the maximum and minimum temperature of the climatology (1950-2010) for Spanish mainland (*Peña Angulo et al.* 2016a).

The first method studied was the Local Weighted Linear Regression (LWLR). It estimates locally the relationship between temperature and elevation (*Brunetti et al*. 2014) separately for each grid cell of the Digital Elevation Model (DEM), giving more importance to any nearby station with topographical characteristics similar to those of the grid cell itself. For its calculation it will consider at least 15, and no more than 35, stations being 35 the number of neighboring stations that minimizes the error. Finally, for each grid point, a number of neighboring stations with the highest weights were used to estimate the regression. These weighting factors were: coordinates, height, distance from the sea, slope steepness and slope orientation, and they were calculated by using a Gaussian function.

The second methods analyzed was the Regression-Kriging (RK), which combines a regression model with a kriging of the regressions residuals (*Tveito et al.* 2008, *Di Piazza et al.* 2011, *Brunetti et al.* 2014). Then firstly we estimated the relationship between temperature and elevation, and then a Kriging interpolation was applied to the residuals from this model. This technique can be used to obtain a variogram that provides information on the spatial correlation of the analyzed residuals. In this case, we took into account all pairs of stations in the range of 250 km, and grouped them according to distance intervals of 10 km.

Finally, we studied the Regression-Kriging with Stepwise selection (SRK). The method is in fact a variation of the previous one, but in this case the kriging is used to interpolate the residuals from a multilinear regression model with stepwise selection. The stepwise selection method allows us to choose the optimum independent variables that will be used in the multilinear regression model for each month. Therefore, with this method we integrate the variables in an iterative way. Thus, in each step it evaluates which set of variables should be included in the model, and the algorithm stops when the model does not make any further improvements (either by introducing or by removing variables). As in the previous method, a Kriging interpolation was applied to the residuals from the multi-linear regression. And also all the pairs of stations in the range of 250km were taken and grouped according to distance intervals of 10km. finally, we selected the exponential variogram to model the dependency between the semivariance and the distance (*Peña Angulo et al.* 2016a).

In the previous research *Peña et al.* (2016a) assessed the effect of altitude in prediction performance on a monthly basis using the BIAS (*Pielke*, 1984). This indicator of error provided information on the tendency of the model to overestimate or underestimate a variable accordingly elevation (intervals of 200 m until 1200 m), thermometric measurements and models. In the present study we conduct a similar analysis but using the MEAN Absolute Error (MAE) that we used to measure the absolute difference between values predicted by the model and observed values.

Finally, to support the findings from MAE versus elevation, we calculate the coefficient of determination of residue against elevation (RSQR) (*Steel and Torrie*, 1969). This measure is used to identify the quality of the model, its ability to predict, and evaluate variability of the results that model can explain at monthly, thermometric measurements and accordingly altitudinal intervals. In brief, it provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. Again, we split stations according to elevation and over/under estimation. Normality in the residuals was tested using the Shapiro-Wilk test.

The existence of spatial autocorrelation in the residuals is tested using the Moran's I. We analyze changes in the spatial pattern of the residue by means of cluster and outlier analysis using the Anselin's Local Moran's I (*Anselin*, 1995) using a 35 K-Nearest Neighbors. This is done solely for LWLR to provide further insights into the spatial behavior the method.

3. RESULTS AND DISCUSSION

We analyzed the spatial distribution of annual BIAS is presented in *Figure 2*. The overall result are very similar and apparently there is not special differences in the spatial distribution.

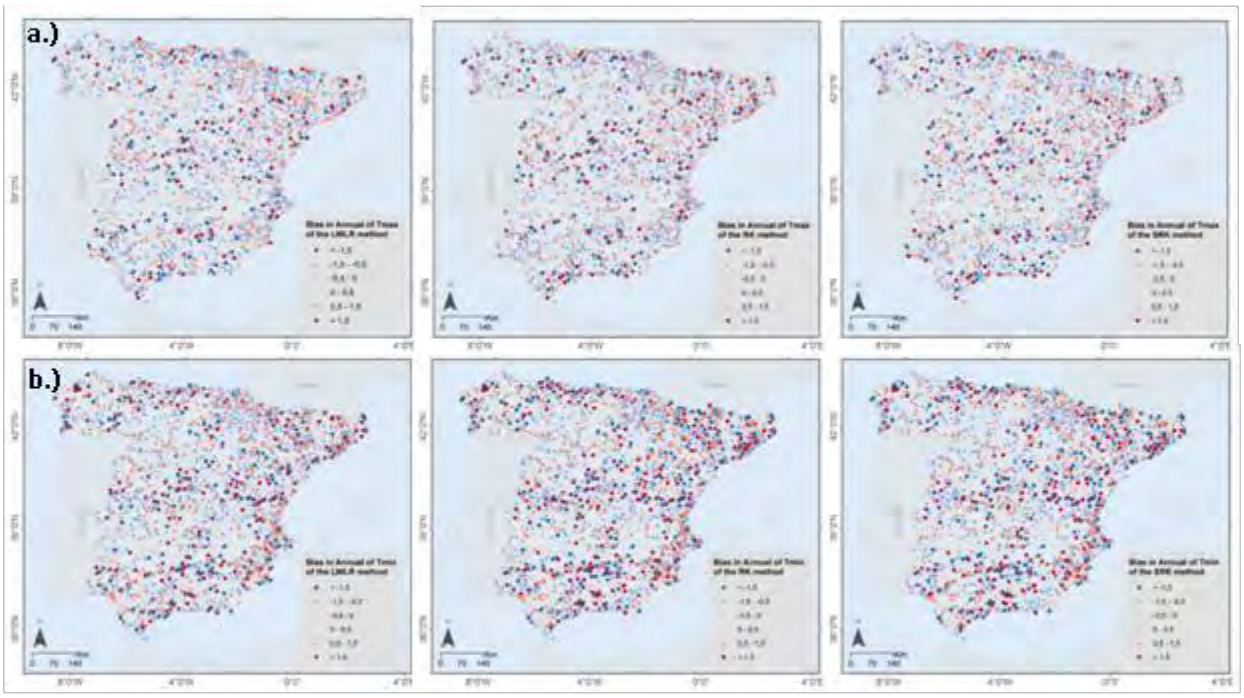


Fig.2. Spatial distribution of annual BIAS. a.) Tmax b.) Tmin

However, as said at the beginning, the Spanish mainland is a very complex territory and therefore these results may be affected by distance from water bodies, altitude or latitude. In his case in particular, we analyze the accuracy of all the models for Tmax and Tmin changes with altitude using MAE error (*Figure 3*).

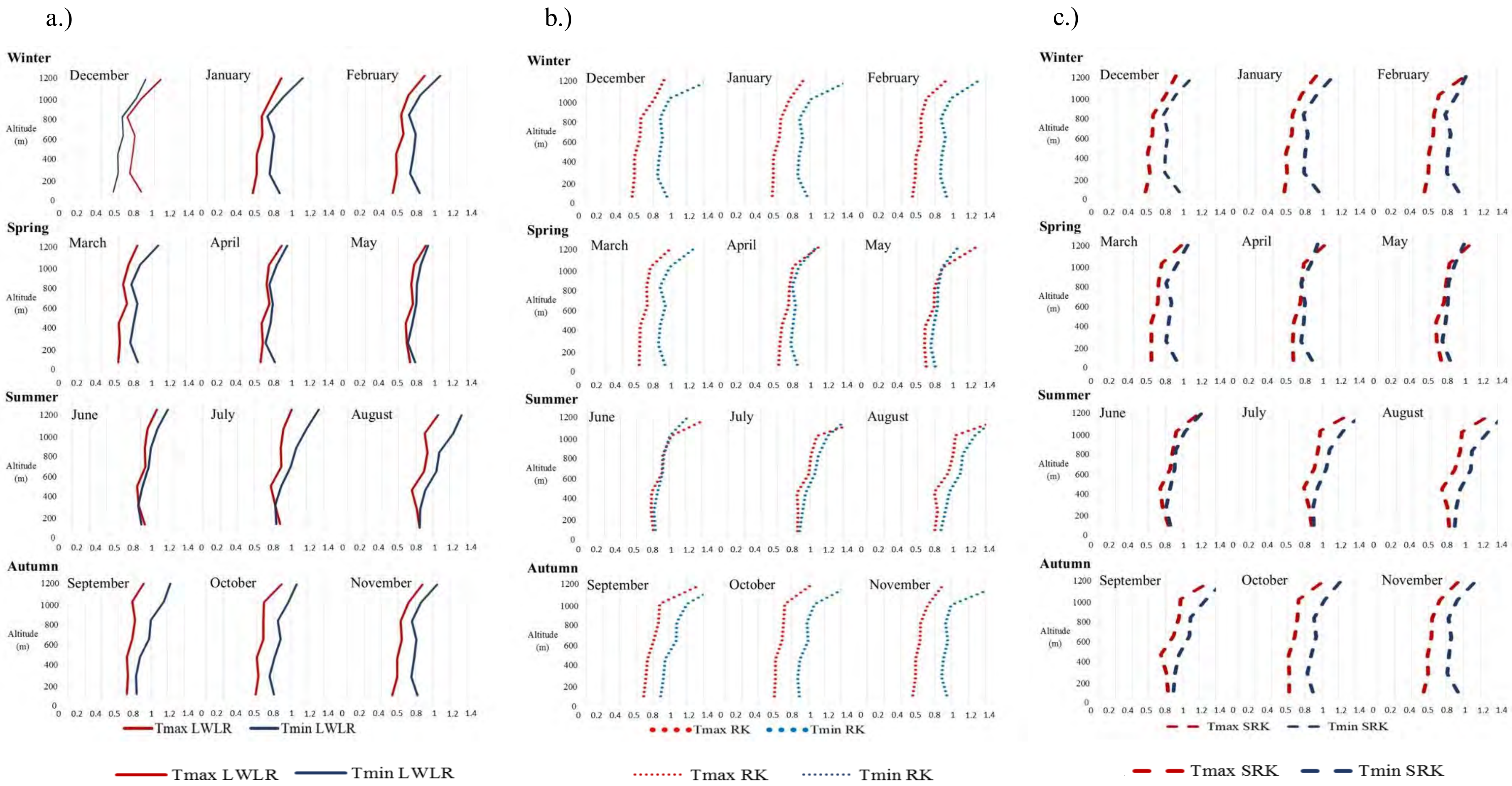


Fig. 3. MAE for different elevation bands in each month of Tmax and Tmin; a.) LWLR, b.) RK, c.) SKR)

Figure 3 shows the MAE values in different altitudinal bands (0-200, 200-400, 400-600, 600-800, 800-1200, greater than 1200 meters) for each month, for Tmax and Tmin and for each interpolation method applied (a. LWLR, b. RK, c. SRK). In general, it can be seen that the LWLR method offers better results, i.e. the MAE values are lower with the LWLR than with the other two methods. Also, it clear that Tmin has higher values of MAE than Tmax, which suggests that Tmin data are more difficult to estimate from the interpolation models tested. From the seasonal point of view, in the three methods of interpolation the summer months offer the worst results since the MAE values are higher with respect to the other months of the year. Finally, the MAE values increase with the altitude with the three interpolation methods, and the highest errors are observed from elevations of 1000 m (*Peña-Angulo et al.*, 2016a). Nevertheless in areas with higher elevations, a lower number of records is available, a classical problem of temperatures data in mountain regions.

In particular, the MAE values obtained with the LWLR method show monthly and altitudinal differences. In the cold months of autumn and winter, the Tmax and Tmin present their greatest differences in the low altitudinal levels, while in higher altitude areas, the MAE values of both variables are very similar. On the contrary, in the summer months, the MAE values at the lower levels are very similar for Tmax and Tmin; while in the high levels both thermal variables have their greatest differences. In the case of the SRK method the same pattern is observed, but the differences between the Tmax and Tmin are much more evident. Finally, the RK method shows the largest differences with respect to the other two methods. The MAE values obtained with the RK does not show a clear difference between high and low altitudes for Tmax and Tmin, and their MAE values are consistently higher than the other previously discussed interpolation methods. These results suggest that interpolation methods that use secondary variables to estimate the values of the dependent variable (in this case, the temperature) not only offer better results (*Vicente-Serrano et al.,* 2003), but also provide information on how these secondary variables affect diurnal and nocturnal temperatures.

The relationship between altitude and residuals was studied from de difference between the values predicted by the model and the observed values (BIAS). This analyses allow us to verify the validity of each method of interpolation and its capacity to model the temperature in function of the elevation. The analysis was performed using the regression model and its corresponding coefficient of determination (RSQR). Three regression models were performed for the whole dataset, also we present the results for those data overestimated by the model, and underestimate. The results of these three regression models are shown in *Figures 4 a, b* and *c* for each month, interpolation method, and Tmax and Tmin variable.

In general, the same patterns described for the MAE error are observed. In the Tmin, during the summer months, and in the high altitudes, we detected the worst results as those observed by *Peña-Angulo et al.* (2016a). The most interesting new results are shown in *Figure 4a* where we show the low values of RSQR in all cases, especially in the LWLR method. This low value indicates that the model is able to capture the variability that altitudes generate on temperatures. However, this capacity of the model decreases at higher altitude, which is attributed to the scarce availability of information. In *Figures 4b* and *4c* we show how that the Tmax is systematically overestimated (*Figure 4b*) while the Tmin is underestimated (*Figure 4c*), in all months and by the three models. This overestimation and underestimation of the Tmax and Tmin, respectively, is enhanced in the mountain areas, and affects especially overestimating Tmax during the warmest months, and underestimating Tmin during the coldest months. This disparity between the thermometric measurements indicates that models produces higher thermal gradient to real ones in Tmax, whereas the reverse is true for Tmin.

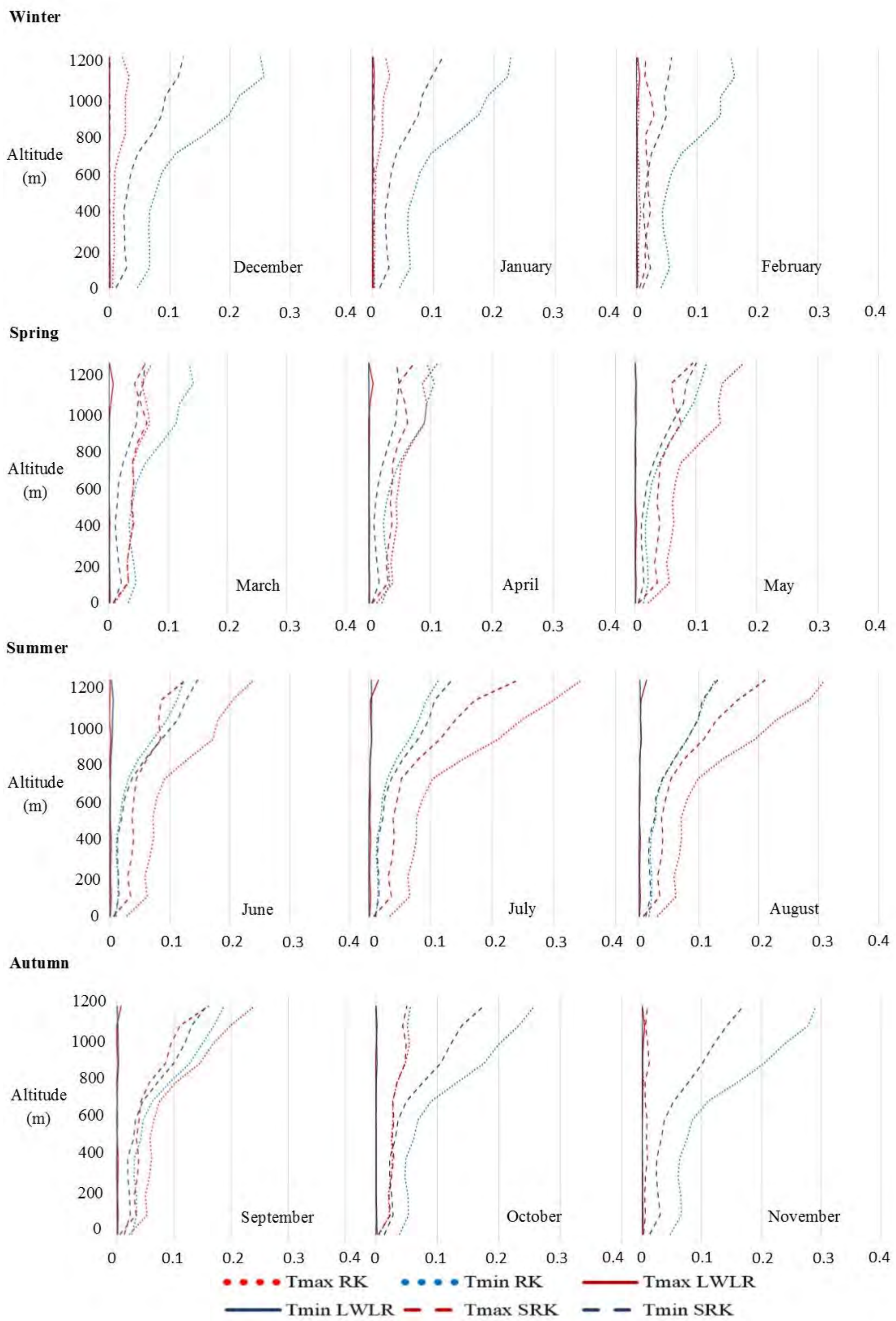


Fig. 4.a RSQR in each manual of Tmax and Tmin of the relationship between error and altitude (ALL)

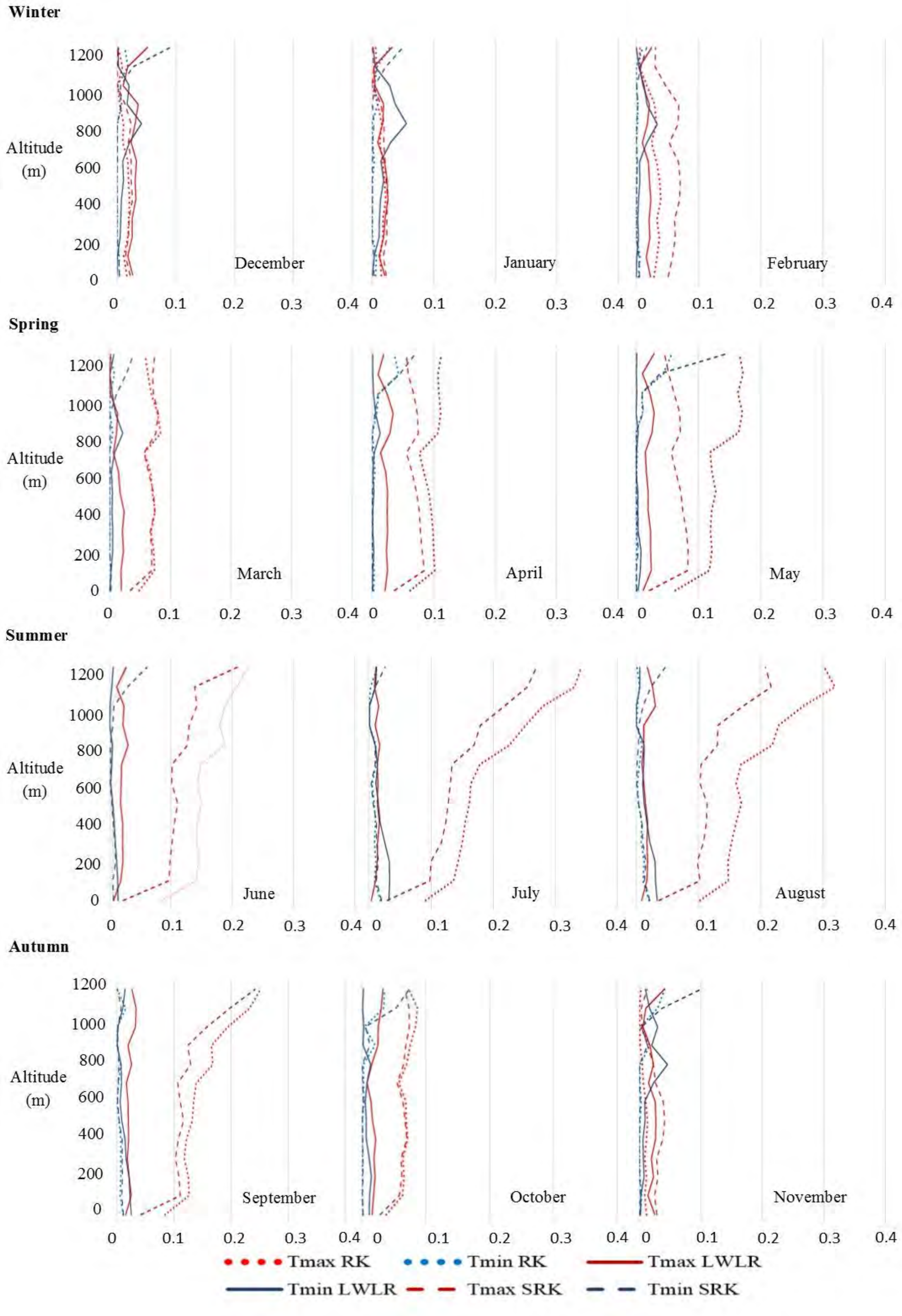


Fig. 4b. RSQR in each manual of Tmax and Tmin of the relationship between error and altitude (OVERESTIMATE)

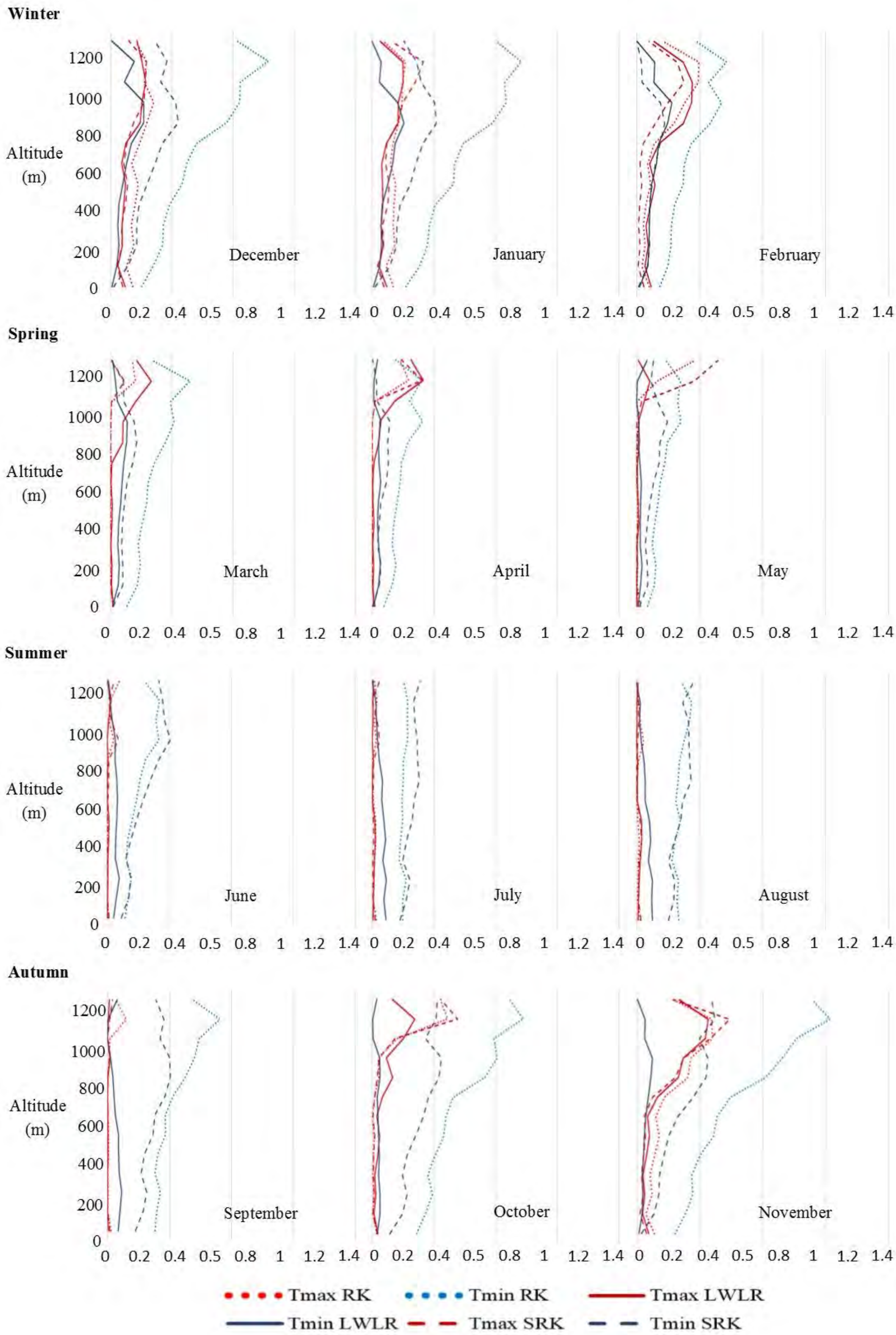


Fig. 4c. RSQR in each manual of Tmax and Tmin of the relationship between error and altitude (UNDERESTIMATE)

Spatial autocorrelation was tested in the residues from LWLR, RK and RSK. According to Moran's I spatial autocorrelation index there is no global autocorrelation among the residuals of either method or month. Additionally, local spatial autocorrelation in the residues of LWLR is addressed as well using a local disaggregation of the Moran's I (Anselin's Local I Moran for cluster and outlier analysis). This analysis allows to identify observatories with higher prediction error (outliers) compared to their respective nearest neighbors. Local Moran's I revealed some spatial pattern in the residue distribution. Tmax error clusters tend to be grouped in small spots along the coastline, especially in the Mediterranean coast from spring to early summer (*Figure 5*). This suggests that LWLR is not able to fully address thermal regulation effect due to proximity to the sea. Tmin error is more homogeneous from a spatial standpoint, although several clusters are detected with no clear pattern, except a tendency to move towards the hinterland during summer (*Figure 6*). High thermal amplitude during summer in the Spanish hinterland might be related with this behavior.

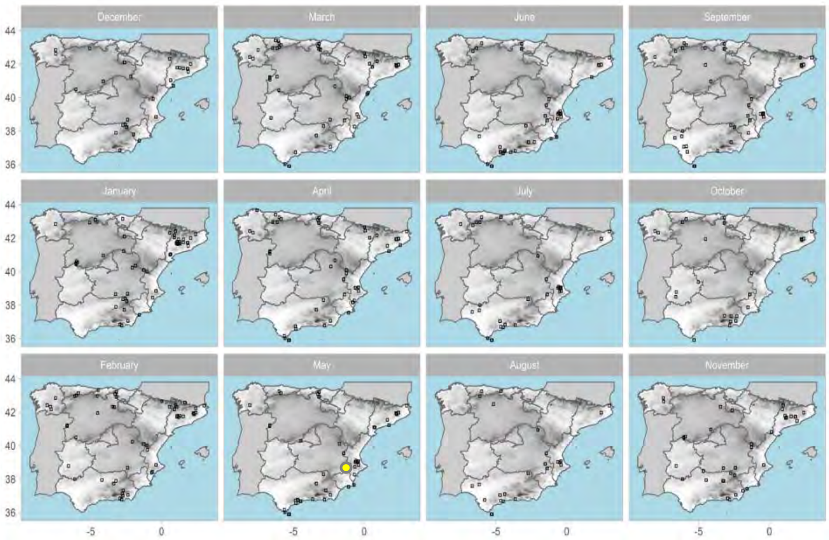
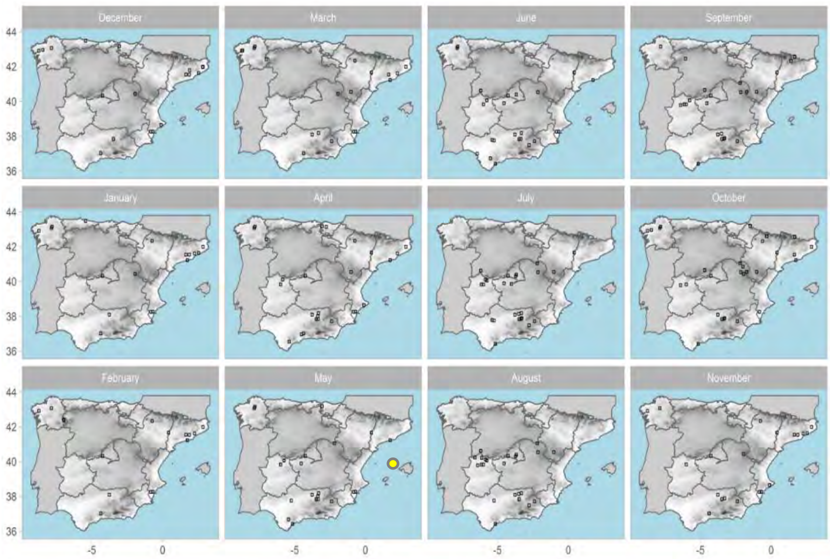


Fig. 5. Spatial autocorrelation in the residuals is tested using the Moran's I of Tmax



Outliers

Fig. 6. Spatial autocorrelation in the residuals is tested using the Moran's I of Tmin

4. CONCLUSIONS

The results of the exhaustively comparison of three interpolation methods have suggested that a better performance is obtained for maximum temperatures (diurnal records) rather than for minimum temperatures (nocturnal records) at monthly and annual scales. Furthermore, in all the cases analyzed the error is higher in summer months.

The comparative study of the three methods of interpolation suggests that the LWLR offers better results than the other two and suggest that this local method that uses secondary variables in the estimation of the temperatures is the most appropriate to obtain the climatology for the Spanish territory. As a consequence, we suggest that interpolation approach in environments with high spatial variability factors as the Iberian Peninsula should consider more than global those local factor to capture as much as possible the effect of local factors.

Finally, note that the worst estimates, from the three methods of interpolation, are obtained in mountain areas (over 1000 m), which coincides with areas with lowest density of information. In addition, it is important to note that in these areas the three interpolation methods tend to overestimate the Tmax; while the Tmin is underestimated.

FUNDING

This study was supported by the Ministry of Science and Innovation (Spanish Government), projects: the Development of Drought Index in Spain (DESEMON, CGL2014-52135-C3-3-R) and Gobierno Regional de Aragón DGA-FSE (Grupo de Investigación Consolidado 'Clima, Agua, Cambio Global y Sistemas Naturales'). Original data from AEMet (Spanish National Meteorological Agency). Celia Salinas Sole is a PhD student from FPI Program (Spanish Government).

References

Anselin L (1995) Local Indicators of Spatial Association-LISA. *Geographical Analysis* 27, 93-115.

Brunetti, M., et al., 2014. High-resolution temperature climatology for Italy: interpolation method intercomparison. *International Journal of Climatology*, 34 (4), 1278–1296. doi:10.1002/joc.2014.34.issue-4

Dai, A., Fung, I.Y., and Del Genio, A.D., 1997. Surface observed global land precipitation variations during 1900-88. *Journal of Climate*, 10 (11), 2943–2962. doi:10.1175/1520-0442(1997)010<2943:SOGLPV>2.0.CO;2

Daly, C., et al., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28, 2031–2064. doi:10.1002/joc.1688

Di Piazza, A., et al., 2011. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 13 (3), 396–408. doi:10.1016/j.jag.2011.01.005

Font Tullot, I., 1983. Climatología de España y Portugal. Ediciones Universidad de Salamanca Vol. 2007 (ISBN: 978-84-7800-944-2).

Gonzalez-Hidalgo, J.C., et al., 2015a. Recent trend in temperature evolution in Spanish mainland (1951–2010): from warming to hiatus. *International Journal of Climatology*. doi:10.1002/joc.4519.

Gonzalez-Hidalgo, J.C., et al., 2015b. MOTEDAS: a new monthly temperature database for mainland Spain and the trend in temperature (1951-2010). *International Journal of Climatology*, 35, 4444–4463. doi:10.1002/joc.4298

Hofstra, N., et al., 2008. Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*: Atmospheres, 113 (D21), D21110. doi:10.1029/2008jd010100.

Jones, P.D. and Hulme, M., 1996. Calculating regional climatic time series for temperature and precipitation: Methods and illustrations. *International Journal of Climatology*, 16, 361–377.doi:10.1002/(ISSN)1097-0088

Lionello, P., 2012. The Climate of the Mediterranean Region from the Past to the Future. Elsevier, Amsterdam, The Netherlands, p. 592

Martín, J., Olcina, J., 2001. Climas y tiempos de España. Historia y Geografía. Alianza Editorial, S.A., Madrid (264 páginas). ISBN: 84-206-5777-8).

Mitchell, T.D. and Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, 25 (6), 693–712. doi:10.1002/(ISSN)1097-0088

New, M., Hulme, M., and Jones, P., 2000. Representing twentieth-century space-time climate variability. Part II: development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate*, 13 (13), 2217–2238. doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2

Peña-Angulo, D., et al., 2016a. A new climatology of maximum and minimum temperature (1951-2010) in the Spanish mainland: a comparison between three different interpolation methods. *International Journal of Geographical Information Science*, 30(11), 2109-2132. doi: 10.1080/13658816.2016.1155712

Peña-Angulo, D., Trigo, R.M., Cortesi, N., González-Hidalgo, J.C., 2016b. The influence of weather types on the monthly average maximum and minimum temperatures in the Iberian Peninsula. *Atmos. Res*. 178-179, 217–230. http://dx.doi.org/10.1016/j.atmosres.2016.03.022.

Pielke, R.A., (1984): Mesoscale Meteorological Modeling. Quaterly *Journal of the Royal Meteorological Society*, 111, 671-672 pp.

Steel, R.G.D. and Torrie, J.H., (1960): Principles and procedures of statistics. McGrawHill Book Company, New York. 481 pp.

Tveito, O.E., et al., 2008. The use of geographic information systems in climatology and meteorology. COST action 719—ESSEM [online]. Office for Official Publications of the European Communities. ISBN: 978-92-898-0045-7. Available from: http://bookshop.europa.eu/uri?target=EUB:NOTICE:QSNA23461:EN:HTML [Accessed 26 Feb 2016]

Vicente-Serrano, S.M., Saz-Sánchez, M.A., and Cuadrat, J.M., 2003. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research*, 24 (2), 161–180. doi:10.3354/cr024161

Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63 (11), 1309–1313. doi:10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

EXPERIENCES WITH SNOW LEVEL ESTIMATION FOR SPATIAL ANALYSE OF NEW SNOW DEPTH BASED ON PRECIPITATION DATA

Květoň, Vít and Žák, Michal

Dept. of General Climatology
Czech Hydrometeorological Institute
Na Šabatace 17
CZ 143 06 Prague
Czech Republic

Department of General Climatology of the Czech Hydrometeorological Institute (CHMI) issues special winter road maintenance index since winter season of 2004-2005. This index is issued for roads and highways in about 100 regions of the Czech Republic (see *Figure 1*) and has two parts – index of plowing (*Figure 2*) and scattering. The index is provided for CROSS company for further processing and then delivered to Road and Motorway Directorate. The Index is rated as an excellent instrument of objective indication both road maintenance winter severity and retrograde supervision of selected road winter maintenance performances adequacy. It is very important for economic rating of winter maintenance centres.

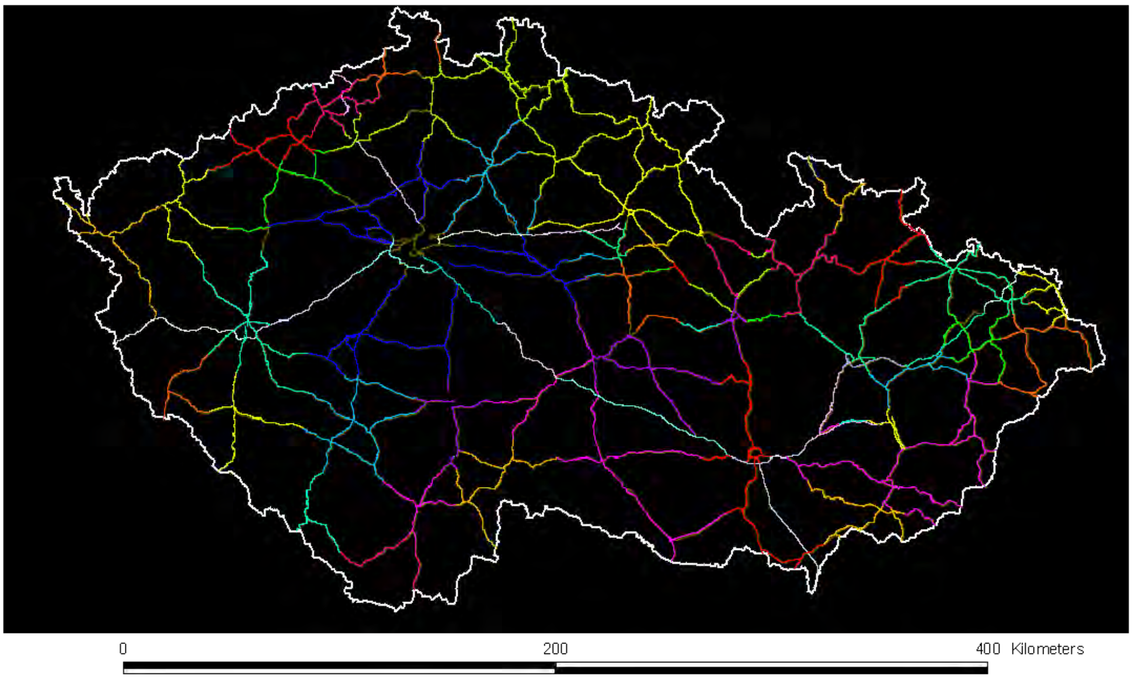


Fig. 1. Division of roads/highways network into individual maintenance centres

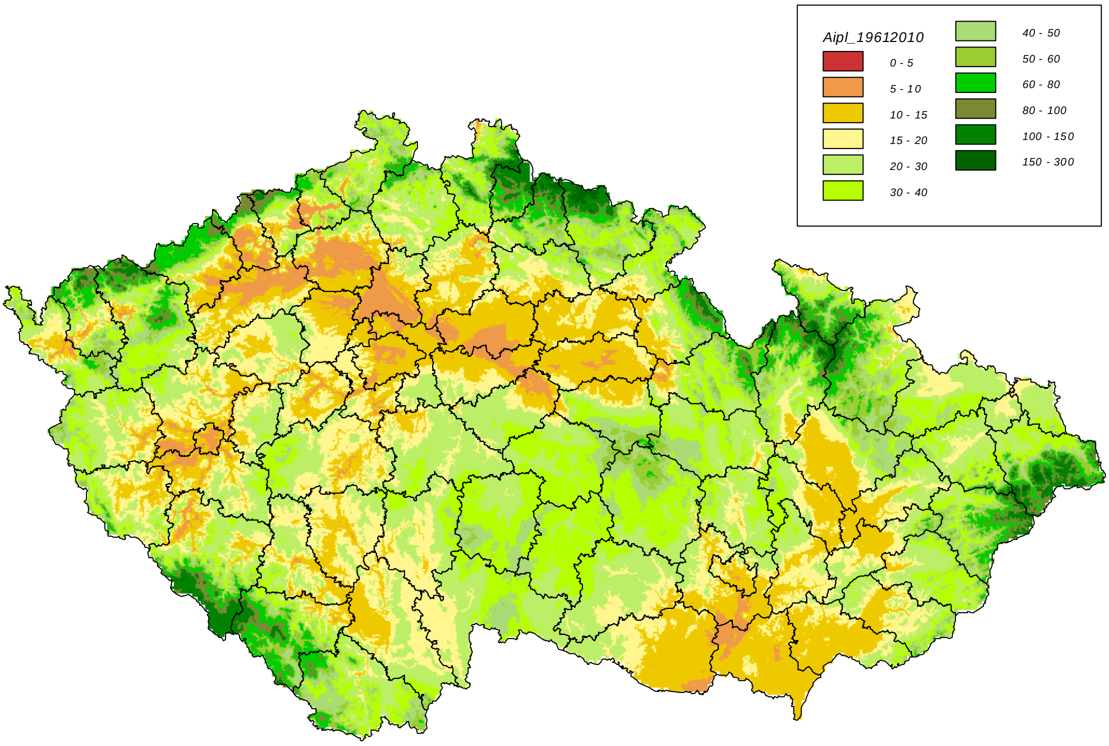


Fig. 2. Map of seasonal index of plowing (period 1961-2010)

Winter road maintenance index is based on daily values of selected meteorological parameters such as snow, icy road conditions (black ice, glaze, frost deposit) and snow drifts. New snow depth values are essential parameter for computation of the index. Computation of the index is based on spatial interpolation of station data.

Although the knowledge of the new snow depth is essential, only smaller part of stations send snow data every day to the CHMI's central database (so called "INTER" stations). The larger part of snow data is available with delay, usually around 10^{th} of the next month (let's call these stations „Snow"). Furthermore, some precipitation stations don't measure snow data (stations labelled as „Precipitation"). The number of the respective stations group is given in *Table 1* and stations positions in *Figure 3*.

Table 1. Number of station for given stations group

Station type	Count
INTER	97
Snow	249
Precipitation	51
Total	397

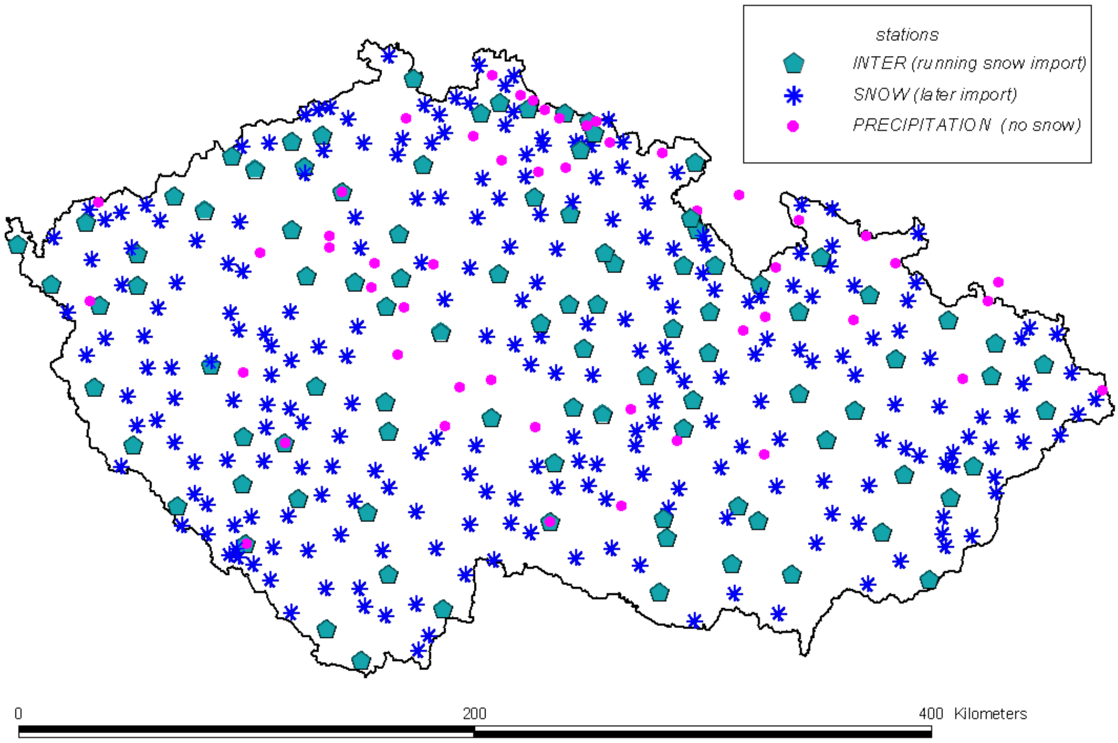


Fig. 3. Positions of all stations used for the index processing

Preliminary indices are issued every week with time resolution of 1 day. These indices are based on accessible snow data in the current week. Final indices are issued around 10th of the next month based on all climatological data (i.e. manual stations measurement, too). This can lead to discrepancies between preliminary and final winter maintenance indices. And that is also our motivation to estimate missing (or questionable) new snow values.

Following steps are used:

1) Determination of snow level

2) Snow density estimation

3) New snow estimation

The **snow level** is specified as the altitude of the lowest station with SNO in defined surrounding (region) (see *Figure 4*). This level is computed for 10 regions that were selected according to the orographic features in the frame of the Czech Republic (*Figure 5*).

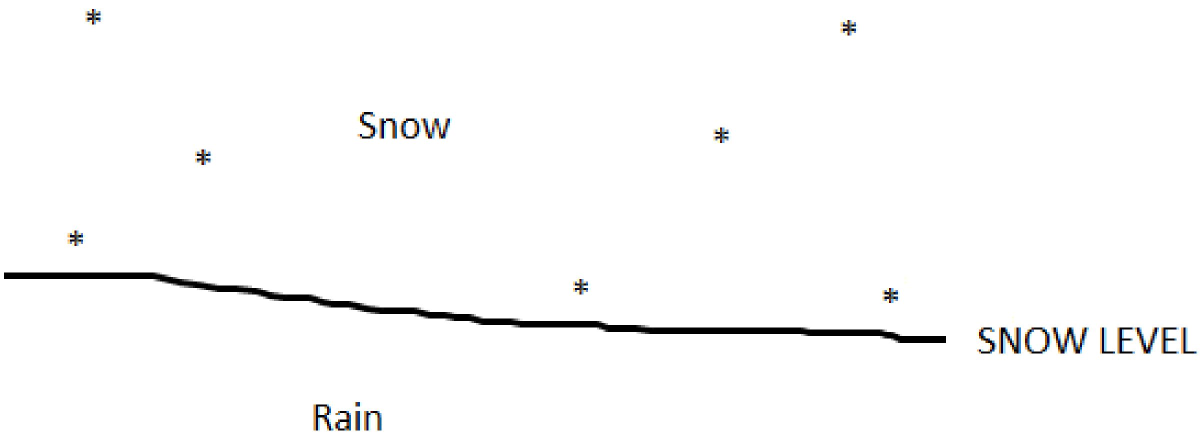


Fig. 4. Determination of snow level

Fig. 5. Division of the Czech republic into 10 regions for snow level estimation

Regarding **snow density** (SNOdensity), it has been computed as ratio of new snow value to precipitation total at the nearest station (refstation) with the least altitude difference but still situated above the snow level. If the new snow value at the refstation is smaller than 4, snow density is taken as 0.2 with respect to lower validity of snow density in these cases. It has to be stressed that snow density is important both from a view of maintenance difficultness and for estimation of new snow value based on daily precipitation sum.

As for **new snow estimation**, following steps are applied:

a) spatial interpolation of SRA (daily precipitation sum)
b) spatial interpolation of SNO (daily new snow depth)
c) estimation of missing or questionable new snow values in pixels:
    a. If (SRA=0) then SNO=0
    b. If SNO <= 0 in given pixel and SRA>0, then
        i. if (Altitude of pixel < Snow level) then SNO=0
        ii. else SNO=SRA*SNOdensity
d) Daily averages over pixel values for each maintenance centre were done

Results of our approach can be seen on *Figure 6* (for one selected highways maintenance centre), *Figure 7* (for first class road maintenance centre) and on *Figure 8* (for all maintenance centres in the Czech Republic) where differences between maintenance centre new snow values based on measured values (from INTER stations) and estimated snow values (labelled as SNO(COR) in graphs) are depicted for every day of the winter season 2016/2017 (starting on November 1^{st}, finishing on February 28^{th}).

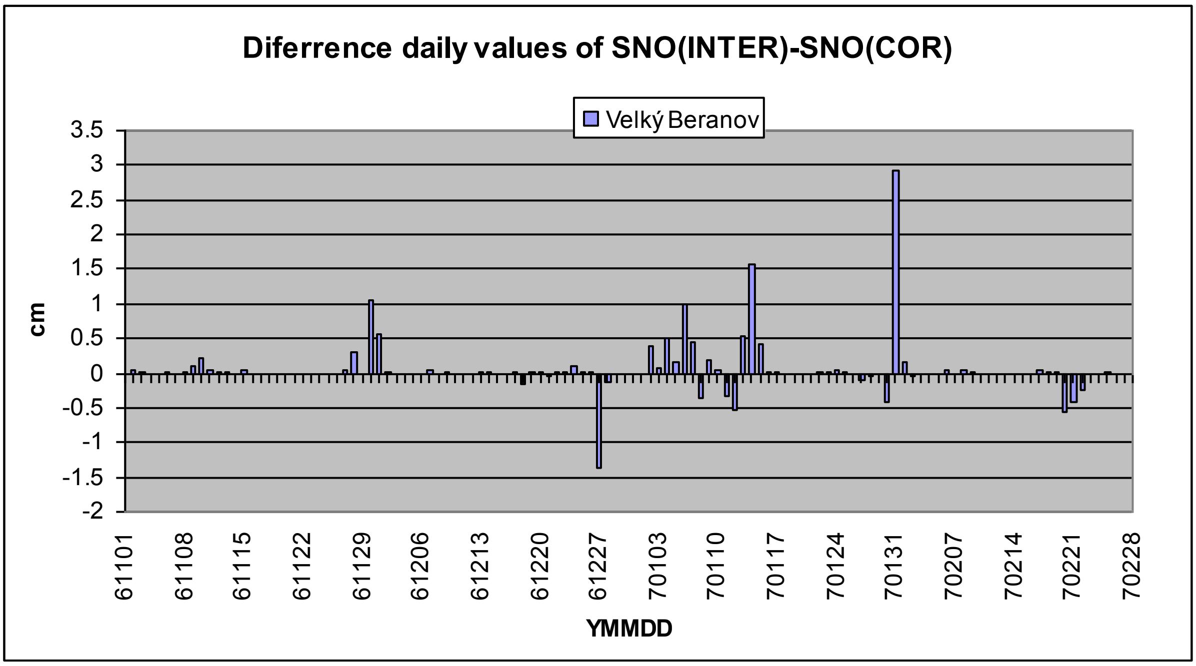


Fig. 6. Differences of daily values of snow values based on INTER stations and estimated values for highway maintenance centre

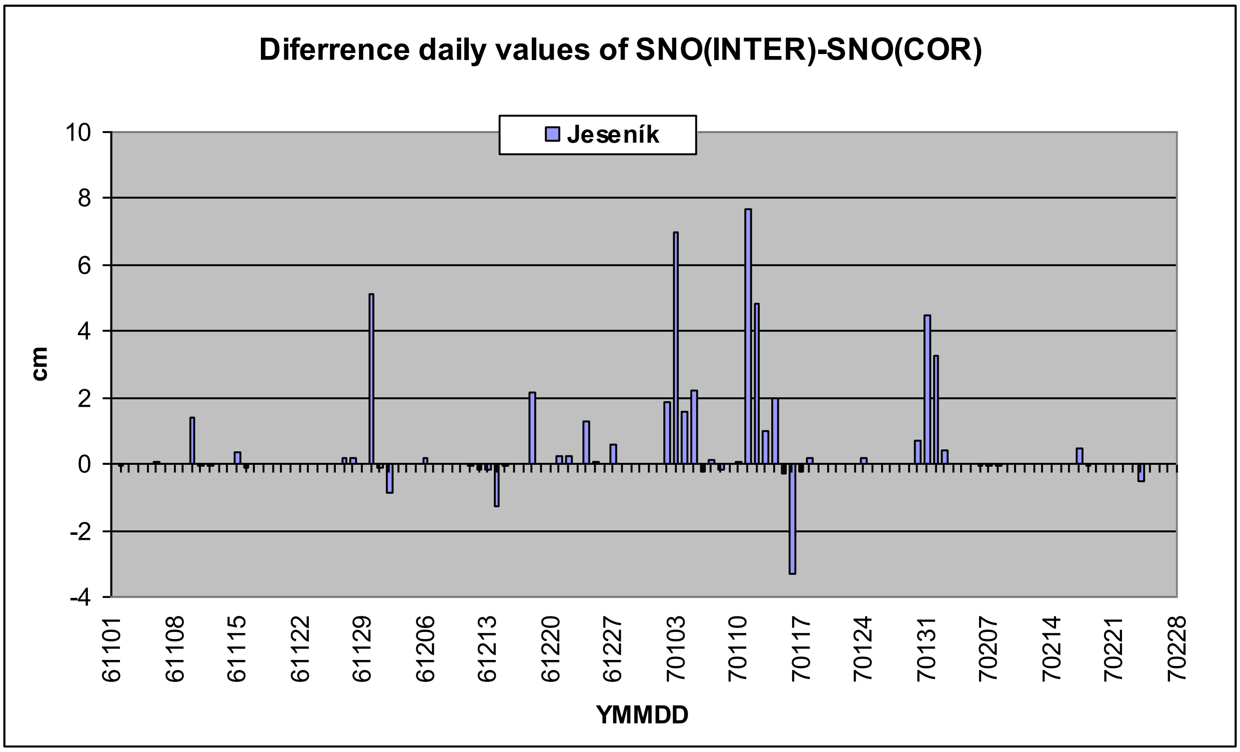


Fig. 7. Differences of daily values of snow values based on INTER stations and estimated values for 1st class road maintenance centre

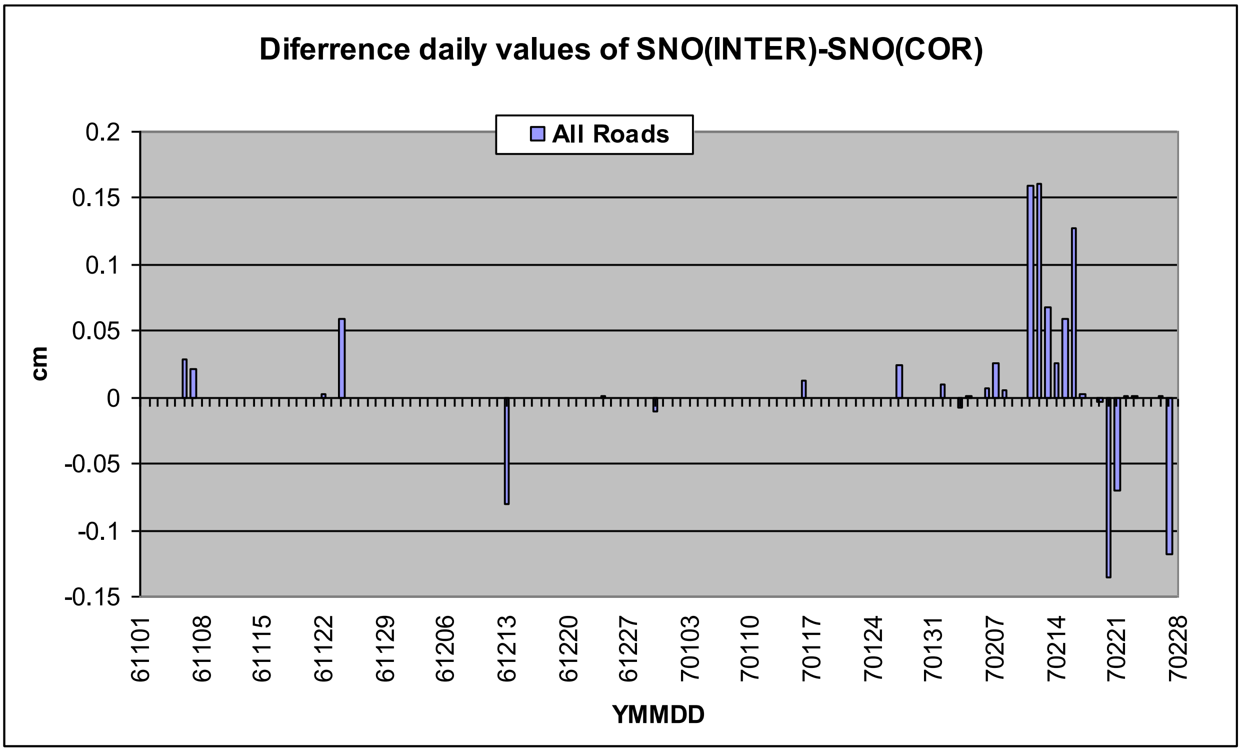


Fig. 8. Differences of daily values of snow values based on INTER stations and estimated values averaged for all maintenance centres in the Czech Republic

As can be seen from the above graphs, the differences between the values are up to several centimetres for various maintenance centres, usually our correction tends to lower a bit new snow values. According to CROSS company evaluation, the winter road maintenance index (preliminary values) based on corrected values of new snow cover provides better results than without this correction (it means based only on smaller count of new snow measuring stations). We plan to make further results verification and based on it, further development of this method is underway.

HOMOGENIZATION AND GRIDDING OF THE GREEK TIME SERIES

A. Mamara[1], M. Anadranistakis[2], A.A. Argiriou[3]

1. Hellenic National Meteorological Service, El. Venizelou 14, GR-167 77 Elliniko, Greece, tel. +30 210 9699101, fax: +30 210 9649646, e-mail: anna.mamara@hnms.gr
2. Hellenic National Meteorological Service, El. Venizelou 14, GR-167 77 Elliniko, Greece, tel. +30 210 9699101, fax: +30 210 9649646, e-mail: anad@hnms.gr
3. Laboratory of Atmospheric Physics, University of Patras. GR-265 00 Patras, Greece, tel: +30 2610 996078, e-mail: athanarg@upatras.gr

Abstract

Weather observations are frequently exposed to artificial influences caused by station relocations, changes in the instrumentation, etc. introducing inhomogeneities. Data series of temperature, precipitation and sunshine duration were homogenized on a monthly time scale. Data from the Hellenic National Meteorological Service station network and of the Public Power Corporation S.A., cover the period from 1971 to 2000.

The homogenized derived data series were used to create a high resolution (0.0083333° × 0.0083333°) gridded data set. The method MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) developed at the Hungarian Meteorological Service was used for the spatial interpolation of homogenized monthly values. The potential to use the elevation and the first 15 AURELHY principal components as temperature predicators was investigated. It was found that in the Mediterranean with an important coastline, the use only of elevation and the AURELHY variables cannot describe temperature. Additional topographical and geographical variables, namely the land to sea percentage and the expected solar irradiance are required. Compared to previous climatologies, the proposed database has the following improvements: data are provided at a higher spatial resolution, temperature data were homogenized, improved geographical and topographical data were used, an interpolation method appropriate for meteorological parameters was applied and the statistical results of the observed versus predicted values were better.

1. INTRODUCTION

Modern climatology requires high resolution climate atlases, a requirement reflected also by the WMO requirement (WMO, 2014), to which many meteorological services have already complied (e.g. the German Climate Atlas http://www.dwd.de/bvbw/appmanager/bvbw/dwdwww Desktop?_nfpb=true&_pageLabel=dwdwww_result_page&gsbSearchDocId=1070304; Digital Climatological Atlas of Spain http://opengis.uab.es/wms/iberia/mms/, Finnish wind Atlas http://www.tuuliatlas.fi/creators/index.html. To cover specific needs, high resolution climatologies have also been published, like for the greater Alpine area (*Auer et al.*, 2008; *Hiebl et al.*, 2009) and for the Carpathian Mountains (e.g.; *Lakatos et al.*, 2013).

For Greece, until recently, the only climate atlases were that of *Mariolopoulos and Livathinos*, (1935) and of the project GEOCLIMA (http://www.geoclima.eu). To fill the gap, the Hellenic National Meteorological service, in collaboration with the Laboratory of Atmospheric Physics of the University of Patras, decided to develop a high-resolution climate atlas for Greece (*Mamara et al.,* 2017), based on homogenized data (*Mamara et al.,* 2013; *Mamara et al.,* 2014).

2. DATA

The data sets cover a climatological normal period of 30 years (1971-2000) and come from the Hellenic National Meteorological Service station network, which is part of the WMO Global Observing System. The temperature data come from 52 meteorological stations and the sunshine duration data from 44 meteorological stations of the HNMS's network. The precipitation data come from 157 meteorological stations, 89 stations belong at the Public Power Corporation S.A. Hellas and 68 stations at the HNMS. The position of the stations is shown in *Figure 1*.

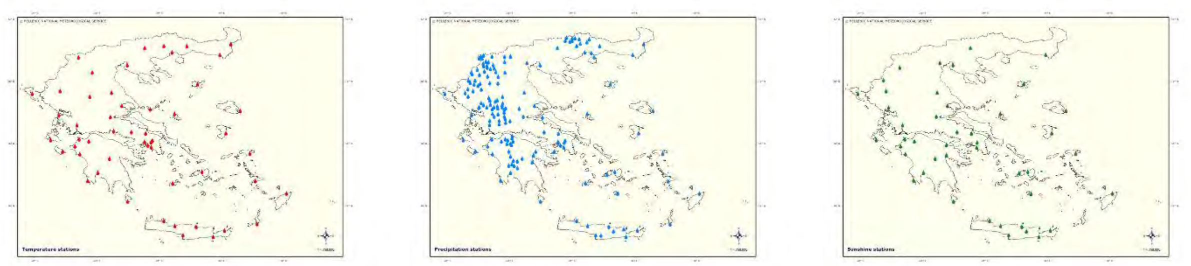


Fig. 1. Location of the meteorological stations measuring (a) temperature (left), (b) precipitation and (c) sunshine duration.

3. HOMOGENISATION

It is well known that meteorological time series are subjected to artificial influences due to various causes, like station relocations, changes in the instrumentation, etc. introducing inhomogeneities. The data series of temperature, precipitation and sunshine duration used for the climatic atlas, were homogenized on a monthly time scale basis using the HOMER software (*Venema et al.*, 2012). Homogenization results were then cross-checked using the MASH, ACMANT and CLIMATOL homogenization methods. Due to the complex topography of Greece, the homogenization exercise was not applied once to all available time series, but it was performed separately to the time series belonging to the climatic regions shown in *Figure 2*.

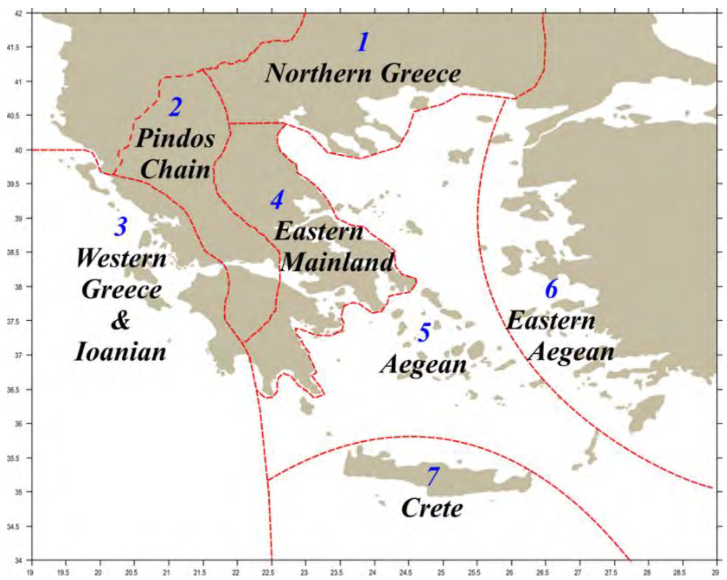


Fig. 2. Climatic regions used during the homogenization.

The homogenization exercise detected 56 breaks over 52 monthly mean temperature time series, 75 breaks over 52 monthly minimum temperature time series, 51 breaks over 52 monthly maximum temperature time series, 10 breaks over 44 monthly sunshine duration time series and 65 breaks over 157 monthly precipitation time series. The identified break points were corrected accordingly, based on meta-data. A correction example is shown in the *Figure 3*, which shows the mean winter temperature in the island of Zante, in the Ionian Sea, Western Greece.

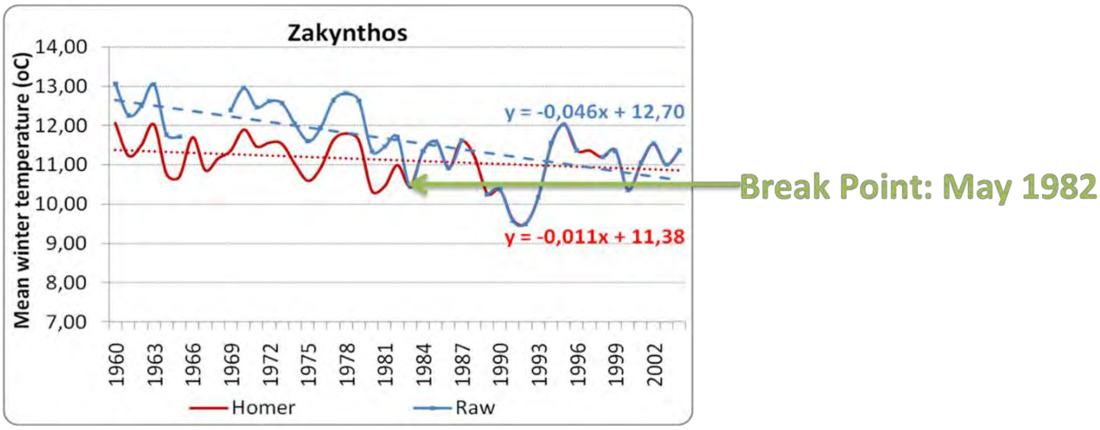


Fig. 3. Break point detection and correction in the monthly mean temperature time – series, Zante, Greece.

The general conclusions of the homogenization exercise were that the most frequent identified reasons for inhomogeneities were station relocation from towns to airports and the change of observation practices. Most breaks (for both temperature and precipitation series) occurred in the '80s; this was confirmed by meta-data showing that during that period many stations were relocated to airports to cover aviation needs. Finally, no clear improvement of precipitation series, most probably due to smaller correlations between stations.

4. SPATIAL INTERPOLATION

The homogenized time-series were used to create a high resolution (0.0083333° × 0.0083333°) gridded data set. Elevation data coming from the DEM originating from the NASA (SRTM) 90 × 90 m (http://srtm.csi.cgiar.org) were used as predictors of temperature, precipitation and sunshine duration. The method MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) developed at the Hungarian Meteorological Service was used for the spatial interpolation of homogenized monthly values (*Szentimrey & Bihari*, 2014). The main advantages of MISH method are that it requires homogenized data series, it gets information of long term data series (usually geostatistical methods use a single realization in time for modelling but meteorological data are spatio-temporal data!), it calculates the optimum interpolation parameters which are certain known functions of the climate statistical parameters and it uses an interpolation formula depending on the climate parameter.

Also, the potential to use the elevation and the first 15 AURELHY principal components as temperature predicators was investigated. AURELHY is a method that models the spatial interpolation of climatic variables using topography (*Benichou and Le Breton*, 1987). AURELHY proposes an objective method based on principal component analysis for the determination of topographic variables It was found that in the Mediterranean with an important coastline, the use only of elevation and the AURELHY variables cannot describe temperature. Additional topographical and geographical variables, namely the land to sea percentage (*Feidas et al.,* 2013) and the expected solar irradiance, calculated using libRadtran, a detailed radiative model for solar and terrestrial radiation in the atmosphere (*Mayer and Kylling,* 2005) were used.

The results revealed that elevation, land to sea percentage and solar irradiance should be used as independent model variables.

5. RESULTS AND CONCLUSIONS

The results revealed that elevation, land to sea percentage and solar irradiance should be used as independent model variables. Also, the east-west slopes seem to be associated with mean temperature during all months except between April to July; latitude seems to affect mean temperature as well. North-south saddles are related to mean temperature only during the winter months.

As shown by the correlation coefficient between the observed and predicted values (*Figure 4*), the model performs very well from October to March, quite good for April and September and relatively good from May to August. The lower correlation results mainly in summer could be attributed to the lower spatial variability of mean temperature. However, the modelling results of the spatial trend in August in the main continental part plus a limited part of the coastal area were much better, showing that the dependence of mean temperature from the selected topographical and geographical variables is sufficient.

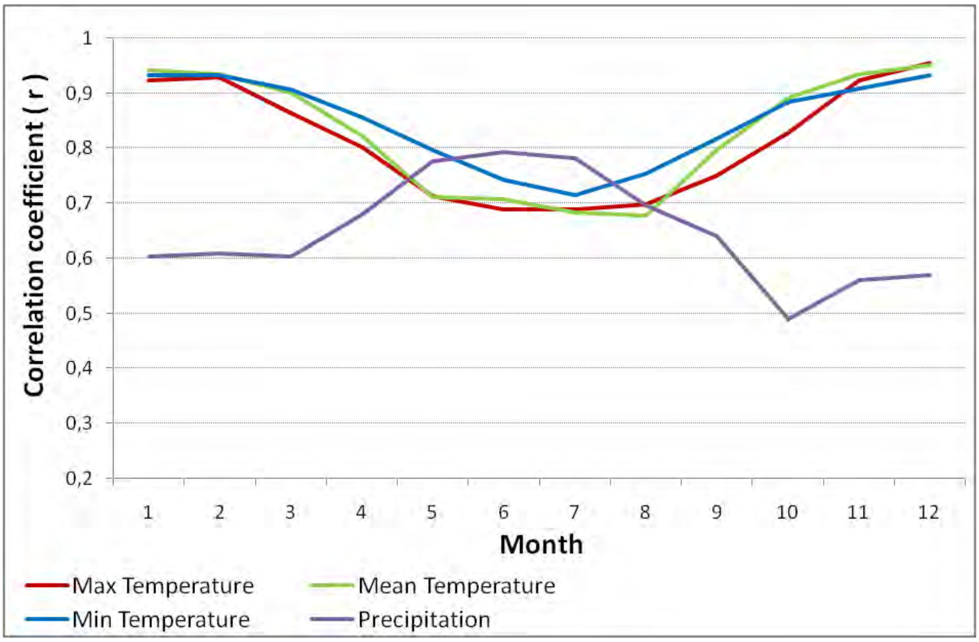


Fig. 4. Correlation coefficients between observed and predicted, by the spatial model, values, as a function of the month.

The mapping of the monthly temperature normals (1971-2000) is very close to reality. Lower temperatures, mainly in winter, at mounts Olympus and Voras (north-central Greece), on the Pindos mountain chain (western Greece) and on the Rhodopes mountain range (north-eastern Greece) can be clearly distinguished; also, the low temperatures on other mountains such as the White Mountains and mount Psiloritis in Crete, mount Parnassus in central Greece and mount Taygetos in the Peloponnese in southern Greece are clearly shown. The highest temperatures in summer can be observed in the plains of Thessaloniki, of Thessaly, of Kopais, of Argolis and of Aitoliki, as it happens (*Figure 5*).

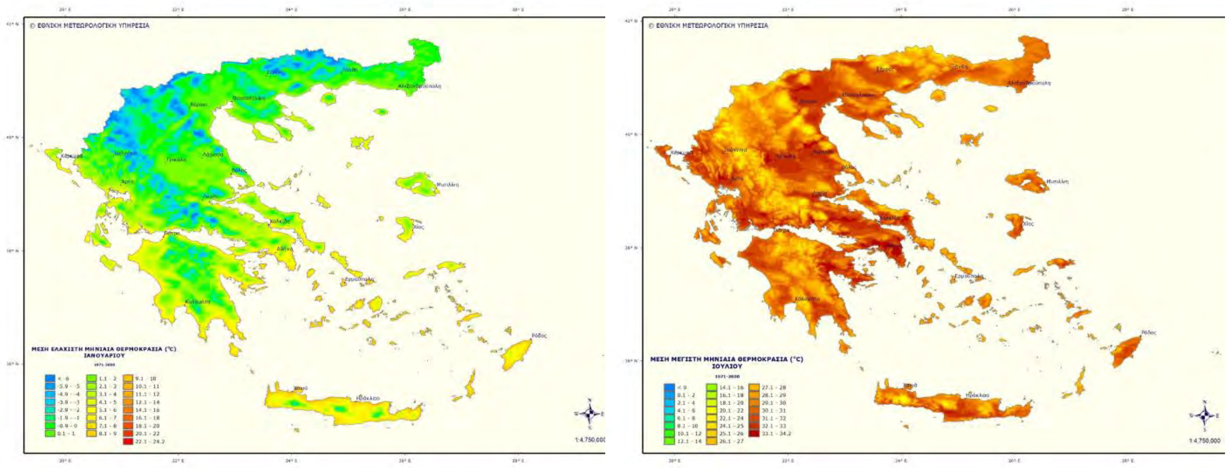


Fig. 5. Monthly minimum temperature in January (left) and monthly maximum temperature in July (right) for the normal period 1971 – 2000.

Examples of mean monthly precipitation and monthly sunshine duration are shown in *Figure 6.* As it can be shown in the left pane of *Figure 6*, the proposed model describes well characteristics of the distribution of precipitation; the rain shadow on the eastern part of the mainland is characteristic.

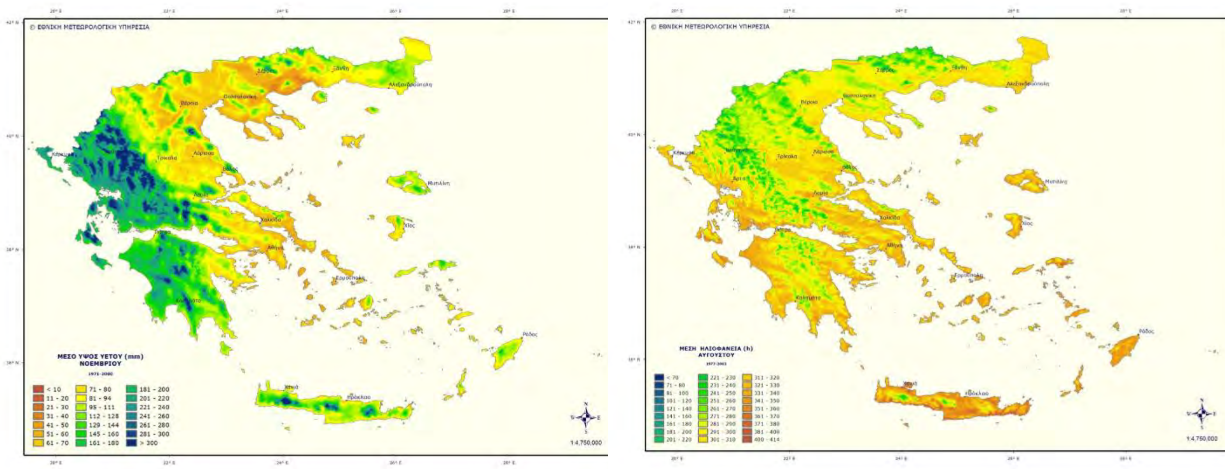


Fig. 6. Monthly mean precipitation in November (left) and monthly mean sunshine in July (right) for the normal period 1971 – 2000.

Compared to a previous climatology, the proposed atlas has the following improvements: data are provided at a higher spatial resolution, all data were homogenized, improved geographical and topographical data were used, an interpolation method appropriate for meteorological parameters was applied and the statistical results of the observed versus predicted values were better. In a future work, additional climatic data will be homogenized and interpolated.

References

Auer I, Böhm R, Hiebl J, Schöner W, Spinoni J, Lentini G, Maugeri M, Brunetti M, Nanni T, Perčec Tadić M, Bihari Z. 2008. ECSN – HRT/GAR High Resolution Temperature Climatology in Complex Terrain – demonstrated in the test area Greater Alpine Region GAR Final Report. Vienna.

Bénichou P, Le Breton O. 1987. AURELHY: une method d'analyse utilisant le relief pour les besoins de l'hydrométéorologie. In: Deuxièmes journées hydrologiques de l'ORSTOM à Montpellier. Paris: *ORSTOM*, p. 299-304. (Colloques et Séminaires). ISBN 2-7099-0865-4.

Feidas H, Karagiannidis A, Keppas S, Vaitis M, Kontos T, ZanisP, Melas D and Anadranistakis E. 2013. Modelling and mapping temperature and precipitation climate data in Greece using topographical and geographical parameters. *Theor. Appl. Climatol.*118:133-146.

Hiebl J, Auer I, Böhm R, Schöner W, Maugeri M, Lentini G, Spinoni J, Brunetti M, Nanni T, Perčec Tadić M, Bihari Z, Dolinar M, Müller-Westermeier G. 2009. A High Resolution Temperature Climatology for the Greater Alpine Region (GAR) *Meteorol. Z.* 18:507-530.

Mamara A, Argiriou AA, Anadranistakis M. 2013. Homogenization of mean monthly temperatures of Greece. *Int. J. Climatol*. 33: 2649–2666. DOI: 10.1002/joc.3614

Mamara A, Argiriou AA, Anadranistakis M. 2014. Detection and correction of inhomogeneities in Greek climate temperature series. *Int. J. Climatol*. 34: 3024–3043. DOI: 10.1002/joc.3888

Mamara A, Argiriou AA, Anadranistakis M. 2017. High resolution air temperature climatology for Greece for the period 1971–2000. *Meteorol. Appl*. 24: 191–205. DOI: 10.1002/met.1617

Mariolopoulos EG and Livathinos AN. 1935. Atlas Climatique de la Grece. Observatoire National d' Athenes. Athenes, Grece.

Mayer B and Kylling A. 2005. Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmos. Chem. Phys*. 5:1855-1877.

Szentimrey T. and Bihari Z. 2007. Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis). Proceedings from the Conference on Spatial Interpolation in Climatology and Meteorology. Budapest. Hungary. 2004. COST Action 719. COST Office.17–27.

Venema V, Mestre O, Aguilar E, Auer I, Guijarro JA, DomonkosP,Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Muller-Westermeier G, Lakatos M, Williams CN, Menne M, LindauR, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L,Acquaotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, GruberC, Duran MP, Likso T, Esteban P, Brandsma T. 2012. Benchmarkingmonthly homogenization algorithms. *Clim. Past* 8: 89–115.

WMO-No 1131 2014. Climate Data Management System Specifications. Geneva. Switzerland.

COMPARISON OF THE E-OBS AND THE CARPATCLIM GRIDDED DATASETS FOR MINIMUM TEMPERATURES, MAXIMUM TEMPERATURES AND PRECIPITATION BY THE ANALYSIS OF VARIANCE (ANOVA) METHOD

Mónika Lakatos, Tamás Szentimrey, Beatrix Izsák, Lilla Hoffmann

Hungarian Meteorological Service
lakatos.m@met.hu

Abstract

Several observational datasets are available to provide background for climate studies and for derivation of climate information for a specific area. The E-OBS is the largest available pan-European gridded data set for this purpose. Beside the pan European datasets a high resolution gridded climate dataset is available for the Carpathian Region too. The CARPATCLIM dataset consists of homogenized, harmonized and gridded data on daily scale for basic meteorological variables and several climate indicators from 1961 to 2010 in 0.1° spatial resolution. Some of statistical properties of the CARPATCLIM and E-OBS gridded datasets are compared for the common region in this paper. The seasonal and the annual average maximum and minimum temperatures and precipitation sums are examined for the area covered by the both dataset. Generally used statistical methodology: the Analysis of Variance (ANOVA) was applied for this purpose. The computations summarized here and the results plotted in graphs and maps are available for further analysis.

1. INTRODUCTION

Gridded climate data derived from meteorological measurements are used in climate research, validation of global and regional climate models, in many applications in climate change impacts assessments and derivation of different climate products. Several observational datasets are available to provide background for climate studies and for derivation of climate information for a specific area. One of the source of uncertainty of a gridded data is related to the limited number of the available station observations and the other is the interpolation method were used for estimation of the gridded values from the underlying station network. In this paper the statistical properties of the CARPATCLIM (*Szalai et al.,* 2013) and E-OBS (*Haylock et al.,* 2008) gridded datasets are compared for the Carpathian Region. In this work the seasonal, annual average maximum temperatures, minimum temperatures and precipitation sums are examined for the regions commonly covered by E-OBS and CARPATCLIM in the period of 1961-2010. For the comparison of the datasets are in the focus of this paper the general statistical methodology of Analysis of Variance (ANOVA) was applied.

2. DATA

Gridded data sets can be derived through the interpolation of the data originating from measurements recorded at meteorological observation stations. The largest available pan-European high-resolution gridded data set including daily climate data over Europe is the E-OBS (*Haylock, et al., 2008)*. The E-OBS gridded data set is derived and updated regularly through interpolation of the data of the ECA&D (European Climate Assessment and Dataset) stations. Beside the pan European datasets a high resolution climate dataset is available for the Carpathian Region. The CARPATCLIM consists of daily gridded data, that is homogenized harmonized and gridded by MASH (*Szentimrey*, 2011) and MISH, (*Szentimrey and Bihari*, 2007) methods. The spatial resolution of this dataset is 0.1 degree. The basic meteorological variables and several climate indicators, 37 in total, from 1961 to 2010 are publically available for the Carpathian Region. The area of CARPATCLIM partly includes the territory of the Czech Republic, Slovakia, Poland, Ukraine, Romania, Serbia, Croatia, Austria and Hungary. 415 climate stations and 904 precipitation stations were homogenized and interpolated to a grid covering the Carpathian Region. The E-OBS v14.0 0.25 degree regular grid of daily maximum, minimum temperatures and precipitation were downloaded for the period between 1961 and 2010 for the CARPATCLIM region and was used in the comparison.

3. METHODOLOGY

For the comparison of these datasets we applied a general statistical methodology, namely the Analysis of Variance (ANOVA) method. This methodology can be used effectively for the characterization of the statistical properties of several spatiotemporal datasets like CARPATCLIM and E-OBS. Station data or gridded datasets with different spatial resolution can be compared by analysing the spatiotemporal means and variances (*Szentimrey*, 2016). This methodology is built into the modelling part of method MISH in order to evaluate the modelling results automatically.

ANOVA (Analysis Of Variance) examination

Notations

$Z(\mathbf{s}_j,t)\ (j=1,..,N;\ t=1,...,n)$ – data series $(\mathbf{s}_j: \text{grid}; t: \text{time})$

$$\hat{E}(\mathbf{s}_j)=\frac{1}{n}\sum_{t=1}^{n}Z(\mathbf{s}_j,t) \quad (j=1,..,N) \text{ – sample mean at grid } \mathbf{s}_j$$

$$\hat{D}^2(\mathbf{s}_j)=\frac{1}{n}\sum_{t=1}^{n}\left(Z(\mathbf{s}_j,t)-\hat{E}(\mathbf{s}_j)\right)^2 \quad (j=1,..,N) \text{ – sample variance at grid } \mathbf{s}_j$$

$$\hat{E}(t)=\frac{1}{N}\sum_{j=1}^{N}Z(\mathbf{s}_j,t) \quad (t=1,..,n) \text{ – sample mean at moment } t$$

$$\hat{D}^2(t)=\frac{1}{N}\sum_{j=1}^{N}\left(Z(\mathbf{s}_j,t)-\hat{E}(t)\right)^2 \quad (t=1,..,n) \text{ – sample variance at moment } t$$

$$\hat{E} = \frac{1}{N \cdot n} \sum_{j=1}^{N} \sum_{t=1}^{n} Z(\mathbf{s}_j, t) = \frac{1}{N} \sum_{j=1}^{N} \hat{E}(\mathbf{s}_j) = \frac{1}{n} \sum_{t=1}^{n} \hat{E}(t) \quad \text{– total sample mean}$$

$$\hat{D}^2 = \frac{1}{N \cdot n} \sum_{j=1}^{N} \sum_{t=1}^{n} \left(Z(\mathbf{s}_j, t) - \hat{E} \right)^2 \quad \text{– total sample variance}$$

Partitioning of Total Variance (Theorem)

$$\hat{D}^2 = \frac{1}{N} \sum_{j=1}^{N} \left(\hat{E}(\mathbf{s}_j) - \hat{E} \right)^2 + \frac{1}{N} \sum_{j=1}^{N} \hat{D}^2(\mathbf{s}_j) = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{E}(t) - \hat{E} \right)^2 + \frac{1}{n} \sum_{t=1}^{n} \hat{D}^2(t) \tag{1}$$

As a first step of the analysis of variance the comparison of the magnitude of the components in formulae (1) can perform.

In the next step the plotting of $\hat{E}(t), \hat{D}(t)$ $(t = 1,..,n)$ on graphs and mapping of $\hat{E}(\mathbf{s}_j)$, $\hat{D}(\mathbf{s}_j)$ $(j = 1,..,N)$ can support the visualization of the components of the variance.

Then the calculation of $\hat{E}$ and the different statistics and theirs roots listed in (2) can be applied for analyses of variance:

$$\hat{D}^2, \quad \frac{1}{N} \sum_{j=1}^{N} \left(\hat{E}(\mathbf{s}_j) - \hat{E} \right)^2, \quad \frac{1}{N} \sum_{j=1}^{N} \hat{D}^2(\mathbf{s}_j), \quad \frac{1}{n} \sum_{t=1}^{n} \left(\hat{E}(t) - \hat{E} \right)^2, \quad \frac{1}{n} \sum_{t=1}^{n} \hat{D}^2(t) \tag{2}$$

Built in ANOVA examination in MISH

The MISH method is developed for interpolation of meteorological data. An adequate mathematical background was also developed (*Szentimrey et al.*, 2011) for the purpose of efficient use of all the valuable meteorological and auxiliary model information. The main difference between MISH and the usual geostatistical interpolation methods is the application of the meteorological data series for modeling. In geostatistics (*Cressie*, 1991), the sample used for modeling is only the predictor data, which is a single realization in time, while in meteorology there are long term data series, i.e., sample is available in time and space as well.

Partitioning of Total Variance of station data series as it is applied in MISH is the following:

$$\hat{V} = \frac{1}{M} \sum_{i=1}^{M} \left(\hat{E}(\mathbf{s}_i) - \hat{E} \right)^2 + \frac{1}{M} \sum_{i=1}^{M} \hat{D}^2(\mathbf{s}_i) = \hat{S}^2_{space} + \hat{D}^2_{time} \quad ,$$

$\hat{S}^2_{space}$ is the variance of spatial trend, $\hat{D}^2_{time}$ is the mean temporal variance.

The *Table 1* shows an example of automatically generated output of MISH for Hungarian meteorological stations were used for derivation of CARPATCLIM over Hungary for daily maximum (Tx), daily minimum (Tn) temperatures and daily precipitation sum (R).

Table 1. MISH output for ANOVA, monthly results for Hungary (part of the CARPATCLIM)

	1	2	3	4	5	6	7	8	9	10	11	12
Tx												
D_t:	2.67	3.24	2.69	1.87	1.96	1.64	1.71	1.98	1.96	1.83	2.43	2.11
S_s:	1.00	1.23	1.33	1.21	1.31	1.34	1.37	1.39	1.43	1.34	1.21	1.02
Tn												
D_t:	2.76	2.88	1.86	1.35	1.20	1.12	1.21	1.18	1.29	1.67	1.97	2.12
S_s:	0.85	0.85	0.83	0.88	0.91	0.87	0.90	0.88	0.82	0.77	0.70	0.80
R												
D_t:	22.5	22.9	21.3	25.6	36.2	39.0	39.3	40.7	36.5	35.7	33.3	27.9
S_s:	7.1	5.9	6.9	7.8	7.8	8.9	9.3	10.5	10.2	8.3	10.8	8.6

4. THE AREA IN FOCUS

The target area of the CARPATCLIM between latitudes 50°N and 44°N, and longitudes 17°E and 27°E approximately, partly includes nine countries (*Figure 1*) in the region.

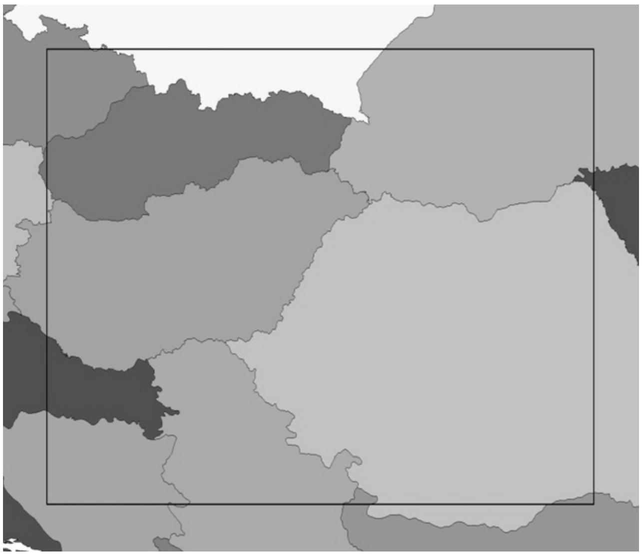


Fig. 1. The target area of the CARPATCLIM and the political boundaries.

The spatial resolution of the datasets in question are different. The E-OBS ***v14.0*** dataset is available on 25 km resolution grid although the CARPATCLIM dataset is derived for 0.1° (10 km approximately). To make the resolution similar for better comparison the CARPATCLIM dataset was set out for 0.2 degree (*Figure 2*) in the first step. According to the *Table II.* the most important statistical properties of the climate variables available on 0.1° and 0.2° grid are the same or very close, therefore the CARPATCLIM 0.2 degree can be used for comparison with E-OBS, instead of CARPATCLIM 0.1degree.

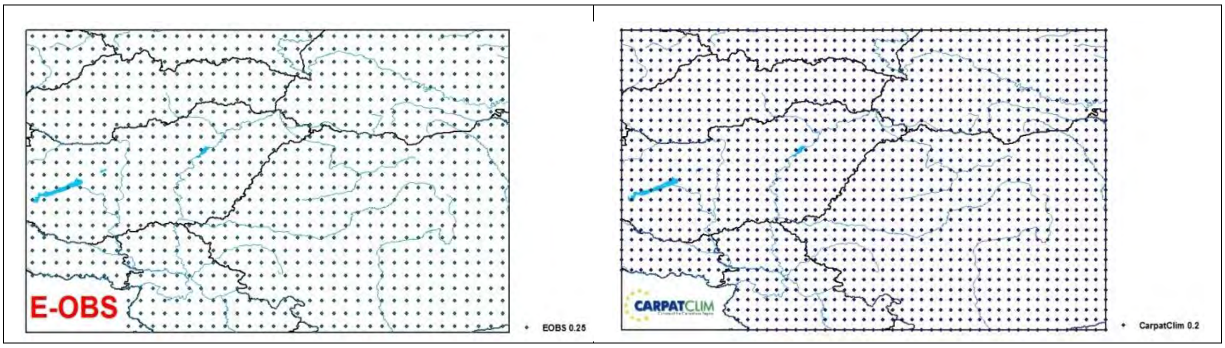


Fig. 2. The target and grid points of area for comparison

Table II. Comparison of different statistics of CARPATCLIM 0.1° and 0.2°, 1961-2010

Tx CC01	**Tx CC02**
Total mean: 13.83	Total mean: 13.82
Total variance: 6.90	Total variance: 6.97
Spatial st. deviation of temporal means: 2.48	Spatial st. deviation of temporal means: 2.49
Root spatial mean of temporal variances: 0.88	Root spatial mean of temporal variances: 0.88
Spatial mean of temporal st. deviations: 0.88	Spatial mean of temporal st. deviations: 0.88
Temporal st. deviation of spatial means: 0.83	Temporal st. deviation of spatial means: 0.83
Root temporal mean of spatial variances: 2.49	Root temporal mean of spatial variances: 2.50
Temporal mean of spatial st. deviations: 2.49	Temporal mean of spatial st. deviations: 2.50

Pr CC01	**PrCC02**
Total mean: 701.21	Total mean: 700.32
Total variance: 40565.66	Total variance: 41227.35
Spatial st. deviation of temporal means: 156.75	Spatial st. deviation of temporal means: 159.01
Root spatial mean of temporal variances: 126.47	Root spatial mean of temporal variances: 126.27
Spatial mean of temporal st. deviations: 124.51	Spatial mean of temporal st. deviations: 124.28
Temporal st. deviation of spatial means: 91.07	Temporal st. deviation of spatial means: 90.57
Root temporal mean of spatial variances: 179.64	Root temporal mean of spatial variances: 181.72
Temporal mean of spatial st. deviations: 178.17	Temporal mean of spatial st. deviations: 180.22

5. RESULTS

In the first step of the comparison of the E-OBS and CARPATCLIM 02, all the components in the formulae (1) were computed for further analysis. *Table III.-V.* show the list of the values can take into consideration in the analysis of variance for different seasons (winter: win, spring: sp, summer: sum, autumn: au) for both dataset CARPATCLIM02 (*C02*) and E-OBS (*EO*) for maximum temperatures (Tx), minimum temperatures (Tn) and precipitation (R) in the period of 1961-2010. Some of them can be illustrated on graphs and some of them on maps. Such as $\hat{E}(t)$, $\hat{D}(t)$ $(t=1,..,n)$ can be demonstrated on graphs. The maps of $\hat{E}(\mathbf{s}_j)$, $\hat{D}(\mathbf{s}_j)$ $(j=1,..,N)$ can be used for interpretation of the components of the variance in space.

Table III. Results of ANOVA for maximum temperatures, 1961-2010

	$\hat{E}$	$\hat{D}^2$	$\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2$	$\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)$	$\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2$	$\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2}$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)}$
winC02Tx	2.07	6.33	3.69	2.64	2.24	4.09	1.92	1.63	1.50	2.02
winEOTx	1.97	6.07	3.49	2.58	2.20	3.87	1.87	1.61	1.48	1.97
spC02Tx	14.46	9.83	7.76	2.07	1.87	7.96	2.79	1.44	1.37	2.82
spEOTx	14.57	7.65	5.61	2.04	1.80	5.85	2.37	1.43	1.34	2.42
suC02Tx	24.40	10.26	8.72	1.55	1.32	8.95	2.95	1.24	1.15	2.99
suEOTx	24.49	7.89	6.36	1.52	1.26	6.63	2.52	1.23	1.12	2.58
auC02Tx	14.36	7.73	6.16	1.57	1.32	6.41	2.48	1.25	1.15	2.53
auEOTx	14.38	6.56	4.91	1.65	1.39	5.17	2.22	1.29	1.18	2.27

Table IV. Results of ANOVA for minimum temperatures, 1961-2010

	$\hat{E}$	$\hat{D}^2$	$\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2$	$\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)$	$\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2$	$\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2}$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)}$
winC02Tn	-4.75	6.53	3.62	2.91	2.40	4.13	1.90	1.71	1.55	2.03
winEOTn	-4.77	6.70	3.84	2.86	2.37	4.34	1.96	1.69	1.54	2.08
spC02Tn	3.94	4.84	4.00	0.85	0.74	4.11	2.00	0.92	0.86	2.03
spEOTn	4.03	4.75	3.89	0.86	0.71	4.04	1.97	0.93	0.84	2.01
suC02Tn	12.74	4.82	4.07	0.75	0.68	4.14	2.02	0.87	0.83	2.03
suEOTn	12.81	4.83	4.07	0.76	0.62	4.21	2.02	0.87	0.79	2.05
auC02Tn	4.83	3.94	2.99	0.95	0.80	3.14	1.73	0.98	0.89	1.77
auEOTn	4.84	4.15	3.17	0.98	0.80	3.36	1.78	0.99	0.89	1.83

Table V. Results of ANOVA for precipitation, 1961-2010

	$\hat{E}$	$\hat{D}^2$	$\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2$	$\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)$	$\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2$	$\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{E}(\mathbf{s}_j)-\hat{E})^2}$	$\sqrt{\frac{1}{N}\sum_{j=1}^{N}\hat{D}^2(\mathbf{s}_j)}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{E}(t)-\hat{E})^2}$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{D}^2(t)}$
winC02R	126.06	3638.55	1469.51	2169.04	1108.65	2529.90	38.33	46.57	33.30	50.30
winEOR	114.03	2946.99	935.48	2011.50	1024.24	1922.72	30.59	44.85	32.00	43.85
spC02R	169.49	4135.46	1534.45	2601.01	1033.17	3102.26	39.17	51.00	32.14	55.70
spEOR	154.75	3438.57	968.89	2469.68	1066.01	2372.55	31.13	49.70	32.65	48.71
suC02R	249.57	9446.81	3870.71	5576.09	2661.01	6785.72	62.22	74.67	51.59	82.38
suEOR	227.67	8403.98	3062.14	5341.84	2507.50	5896.50	55.34	73.09	50.07	76.79
auC02R	155.19	4866.06	1396.40	3469.66	1846.69	3019.40	37.37	58.90	42.97	54.95
auEOR	143.25	4283.28	1017.43	3265.85	1660.77	2622.50	31.90	57.15	40.75	51.21

The total variance can be partitioning as the sum of the spatial variance of the temporal means and the spatial mean of the temporal variances on the left side of the equation (1) on one hand; and on the other hand the sum of the temporal variance of the spatial means and the temporal mean of the spatial variances on the right side of the equation (1). The graphs here show the curves of $\hat{E}(t)$: spatial mean at the moment *t* and $\hat{D}(t)$: spatial standard deviation at the moment *t*. The maps represent the $\hat{E}(\mathbf{s}_j)$: temporal mean at location $\mathbf{s}_j$ and $\hat{D}(\mathbf{s}_j)$: temporal standard deviation at location $\mathbf{s}_j$. The *Figures 3-11* show the results for the annual, spring and summer maximum temperatures for the whole 50 years long time interval. The spatial means are going along together, however the spatial standard deviation are lower in the case of E-OBS, particularly in spring and in summer. The main features of the annual maps for maximum temperatures are similar, though greater differences appear on seasonal maps, mainly in temporal standard deviation (*Figure 8, 11 and 14 right panel*). The curves of spatial mean agree (*Figure 12*). In the case of annual minimum temperatures the spatial standard deviation in E-OBS exceeds the values which characterize the CARPATCLIM (*Figure 13*). The ANOVA results point out greater discrepancies related to precipitation in general. There is higher amount of annual precipitation sum shown in CARPATCLIM than in E-OBS (*Figure 15*). The curves of spatial standard deviation intersect around early nighties and the spatial standard deviation of CARPATCLIM pass the E-OBS spatial variance (*Figure 16*) then. The orographic features more obviously appear in the CARPATCLIM than in E-OBS on the maps representing the temporal mean and standard deviation. Some maps are shown here between 1961-2010 and 1991-2010 annual (*Figure 17-19 and Figure 24-26*) too. The seasonal results for spring, summer and autumn agree on less precipitation in E-OBS and lower temporal standard deviation (*Figure 20-28*).

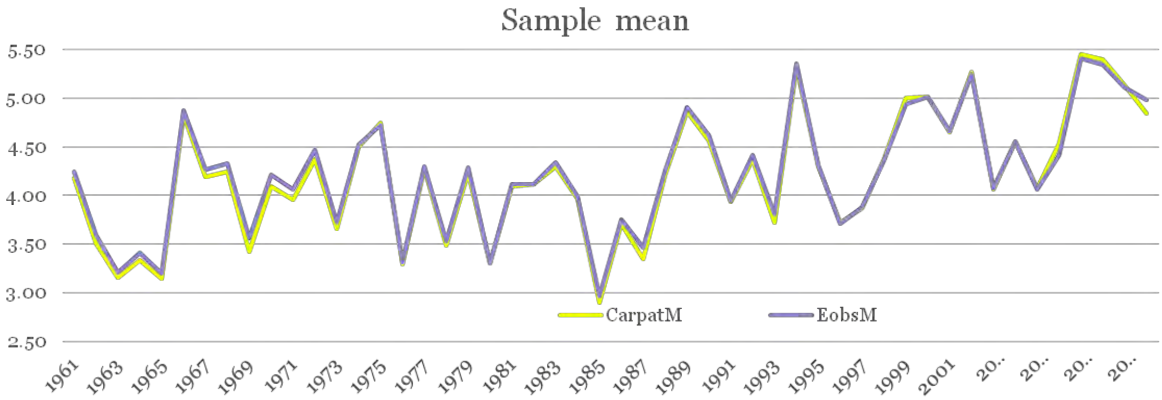


Fig. 3. Spatial mean of annual average maximum temperatures for CarpatClim02 (yellow) and E-OBS (blue) datasets from 1961 to 2010.

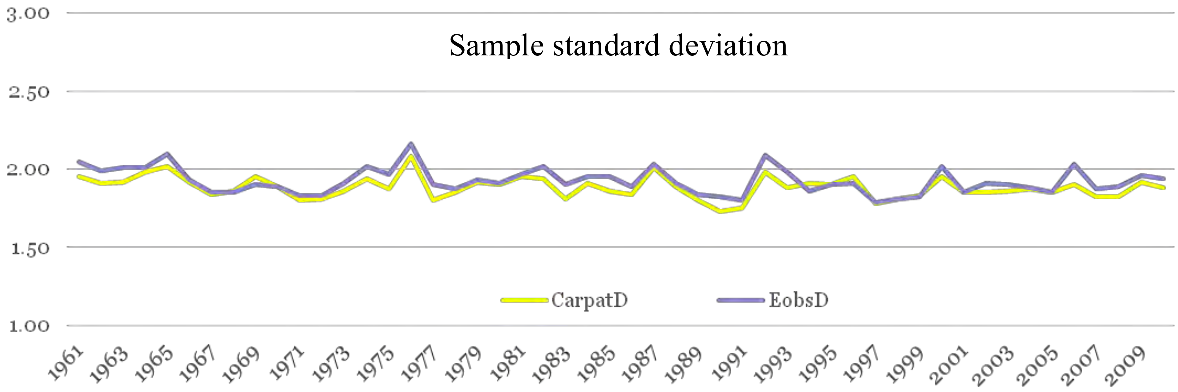


Fig. 4. Spatial standard deviation of annual average maximum temperatures for CarpatClim02 (yellow) and E-OBS (blue) datasets from 1961 to 2010.

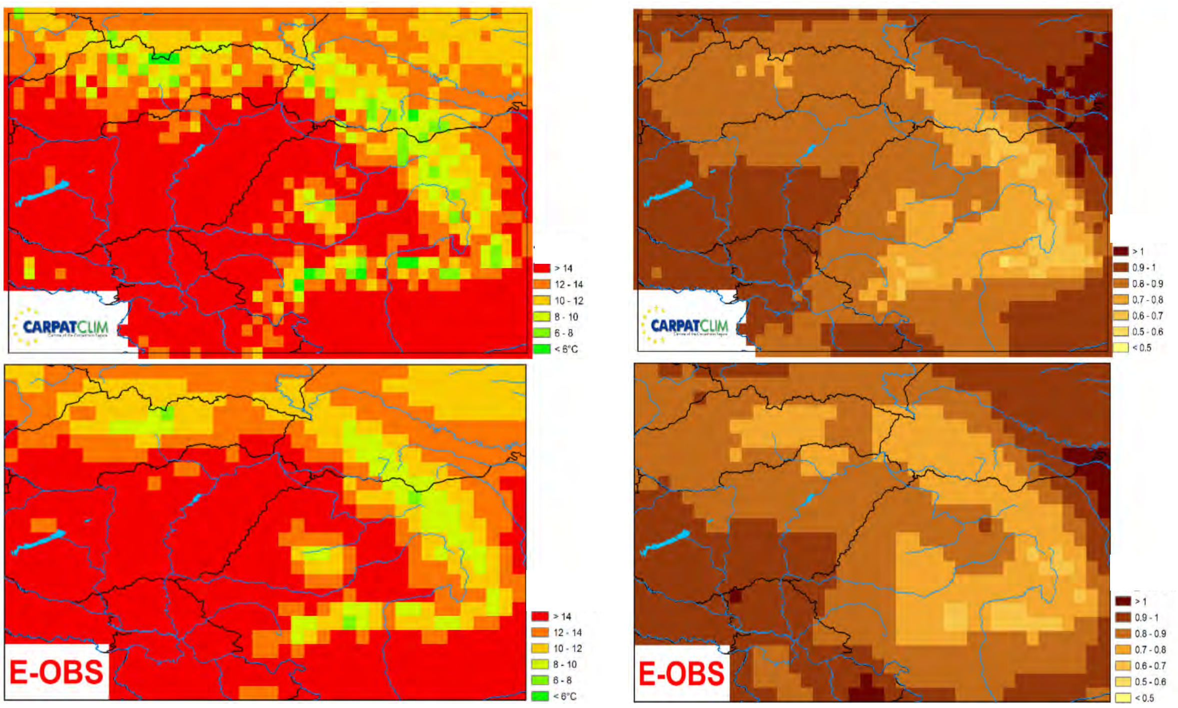


Fig. 5. Temporal mean (left) and temporal standard deviation (right) of the annual average maximum temperatures for CarpatClim02 and E-OBS grid points in the period of 1961-2010.

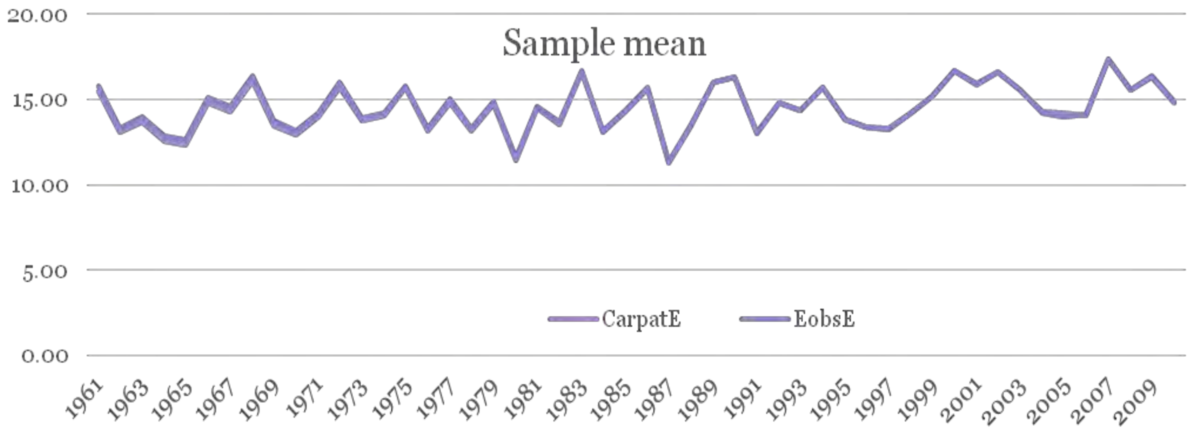


Fig. 6. Spatial mean of the spring average maximum temperatures for CarpatClim02 (purple) and E-OBS (blue) datasets from 1961 to 2010.

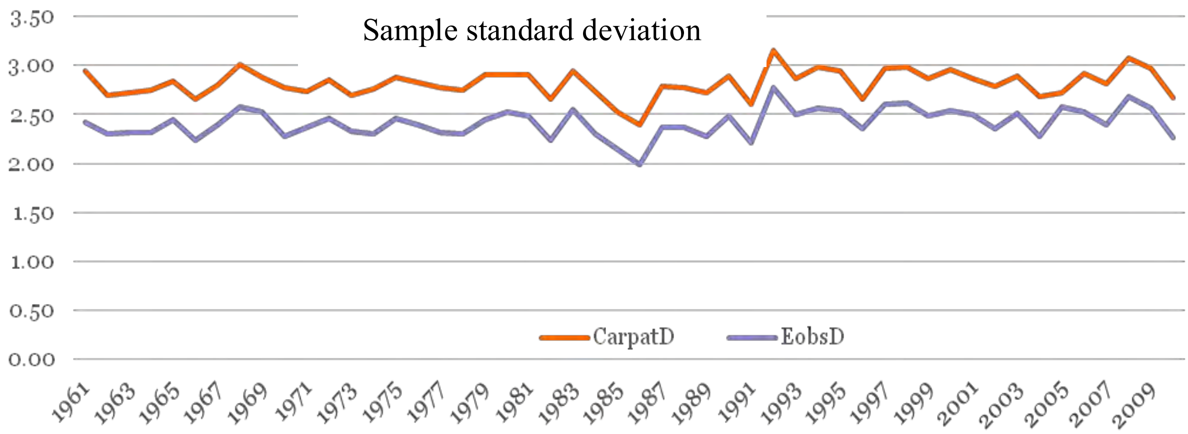


Fig. 7. Spatial standard deviation of the spring average maximum temperatures for CarpatClim02 (orange) and E-OBS (blue) datasets from 1961 to 2010.

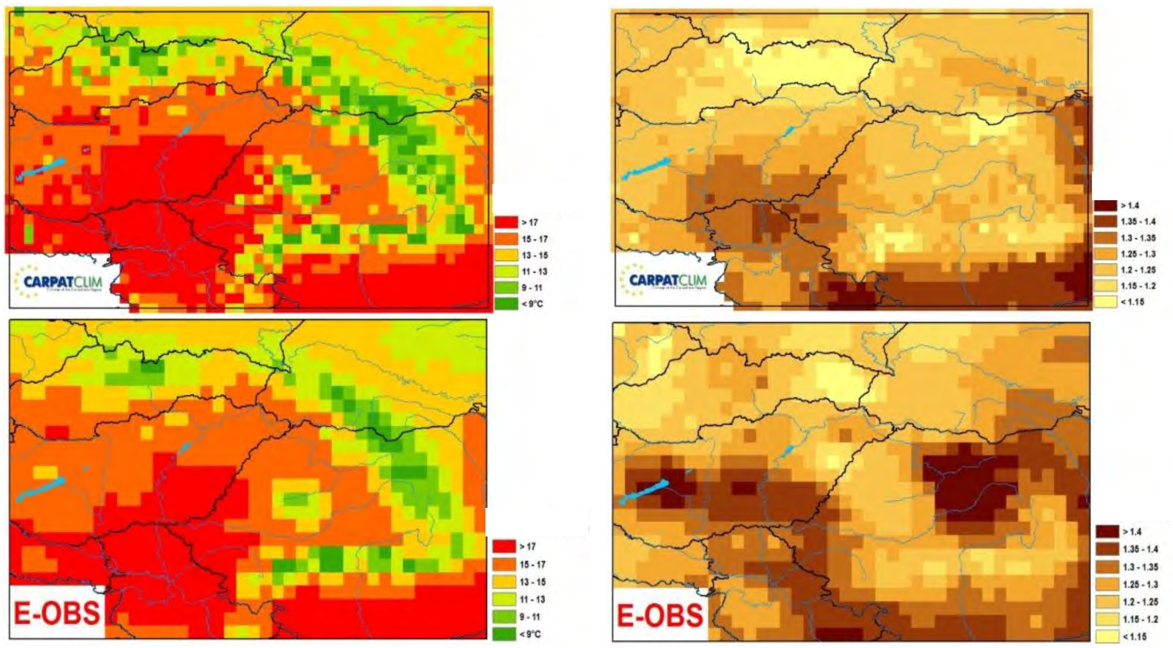


Fig. 8. Temporal mean (left) and temporal standard deviation (right) of the spring average maximum temperatures for CarpatClim02 and E-OBS grid points in the period of 1991-2010.

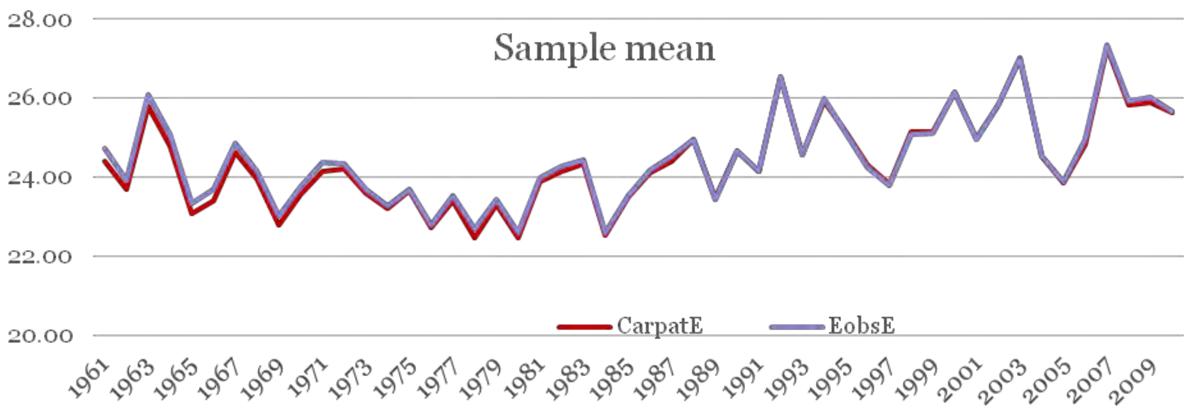


Fig. 9. Spatial mean of the summer average maximum temperatures for CarpatClim02 (dark red) and E-OBS (blue) datasets from 1961 to 2010.

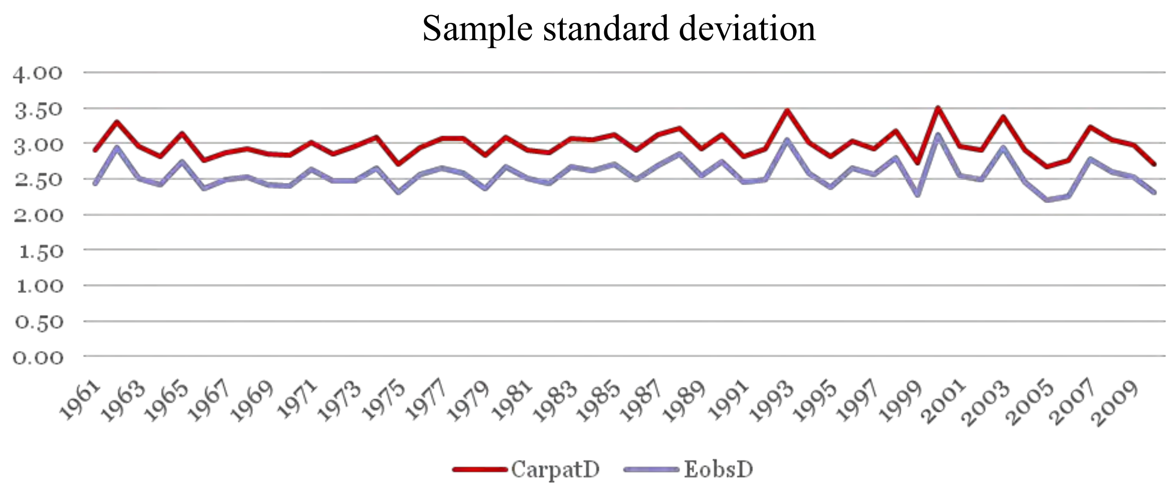


Fig. 10. Spatial standard deviation of the summer average maximum temperatures for CarpatClim02 (dark red) and E-OBS (blue) datasets from 1961 to 2010.

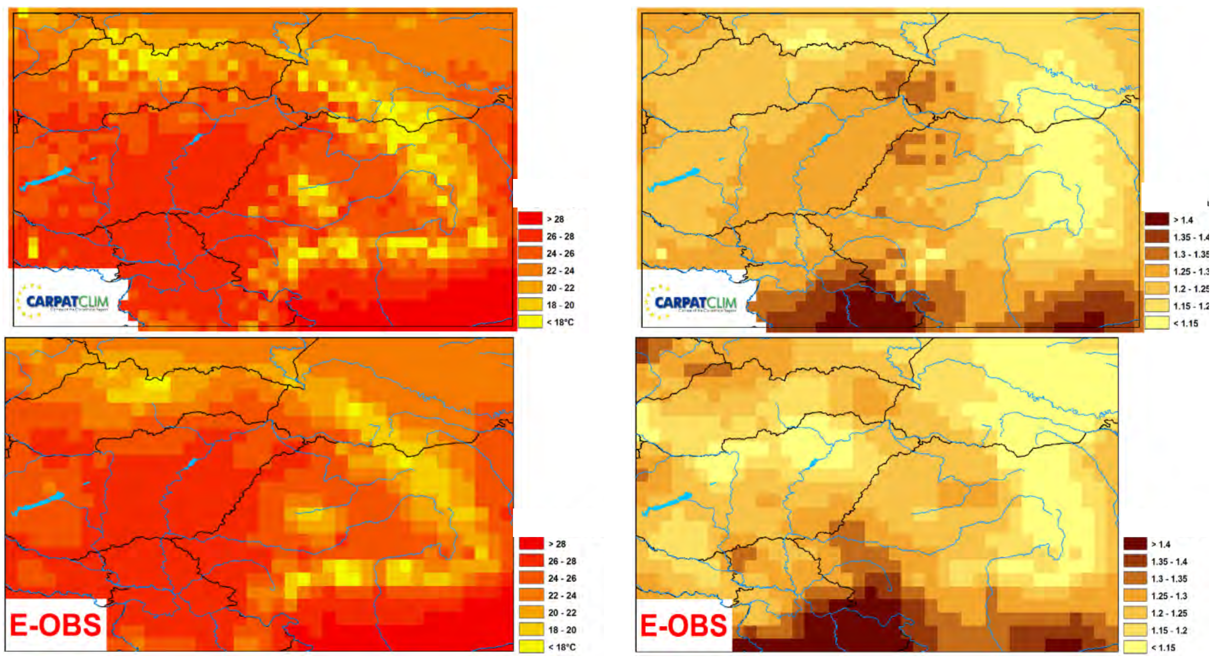


Fig. 11. Temporal mean (left) and temporal standard deviation (right) of the summer average maximum temperatures for CarpatClim02 and E-OBS grid points in the period of 1961-2010.

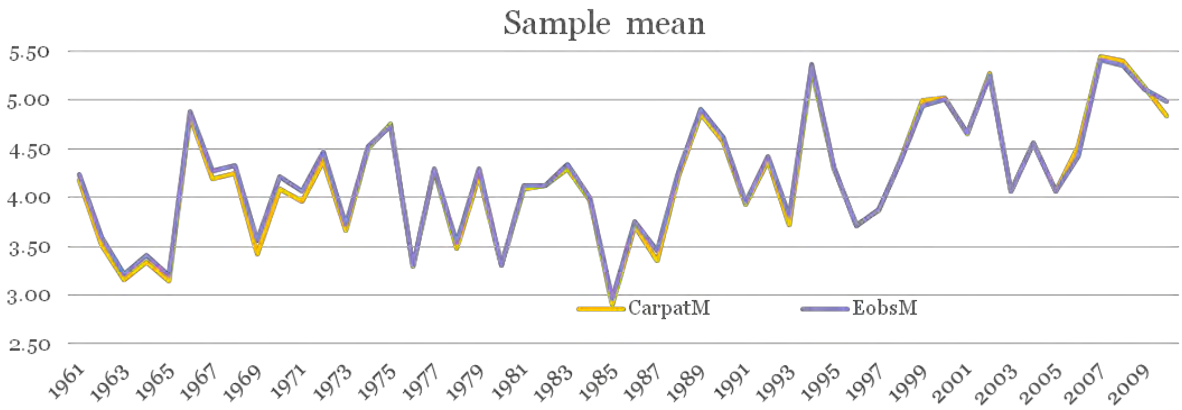


Fig. 12. Spatial mean of the annual average minimum temperatures for CarpatClim02 (yellow) and E-OBS (blue) datasets from 1961 to 2010.

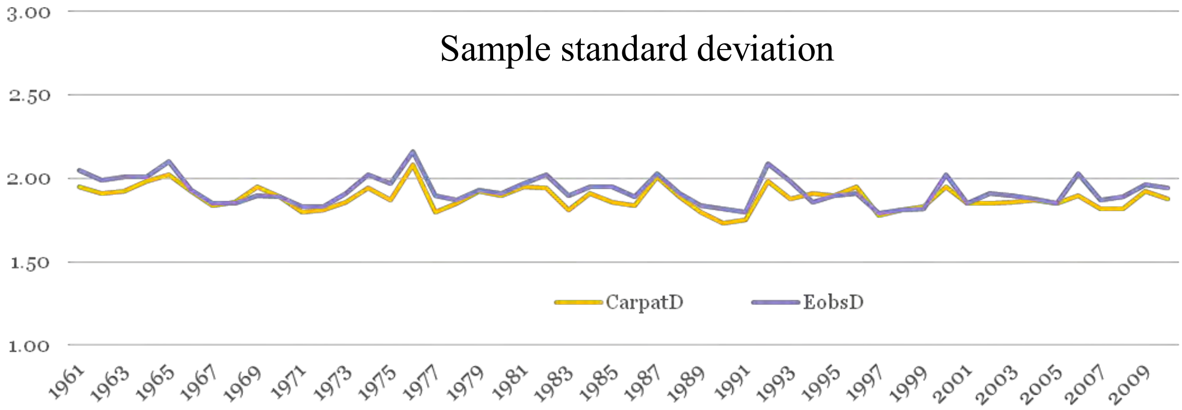


Fig. 13. Spatial varia standard deviation nce of the annual average minimum temperatures for CarpatClim02 (ochre) and E-OBS (blue) datasets from 1961 to 2010.

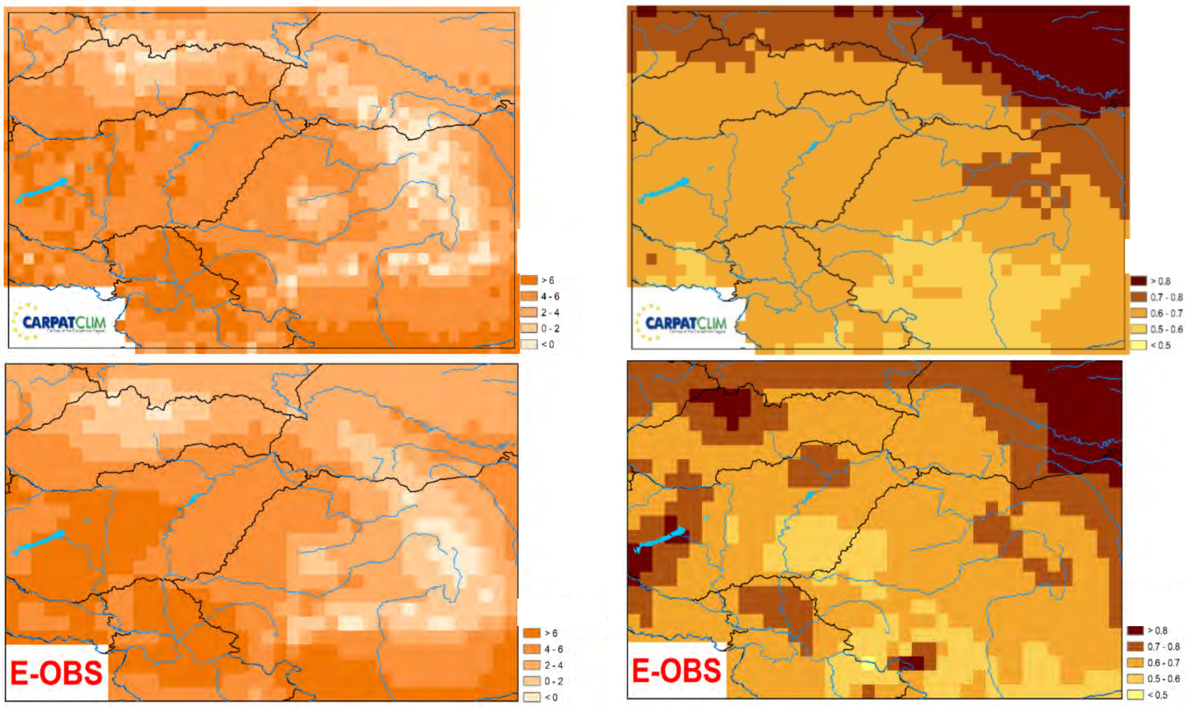


Fig. 14. Temporal mean (left) and temporal standard deviation (right) of the annual average minimum temperatures for CarpatClim02 and E-OBS grid points in the period of 1961-2010.

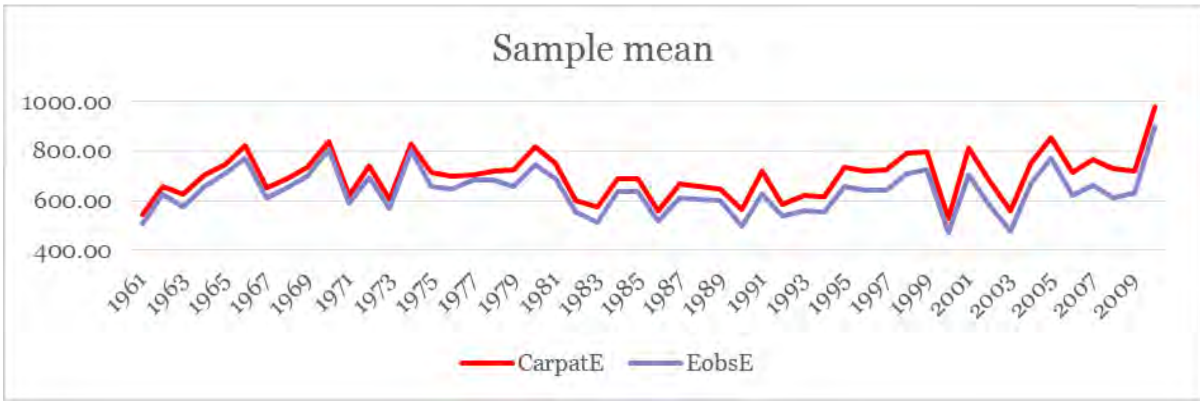


Fig- 15. Spatial mean of the annual precipitation sum for CarpatClim02 (red) and E-OBS (blue) datasets from 1961 to 2010.

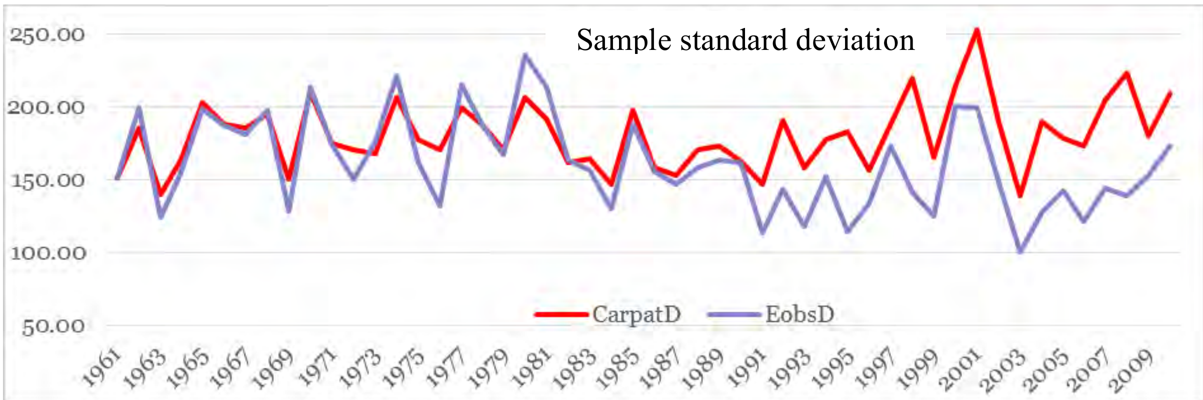


Fig. 16. Spatial standard deviation of the annual precipitation sum for CarpatClim02 (red) and E-OBS (blue) datasets from 1961 to 2010.

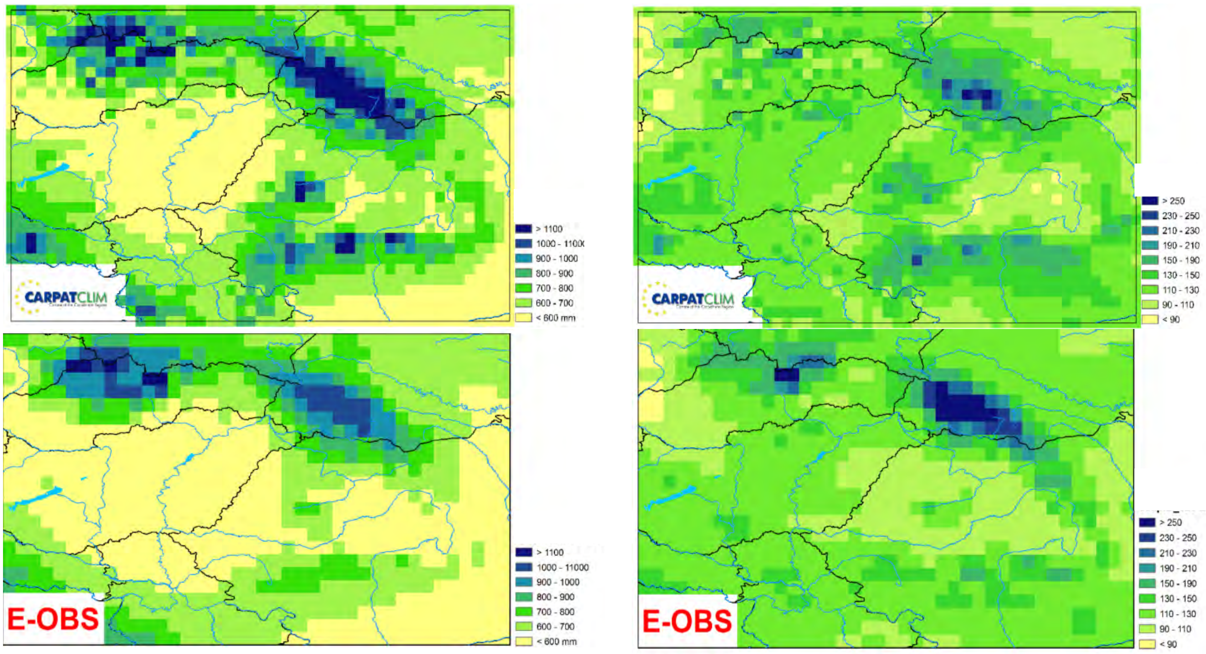


Fig. 17. Temporal mean (left) and temporal standard deviation (right) of the annual average precipitation sum for CarpatClim02 and E-OBS grid points in the period of 1961-2010.

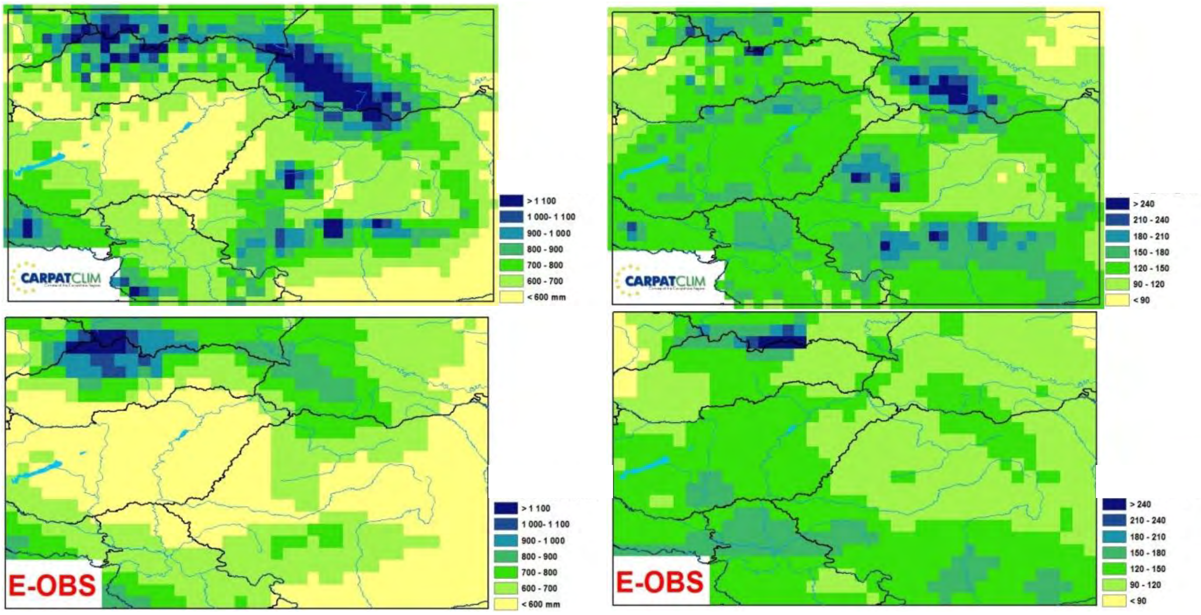


Fig. 18. Temporal mean (left) and temporal standard deviation (right) of the annual average precipitation sum for CarpatClim02 and E-OBS grid points in the period of 1991-2010.

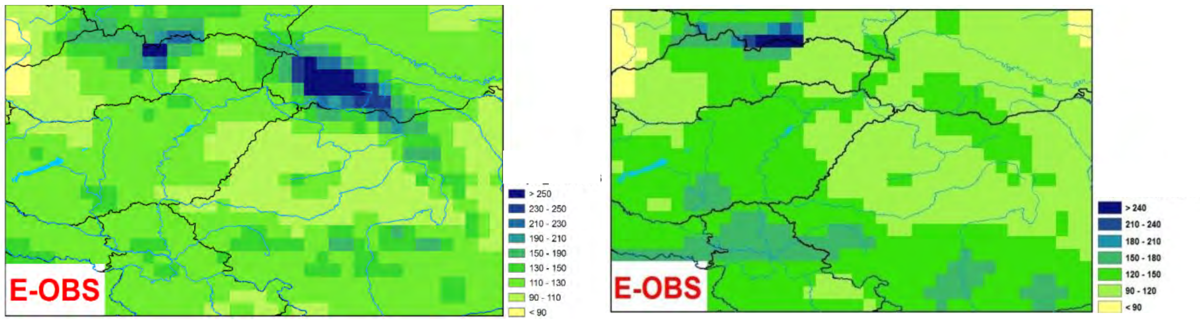


Fig. 19. Temporal standard deviation of the annual average precipitation sum for E-OBS grid points in the period of 1961-2010 (left) and 1991-2010 (right).

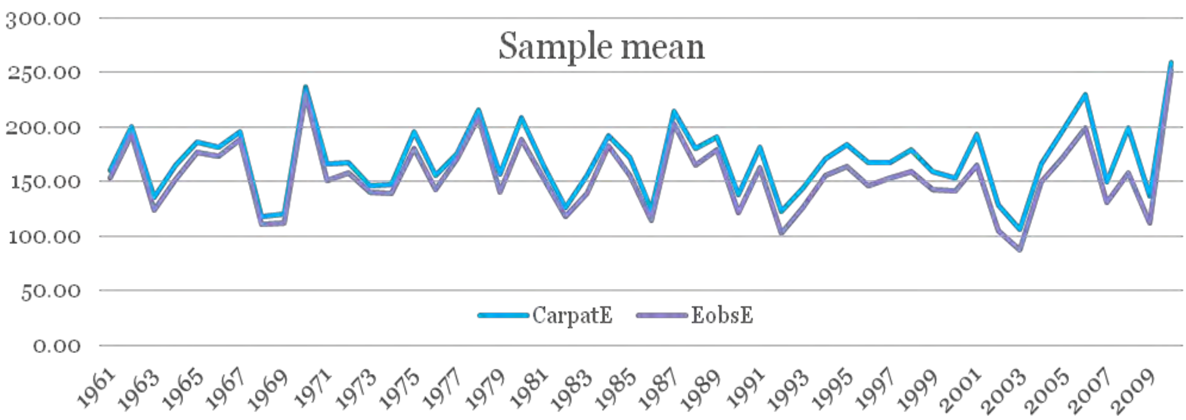


Fig. 20. Spatial mean of the spring precipitation for CarpatClim02 (pale blue) and E-OBS (blue) datasets from 1961 to 2010.

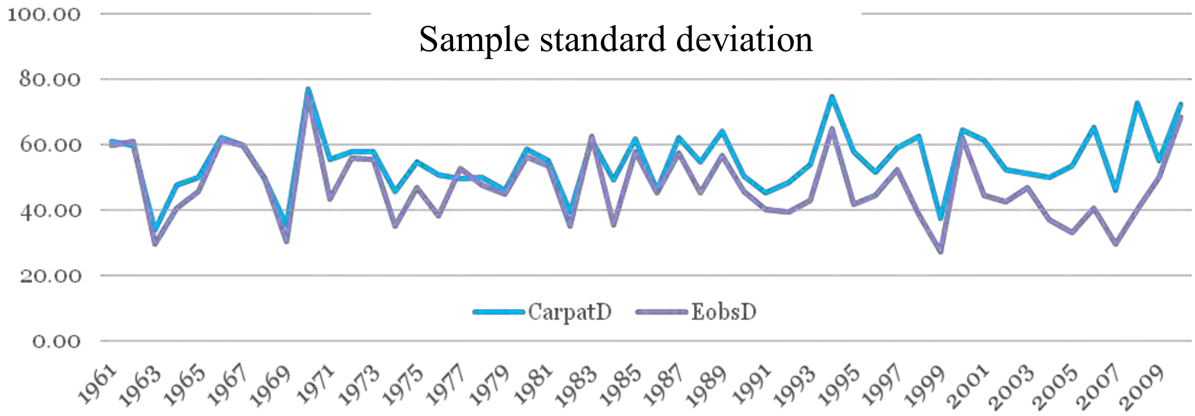


Fig. 21. Spatial standard deviation of the spring precipitation for CarpatClim02 (pale blue) and E-OBS (blue) datasets from 1961 to 2010.

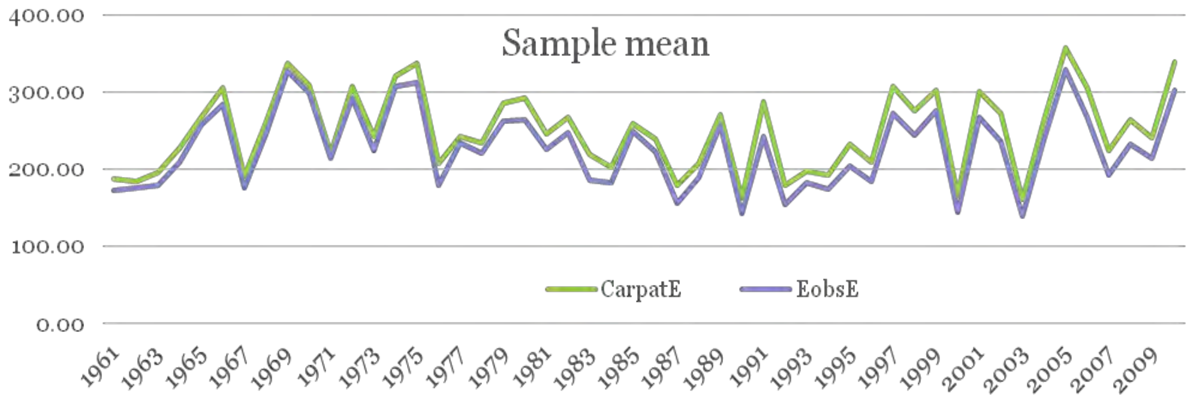


Fig. 22. Spatial mean of the summer precipitation for CarpatClim02 (green) and E-OBS (blue) datasets from 1961 to 2010.

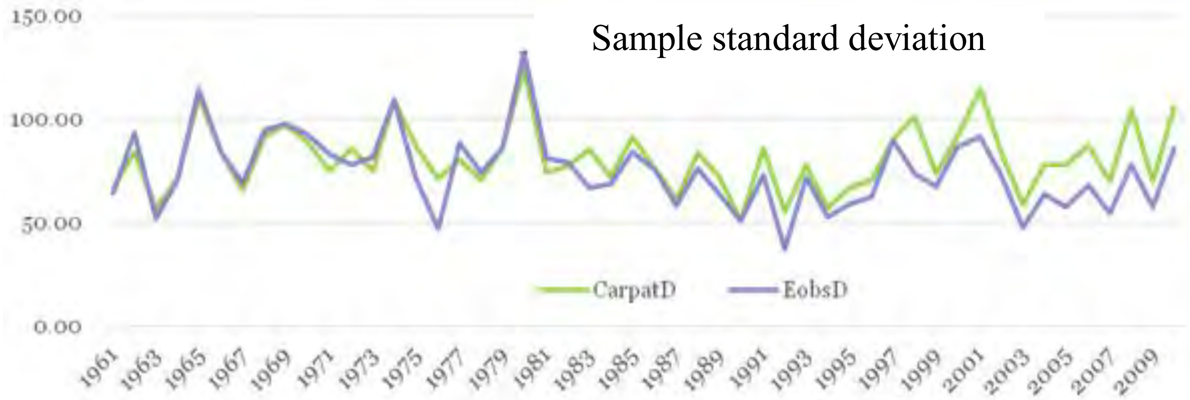


Fig. 23. Spatial standard deviation of the summer precipitation for CarpatClim02 (green) and E-OBS (blue) datasets from 1961 to 2010.

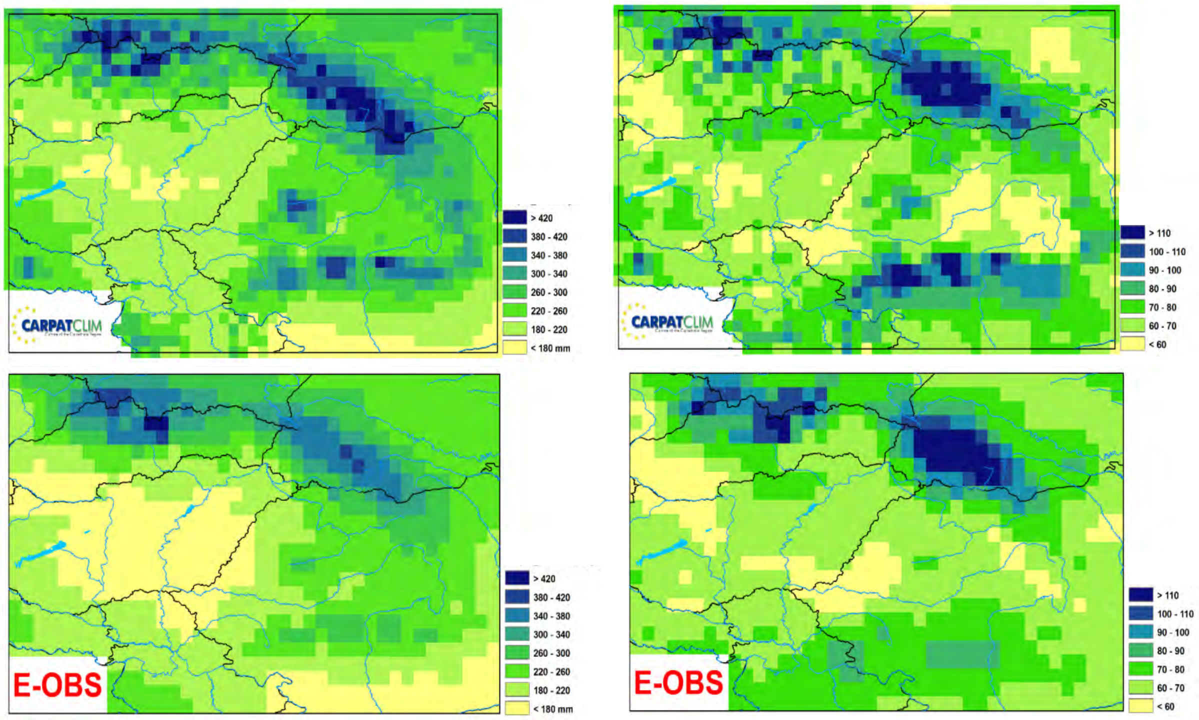


Fig. 24. Temporal mean (left) and temporal standard deviation (right) of the summer precipitation for CarpatClim02 and E-OBS grid points in the period of 1961-2010.

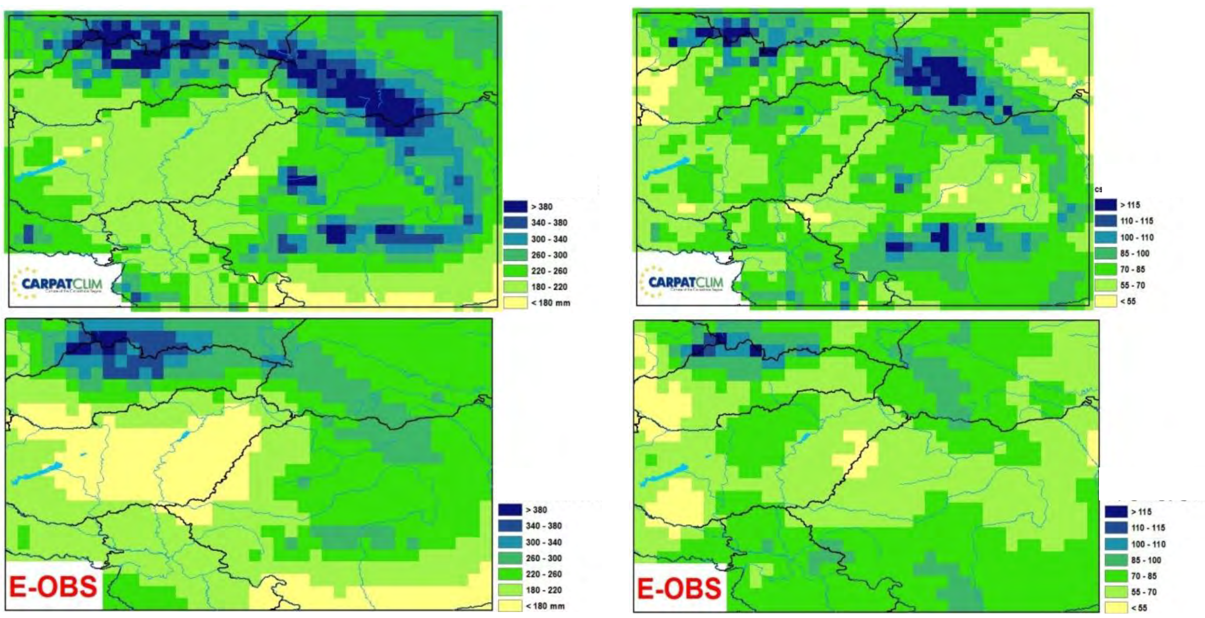


Fig. 25. Temporal mean (left) and temporal standard deviation (right) of the summer precipitation for CarpatClim02 and E-OBS grid points in the period of 1991-2010.

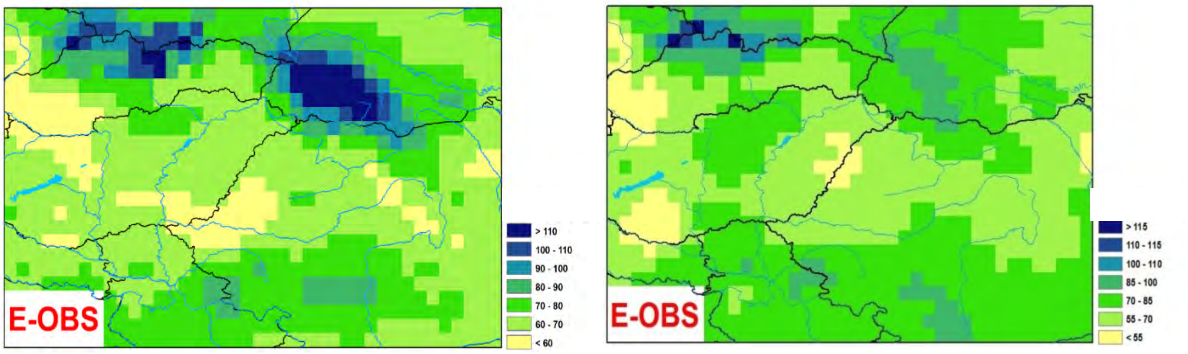


Fig. 26. Temporal standard deviation of the summer precipitation for E-OBS grid points in the period of 1961-2010 (left) and 1991-2010 (right).

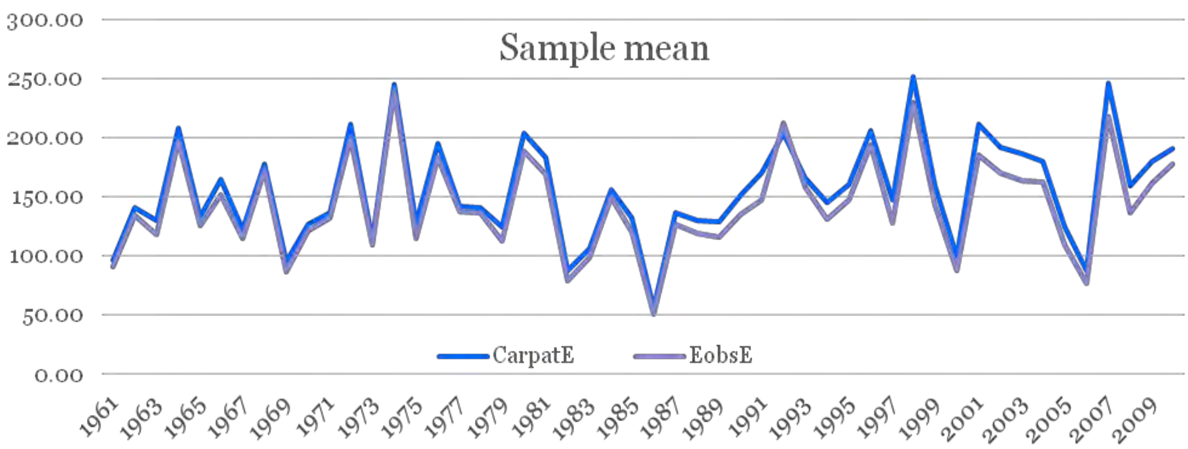


Fig. 27. Spatial mean of autumn precipitation for CarpatClim02 (bright blue) and E-OBS (blue) datasets from 1961 to 2010.

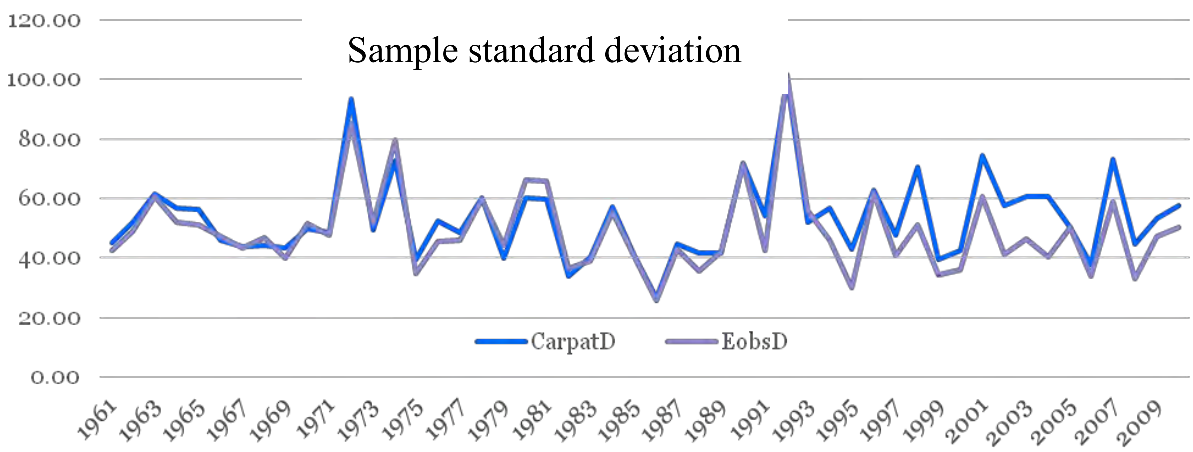


Fig. 28. Spatial standard deviation of autumn precipitation for CarpatClim02 (bright blue) and E-OBS (blue) datasets from 1961 to 2010.

6. CONCLUSION

The statistical properties of the CARPATCLIM and E-OBS gridded datasets were compared for the Carpathian Region on different time scales for the common area in the period of 1961-2010. For the comparison of these datasets a general statistical method: the Analysis of Variance (ANOVA) was applied. The graphs and maps shown in this study let us examine the spatial mean and spatial standard deviation and the temporal mean and temporal standard deviation for both datasets. The largest differences emerged between the standard deviation of the two datasets emerged in seasonal scales and particularly for precipitation. One of the main output of this work is that the graphs and maps generated here are available for further analysis from now. We can conclude that the ANOVA can be used efficiently for characterization and for comparison of certain spatiotemporal statistical properties of the several datasets.

Acknowledgement

Special thanks to Peter Szabo (Hungarian Meteorological Service) for preparation of the E-OBS data for this study.

References

Cressie, N., 1991: Statistics for Spatial Data. Wiley, New York.

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

Szalai, S., Auer, I., Hiebl, J., Milkovich, J., Radim, T. Stepanek, P., Zahradnicek, P., Bihari, Z., Lakatos, M., Szentimrey, T., Limanowka, D., Kilar, P., Cheval, S., Deak, Gy., Mihic, D., Antolovic, I., Mihajlovic, V., Nejedlik, P., Stastny, P., Mikulova, K., Nabyvanets, I., Skyryk, O., Krakovskaya, S.,Vogt, J., Antofie, T., Spinoni, J. 2013: Climate of the Greater Carpathian Region. Final Technical Report.www.carpatclim-eu.org.

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

Szentimrey, T. and Bihari, Z., 2007: Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis). Proceedings from the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2004, COST Action 719, COST Office, 17–27.

Szentimrey, T., 2016: The ANOVA method and its applications. (Internally used document, in Hungarian)

COMPARISON OF MONTHLY SATELLITE, MODELLED AND IN SITU SURFACE RADIATION DATA OVER HUNGARY

Ildikó Dobi

Hungarian Meteorological Service (OMSZ)
1024 Budapest, Kitaibel P. u. 1, 36-1-3464710, fax:36-1-364-4665, e-mail: dobi.i@met.hu

Abstract

Knowledge of accurate surface incoming solar radiation is important for both climate studies and solar power applications. Three kind of radiation data sets are basically available for Hungary: in situ surface observations of OMSZ radiation network, modelled data from CarpatClim project, and satellite data from EUMETSAT Climate SAF. The aim of this paper is to compare these parallel sets of data and to determine their accuracy. Satellite retrievals give high resolution retriavals in space and time. The longest radiation data set dedicated for solar application for Europe is the Surface Solar Radiation Data Set - Heliosat (SARAH) developed by EUMETSAT Climate SAF. This includes a satellite-based climatology of the solar Surface Incoming Shortwave radiation (SIS). The other source is the outcomes of the CarpatClim Project (http://www.carpatclim-eu.org). The monthly global radiation characterizing the incoming solar energy (CC) is one of the fifty-three parameters. Remote sensed and modelled data sets are validated with set of eight surface measurements (OBS) from the network of OMSZ. Comparisons are made by pairs of data sets between 2001 and 2010. Mean Absolute Difference between parallel data set are between 5.4 and 6.5 W/m^2, anomaly corrections are above 0.9 W/m^2in all cases

Keywords: global radiation, EUMETSAT CM-SAF, CarpatClim project

1. INTRODUCTION

Accurate solar information are essential for climate change investigations and for solar energy systems. According to international experiences solar users need data in preliminary stage of investment for feasibility and resource assessment, or system siting and design. Solar applications require accurate direct normal, diffuse horizontal and global horizontal irradiance values (DNI, DHI and GHI) in fine time and space resolution (*Sengupta et. al.,* 2015).

This paper focus on GHI parameter of three main resources: ground measurements, satellite data and modelled data. All have advantages and limitations, which influence the reliability of data. Ground measurements have high accuracy and high time resolutions, but they are scarse in space, and sensor failures, or even spoiling of sensor can cause significant mistakes. Satellite images have high spatial resolution and long term data although characterized by lover time resolution and lower accuracy. Modelled data are particularly useful under limited availability of observed data, taking account that result strongly depends on the model construction.

EUMETSAT (www.eumetsat.org) Climate SAF provides long term radiation data set for solar energy sector. Surface global radiation analogous to surface incoming shortwave radiation (SIS), which are presently included in four product families (MVIRI, CLAAS, CLARA-A and SARAH). CLARA-A use polar orbit (AVHRR) satellites, while all of the others are evolution of Meteosat MVIRI/SEVIRI images, and the latest one, called SARAH (Solar surface Radiation Heliosat). All differs from each other in resolution, in spatial and temporal coverages. Comparison of monthly SIS derived from MVIRI and CLARA-A for eight stations in Hungary (*Rusznyak et. al*, 2014) resulted systematically better estimation in case of geostationary satellite. Consequently SARAH are selected as satellite data set. The other source of grid data set is CarpatClim Project which produced uniformly harmonized and interpolated gridded climate data set for the whole Pannonian Basin including Hungary. The project applied MASH (Multiple Analysis of Series for Homogenization) for homogenization (*Szentimrey*, 2011) and MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) for interpolation (*Szentimrey* and *Bihari*, 2006) the surface sunshine duration data. The aim of the study to quantify the accuracy. The global radiation data measured by OMSZ radiation network are used as reference or "true" data.

2. DATA DESCRIPTION

Full SARAH-E 1.0 SIS data set covers METEOSAT East with resolution 0.05^o x 0.05^o from January 1983 to December 2013. Hungary are covered by the pixels (49°N-45°S, 16°E-24°W).
SIS means the radiation flux (irradiance) reaching a horizontal plane at the Earth surface in the 0.2 – 4 µm wavelength region. Parameter is retrieved using the Heliosat method, which gathers satellite information of atmospheric composition (ozone, water vapour, aerosol) and apply'clear sky' method to calculate fraction of direct and diffuse irradiance. Method calculates the cloud index as the difference between actual reflectivity of the earth as it is seen by the satellite and a reference image which only includes reflectance of the ground (*Müller et al*, 2015).

Global radiation parameter of CarpatClim Project covers (44°N and 50°N, and longitudes 17°E and 27°E) with resolution 0.1^o x 0.1^o from January 1961 to December 2010. Method for area of Hungary – as defined in the project – used accumulated daily records of sunshine duration (hours) and latitudes of the stations (*CarpatClim Project*, 2012). Global radiation was derived using the equation postulated by *Angström* (1924) and modified by *Prescott* (1940).

The observed global monthly radiation amounts of eight radiation stations were extracted from the OMSZ observations ranging from the year 2001-2010. This ten years long period were selected for comparison in case of the three data sets. All the ground stations chosen for this study have the best quality measurements and scattered over the country (*Table 1.*)

Table 1. List of radiation station used in this study

Station	Station number	Latitude	Longitude	Height
Baja, Csávoly	48101	46.18	19.01	112.0
Debrecen	64704	47.49	21.61	107.6
Eger	53206	49.49	20.39	220.9
Győr	23703	47.71	17.67	116.7
Kecskemét K-puszta	46304	46.97	19.55	126.9
Budapest-Lőrinc	44527	47.43	19.18	139.1
Nagykanizsa	17809	46.46	16.97	139.8
Szeged	58113	46.26	20.09	81.8

There are three practical differences between satellite and other data resources:

1.) Satellites detect irradiance, which is the *power* (or flux) of the sunshine received per area (W/m^2) opposed to the surface station (and modelled data) measuring irradiation as *energy* received per area (MJ/m^2/day).

2.) In case of satellite data the monthly averages are calculated from the daily *means* of the given month on pixel basis as arithmetic mean with a required minimal number of 20 existing daily means. The other data sets use monthly amounts, as *sum* of complete hourly data measured during a month (without any missing data).

3.) Satellite retrieval provide an average value of 25 km^2 area. Each pixel (grid) value was compared with observed point data covered by the given area. A grid from CarpatClim represents global radiation value referring to the left upper corner point values of the 10 km * 10 km area.

3. METHODOLOGY

Most of the accuracy assessments uses first order statistics for verifications, like bias (BIAS), mean absolute bias (MAB), standard deviation (SD) and anomaly correlation (AC). The applied quality measures are describes below (*Müller et. al*, 2015). The variable *y* describes the data set to be validated (SARAH or CarpatClim), and *o* denotes the reference observed data set. The individual time step is marked with *k* and *n* is the total number of time steps.

BIAS: The mean error give the mean difference between the two time series and indicate the data set on average over – or underestimates the reference data set.

$$\text{Bias} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)$$

MAB: Mean absolute bias is the average of the absolute values of the differences between each member of the time series. The advantage of it is that there is no cancellation of positive and negative bias values.

$$\text{MAB} = \frac{1}{n}\sum_{k=1}^{n}|y_k - o_k|$$

SD: Standard deviation measures the spread around the expected value of distribution.

$$\text{SD} \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}((y_k - o_k) - (\bar{y} - \bar{o}))^2}$$

AC: Anomaly Correlation Coefficient (ACC) is one of the most widely used measures in the verification of spatial fields. AC will take the maximum value of 1. In turn, if the variation pattern is completely reversed, AC takes the minimum value of -1.

$$\text{AC} = \frac{\sum_{k=1}^{n}(y_k - \bar{y})(o_k - \bar{o})}{\sqrt{\sum_{k=1}^{n}(y_k - \bar{y})^2}\sqrt{\sum_{k=1}^{n}(o_k - \bar{o})^2}}$$

FRAC > 15 W/m^2: A measure of the uncertainty of the data set is the fraction of the time steps that are above 15 W/m^2.

$$\text{Frac} = 100\,\frac{\sum_{k=1}^{n} f_k}{n} \text{ with } \begin{cases} f_k = 1 \; if \; y_k > Th \\ f_k = 1 \; otherwise \end{cases}$$

4. RESULTS AND DISCUSSION

Spatial distribution of the mean global radiation from 1961 to 2010 (see *Figure 1*) illustrates the outcome of CarpatClim project. In southeast part of the country shows the largest solar radiation, contrarily hilly areas at north and eastern part of Hungary characterized by the lowest radiation income. *Figure 2* shows satellite data represented by SIS mean for the full 31 years long period demonstrating increasing insolation from north to south. The map on *Figure 3.* indicates the differences between satellite and modelled global radiation values at 10 km * 10 km spatial resolution. Time range from January 2001 to December 2010 was selected for all verification. The largest radiation is detected at east part of Hungary, near to the borders, while the smallest one occurred at the hilly areas.

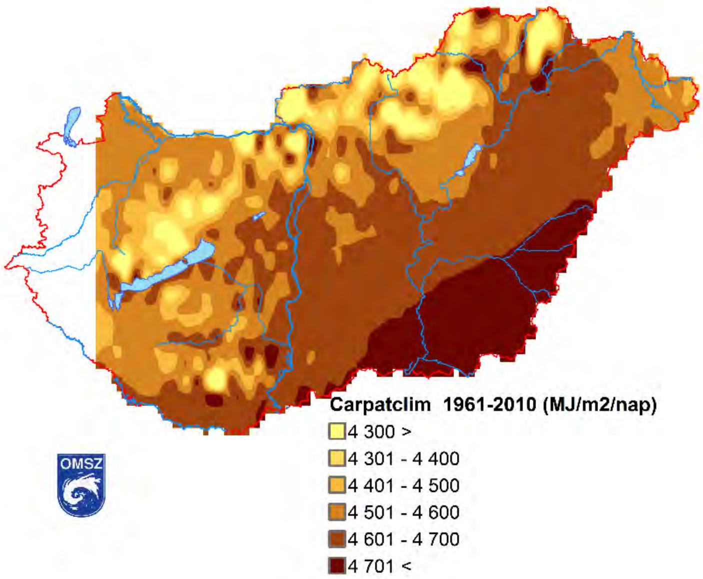


Fig. 1. Multiannual mean surface radiation from monthly data set of CarpatClim project (1961-2010)

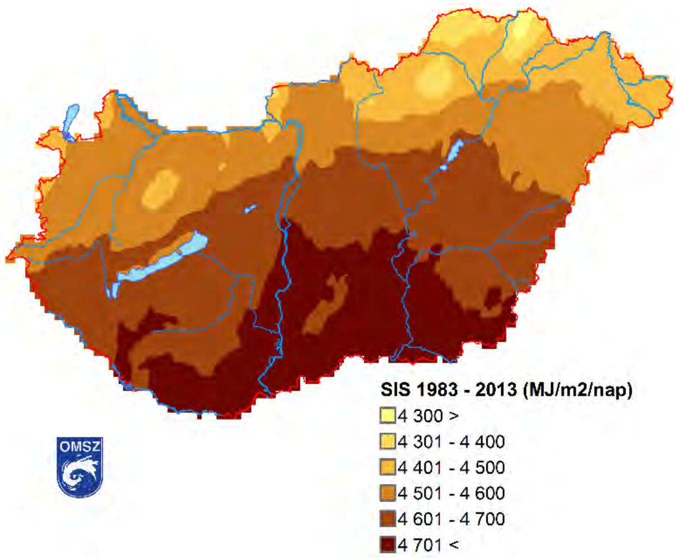


Fig. 2. Multiannual mean SIS from EUMETSAT Climate SAF SIS monthly data set (1983-2013)

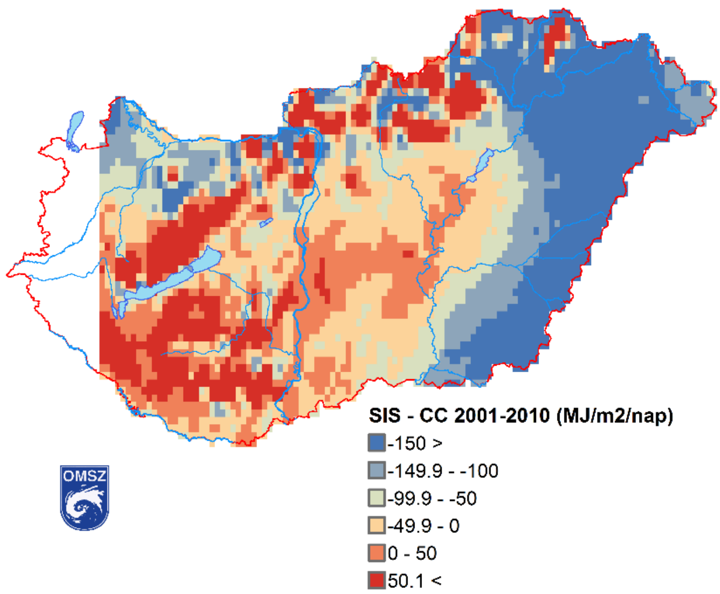


Fig. 3. Difference between satellite SIS and CarpatClim monthly data sets (2001-2010)

Figure 4 shows the time series of satellite (SIS), modelled (CC) and station (Obs) data based on the mean monthly values of the eight selected point (grid) values. Satellite fails at peaks, in summer overestimate and in winter systematically underestimate the "true" radiation values. Next *Figure* (*5.)* also detects significant overestimation in case of Meteosat retrievals. The upper quartiles of the differences are below 10 W/m^2 in all cases.

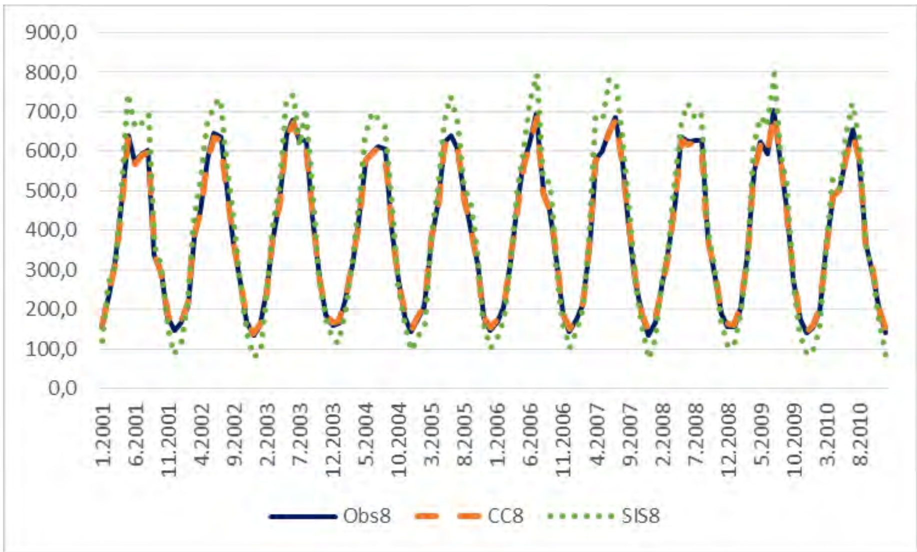


Fig. 4. Annual cycle of parallel monthly data sets between 2001 and 2010
Obs8 – the average of surface observed data for 8 stations; SIS – the average of the monthly satellite data for the 8 pixels covering the selected stations; CC – the average of the monthly modelled radiation data from CarpatClim project for grid points of the measurements.

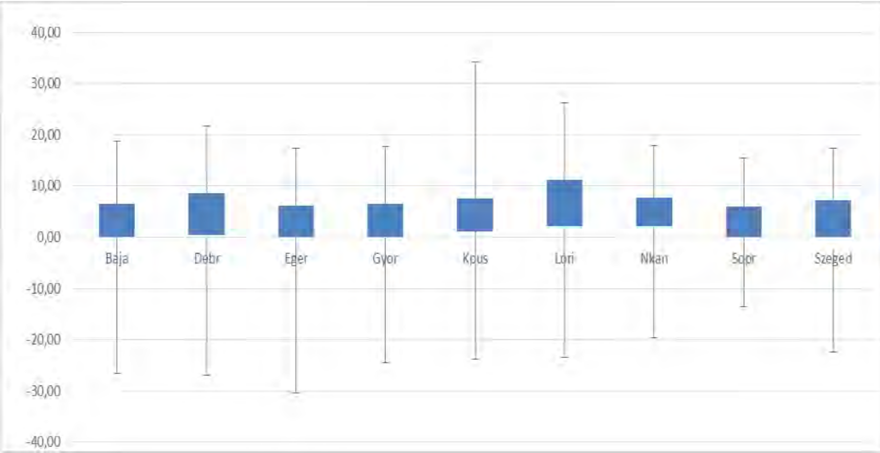


Fig. 5. Difference of satellite and observed data for each station

Numerous studies reported that Meteosat retrievals performed satisfactorily in representing global radiation. *Möser and Raschke* (1984) validated solar radiation over Europe. They used MFG monthly data from Jun. 1979 until Apr. 1982 and found RMSE 5-6%. *Dedieuet et. al* (1987) applied 8*8 km resolution MFG data for verification of western Germany, and get RMSE 6.7%. *Posset et. al* (2012) tested irradiances and cloud albedo from MVIRI climate data records between 1983 and 2005 for monthly, daily and hourly mean at spatial resolution of 0.03°. The mean absolute bias for monthly SIS calculated by Heliosat method were below 10W/m^2. Authors made intercomparison among HelioClim, ERA-Interim, GEWEX SRB, ISCCP FD and SIS CDR. SIS versus Baseline Surface Radiation Network (BSRN) measurements gave the best estimation and smallest spread than the three evaluation datasets. In the paper *of Sanches-Lorenzo et. al* (2013) the same MFG dataset was examined over Europe. The Global Energy Balance Archive (GEBA) was used as reference data. The results show a good agreement between SIS CDR and GEBA with correlation coefficient 0.86 and with a slight overestimation (mean bias difference +5.2 W/m^2). The mean absolute bias difference was 4.4 W/m^{2}and the root mean square difference 11.0 W/m^2 (9.5%), which value below the accuracy threshold of 15 W/m^2 for monthly values defined by EUMETSAT.

Table 2. Summarise verification results for Hungary. The second line contains the validation outputs of *Müller et. al* (2015). They used also SARAH data (the same version) and BSRN network as a reference. Statistics are very similar except BIAS, due to the seasonal overestimation.

Table 2. Results of comparison between the monthly surface observations derived from OMSZ network and CM-SAF surface radiation data set, and Carpatclim radiation time series respectively

Data 1	**Data 2**	**N**	**BIAS**	**MAB**	**SD**	**AC**	**Frac >15 W/m^2**
Müllet et all., 2015	BSRN	1672	1.27	5.46	7.34	0.92	5.60
SIS	OBS	1038	3.17	5.96	7.11	0.90	5.60
SIS	CarpatClim	959	0.92	6.54	8.37	0.97	7.51
CarpatClim	OBS	945	1.69	5.37	6.90	0.99	5.00

The smallest deviances were given by CarpatClim radiation data set, which used observed sunshine duration (point) data as input. After thoroughly homogenization and interpolation the estimated values reflect that applied method is reasonable to evaluate the radiation data.

5. CONCLUSION

This study aimed to compare three long term monthly global radiation time series for Hungary. The eight OMSz stations were selected as a references for verification. Shortwave Incoming solar radiation from EUMETSAT CM SAF SARAH data set indicate some systematical overestimation. CarpatClim project results gave the best agreement with observed data.

Acknowledgements

The SARAH data set was provided data by EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF). The authors are grateful to Judit Kerényi, Andrea Kircsi, Zoltán Nagy and Zoltán Tóth for their support.

References

Riihelä A., Carlund, T., Trentmann , J., Richard Müller and Anders V. Lindfors, 2015: Validation of CM SAF Surface Solar Radiation Datasets over Finland and Sweden. *Remote Sens*.,7(6), 6663-6682; doi:10.3390/rs70606663

Angström, A, 1924. - Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society*, 50:121-125

CarpatClim Project Deliverable D2.10, 2012; Final version of metadata per country of all national gridded datasets created within Module 2; Annex 2 – Computation methodology for several variables (http://www.carpatclim-eu.org/pages/deliverables/)

Dedieu, G.; Deschamps, P.Y.; Kerr, Y.H. (1987). "Satellite Estimation of Solar Irradiance at the Surface of the Earth and of Surface Albedo Using a Physical Model Applied to Meteosat Data." *Journal of Climate and Applied Meteorology* (26); pp. 79–87.

Möser, W.; Raschke, E. (1984). "Incident Solar Radiation over Europe Estimated From METEOSAT Data." Journal of Climate and Applied Meteorology (23); pp. 166–170.

Müller, Richard; Pfeifroth, Uwe; Träger-Chatterjee, Christine; Cremer, Roswitha; Trentmann, Jörg; Hollmann, Rainer. (2015): Surface Solar Radiation Data Set - Heliosat (SARAH) - Edition 1. Satellite Application Facility on Climate Monitoring. DOI:10.5676/EUM_SAF_CM/SARAH/V001. http://dx.doi.org/10.5676/EUM_SAF_CM/SARAH/V001

Müller, R., Pfeifroth, U, Träger-Chatterjee, Ch., Trentmann, J. and Cremer, R., 2015: Digging the METEOSAT Treasure—3. Decades of Solar Surface Radiation, *Remote Sensing*, 7(6), 8067-8101, http://dx.doi.org/10.3390/rs70608067.

Prescott, J.A., 1940. - Evaporation from a water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia*, 64:114-118.

Rimóczi-Paál A., Mika J., Szentimrey T., Csiszár I., Gyarmati G., Domonkos P. and Károssy C., 1997: Estimation of surface radiation balance components from METEOSAT images: Five years statistics. *Advances in Space Research* Vol 19, No. 3, 473-476 0,468

Rusznyák R. 2014: Validation of UMETSAT CM-SAF products with surface global radiation measurements (MsC dissertation in Hungarian)

Posselt, R., R. W. Müller, R. Stöckli, and J. Trentmann, 2012: Remote sensing of solar surface radiation for climate monitoring — the CM-SAF retrieval in international comparison. *Remote Sensing of Environment*, 118, 186-198

Sanchez-Lorenzo, A. and Wild, M., and Trentmann, J., 2013: Validation and stability assessment of the monthly mean CM SAF surface solar radiation dataset over Europe against a homogenized surface dataset (1983-2005), *Remote Sensing of Environment*, 134, 355–366.

M. Shengupta, A. Habte, S. Kurtz, A. Dobos, S. Wilbeer, E. Lorenz, T. Stoffel, D. René, D. Myers, S. Wilcox, P. Blanc, R. Perez: Best practices handbook for the collection and use of solar resource data for solar energy applications. Technical Report NREL/TP-5D00-63112, *National Renewable Energy Laboratory*, Golden, Colorado, USA, 2015

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

Szentimrey, T., Bihari, Z., 2006: MISH (Meteorological Interpolation based on Surface Homogenized Data Basis), COST Action 719 Final Report, The use of GIS in climatology and meteorology, Edited by Ole Einar Tveito, Martin Wegehenkel, Frans van der Wel and Hartwig Dobesch, 2006, pp. 54-56

DAILY SERIES HOMOGENIZATION AND GRIDDING WITH CLIMATOL V. 3

José A. Guijarro

State Meteorological Agency (AEMET), Balearic Islands Office, Spain
<jguijarrop@aemet.es>

1. INTRODUCTION

The need to homogenize observational series before its use to assess climate variability emerged long time ago. Efforts were initially focused on annual, seasonal and monthly series, and the successful COST Action ES0601 allowed the exchange of ideas between homogenization specialists and the improvement of the their methodologies. But now the stress is put on the homogenization of daily series, since the study of the variability of indices and extreme values depends on them. The new version 3.0 of the R *Climatol* package provides functions to facilitate the homogenization of climatological variables at any temporal scale, as long as the data in the series may be considered synchronous (which is doubtful in the case of sub-daily data).

2. CLIMATOL HOMOGENIZATION METHODOLOGY

The homogenization procedure relies on a simple method for estimating data at one point by means of other synchronous data from nearby stations, using a form of orthogonal regression known as Reduced Major Axis (RMA; *Leduc*, 1987). Orthogonal regression is adjusted by minimizing the perpendicular distance of the scatter points to the regression line, instead of minimizing the vertical distance to that line as in Ordinary Least Squares regression (OLS). Although there is an analytic expression to adjust an Orthogonal regression, a simpler and very close approximation is the RMA expression:

$$\hat{y}_i = x_i$$

in which both dependent and independent variables have been previously standardized by removing their mean and dividing by their standard deviation. The Climatol package allows the use of one or more reference data, and hence x_i may be an (optionally weighted) mean of several nearby data. ("nearby data" is used instead of "nearby series" because data availability in the surroundings vary along time when series are incomplete).

This procedure allows great flexibility to use nearby data to fill in missing data in a problem series, since no common period of observation is needed between the two, and the closest reference data can be chosen in every time step adapting to the different availability of data in the other series. Its main drawback is the estimation of the right means (and standard deviations in the default normalization) of the series when they have missing data, which is normally the case. This problem is solved here by computing initial values with the available data in every series, estimating the missing data, recomputing their means (and standard deviations), and repeating the process until the maximum difference between the last and the previous means lies below a prescribed threshold.

Apart from the full standardization, Climatol allows the “normalization” of the data by means of either only subtracting the mean or dividing by the mean. This latter is more appropriate for variables with a zero lower limit and a skewed frequency distribution, as is the case with precipitation or wind, but no series should have mean values lower than 1. (Units may be scaled to avoid this possibility).

Note that the OLS regression with standardized variables can be expressed as $\hat{y}_i = r\, x_i$, where r is the correlation coefficient. Therefore, the lower the correlations between the series, the more will be reduced the variance of the estimated data, while RMA regression is free from this effect, which is important for the analysis of extreme values in the homogenized series.

This method is applied to estimate all series in the studied database, and the estimated data can be used to fill in any missing data and to obtain series of anomalies (observed - estimated data) on which to detect outliers or shifts in the mean through the Standard Normal Homogeneity Test (SNHT; *Alexandersson*, 1986). When the series have missing data, their mean and standard deviation is computed with their available data first, and recomputed after their missing data are filled in, repeating the procedure until a preset degree of convergence is reached. On the other hand, when the SNHT statistic of the series are greater than a prescribed threshold, the series is split at the point of maximum SNHT giving birth to a new series that is incorporated into the data pool. This procedure is done iteratively, splitting only the series with the higher SNHT values at every iteration, until no series is found inhomogeneous. (The SNHT threshold is decided subjectively, since the optimum values depend on the time scale and the variable analyzed. Histograms of SNHT values are provided in an output PDF file to help choosing the thresholds).

3. HOMOGENIZATION OF DAILY DATA

The main problem in the homogenization of daily data is due to the high variability of these series, that lowers the power of detection of shifts in their mean along time. That is why the detection of the inhomogeneities is preferably done on the monthly aggregates of the series, with less inherent variability. In fact the recommended procedure for homogenizing daily temperatures has been to homogenized at the monthly scale and to adjust the daily series with interpolated monthly corrections (*Vincent et al.,* 2002; *Brunet et al.,* 2006), but it did not yield good results when applied to daily peak wind gusts series from Portugal and Spain (*Azorin-Molina et al.,* 2016) and Australia (paper in preparation), nor in experiences with daily precipitation (paper in preparation), that can be attributed to their strongly skewed frequency probability distributions.

But the Climatol package does not need to interpolate monthly corrections, since it just splits the daily series at the break-points detected in the monthly homogenization, and then reconstruct all the series from their homogeneous sub-periods in a final stage by estimating all their missing data. Therefore the final procedure recommended to homogenize daily series with the Climatol 3.0 package consists in the following steps:

1. Prepare the input files in the format for Climatol.
2. First exploratory analysis of the daily data for quality control.
3. Aggregate the daily data into monthly series.
4. Homogenize the monthly series.
5. Adjust the daily series using the monthly detected break-points.

This methodology will be illustrated in the next sub-sections with the example data accompanying the package.

3.1. Preparation of the input files for Climatol

Input data must be prepared as indicated in the package manual. If they are stored in a data-base accessible through the ODBC protocol, the provided function db2dat() can be used to generate the input files. Only for the purpose of running the following examples, these files can be generated in the working directory by means of these commands (anything after # is a comment):

```
library(climatol) # load the functions of the package
data(Ttest) #load the example data into R memory space
write(dat, 'Ttest_1981-2000.dat') #save data file
write.table(est.c, 'Ttest_1981-2000.est', row.names=FALSE,
     col.names=FALSE) #save stations file
rm(dat, est.c) #remove loaded data from memory space
```

3.2. First exploratory analysis of the daily data for quality control

```
homogen('Ttest', 1981, 2000, expl=TRUE)
```

The user should inspect the output graphics in Ttest_1981-2000.pdf to verify that data does not show weird features. The histogram of anomalies (near the end of the file) may help to choose appropriate thresholds for outlier rejection in the final homogenization step.

3.3. Aggregate the daily data into monthly series

```
dd2m('Ttest', 1981, 2000) #daily data to monthly values
```

Monthly input files are saved as Ttest-m_1981-2000.dat and Ttest-m_1981-2000.est

3.4. Homogenize the monthly series

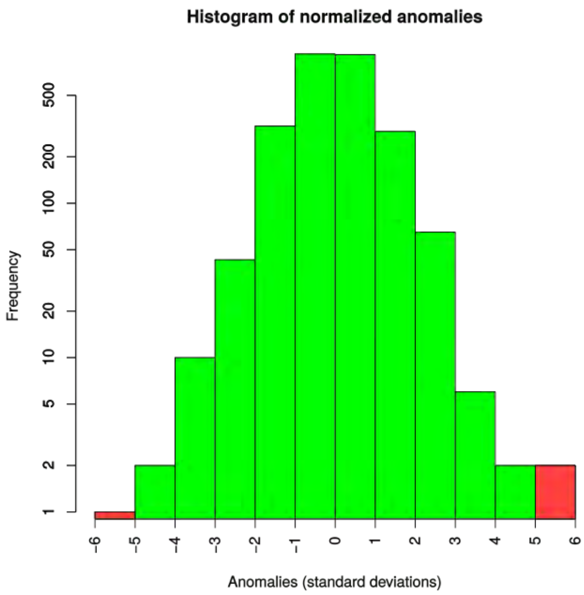
```
homogen('Ttest-m', 1981, 2000) #homogenization of monthly data
```

The user should look at the output graphics in Ttest-m_1981-2000.pdf to check if the default threshold values for outlier rejection (dz.max, *Figure 1 left*) and SNHT (snht1 and snht2, the latter shown in *Figure 1 right*) have been suitable, or re-run the above command specifying better thresholds. If metadata are available, the file Ttest-m_1981-2000_brk.csv containing the break-point list can be edited to refine the dates of the shifts in the mean. (Note that not all changes in the history of the station must necessarily produce inhomogeneities, while changes that produce them are often not reported in the metadata).

3.5. Adjust the daily series using the monthly detected break-points

```
homogen('Ttest', 1981, 2000, metad=TRUE)
```

After every run of the homogen() function, several output files will be generated: a PDF file with lots of diagnostic graphics, lists of breaks and outliers in CSV format, and an R binary file containing raw and homogenized series.



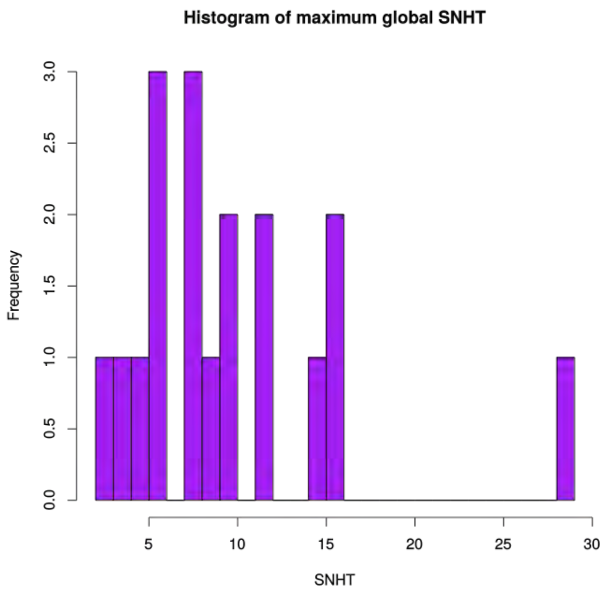


Fig. 1. Histograms of normalized anomalies with rejected outliers (greater than the set dz.max parameter) in red (left) and of maximum SNHT (right).

4. OBTAINING PRODUCTS FROM HOMOGENIZED DATA

The user can load the results of the homogenization as various data objects (explained in the documentation of the package) into the R memory space for any further manual processing by issuing the command:

```
load('Ttest_1981-2000.rda')
```

But a couple of post-processing functions are provided in the package to help in obtaining common products from the homogenized series, either directly from the daily series, or from their monthly aggregates, which can be generated by:

```
dd2m('Ttest', 1981, 2000, homog=TRUE)
```

Examples for getting some statistical products:

```
dahstat('Ttest', 1981, 2000) #means of the daily series
dahstat('Ttest', 1981, 2000, mh=TRUE) #monthly means
dahstat('Ttest', 1981, 2000, mh=TRUE, stat='tnd') #monthly OLS
                                               trends and p-values
dahstat('Ttest', 1981, 2000, stat='q', prob=.2) #first
quintile
                                               of daily values
```

Another function is provided to obtain homogenized grids, but you must define your grid limits and resolution first, as in:

```
grd=expand.grid(x=seq(-109,-107.7,.02), y=seq(44,45,.02))
#grid
library(sp) #load needed package for the following command:
coordinates(grd) <- ~x+y #convert grid into a spatial object
```

Now grids can be generated (in NetCDF format) with:

```
dahgrid('Ttest', 1981, 2000, grid=grd) #daily grids
dahgrid('Ttest', 1981, 2000, grid=grd, mh=TRUE) #monthly grids
```

These grids are built in normalized, dimensionless values. You can obtain a new file with temperatures in degrees Celsius with externals tools, such as the CDO (note that this command must be run on a terminal, outside R):

```
cdo add -mul Ttest-mh_1981-2000.nc Ttest-mh_1981-2000_s.nc \
     Ttest-mh_1981-2000_m.nc Ttest-mu_1981-2000.nc
```

But the new grids in Ttest-mu_1981-2000.nc will be based on geometric interpolations only, and therefore it is better to build new files Ttest-mh_1981-2000_m.nc and Ttest-mh_1981-2000_s.nc through geostatistical methods before using them to undo the normalization of the grids.

5. CONCLUSION

The Climatol package, that can be freely installed through the means provided in any running R environment, is a convenient tool to homogenize monthly and daily series without much effort, and is adapted to use series with a very high amount of missing data. Although most parameters of their functions are set with default values, the user should tune them to the variable under study and its time resolution to optimize the results. (https://cran.r-project.org/web/packages/climatol/climatol.pdf holds the documentation in PDF format).

References

Azorín-Molina C, Guijarro JA, McVicar TR, Vicente-Serrano SM, Chen D, Jerez S, Espirito-Santo F (2016): Trends of daily peak wind gusts in Spain and Portugal, 1961-2014. *Journal of Geophysical Research Atmospheres*, DOI: 10.1002/2015JD024485.

Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, Moberg A, Lister D, Walther A, Lopez D, Almarza C (2006): The development of a new dataset of Spanish daily adjusted temperature series (SDATS) (1850-2003). *Int. J. Climatol.*, 26:1777-1802.

Vincent LA, Zhang X, Bonsal BR, Hogg WD (2002): Homogenization of daily temperatures over Canada. *Journal of Climate*, 15:1322-1334.

ANNEX: MORE EXAMPLES OF THE GRAPHIC OUTPUT

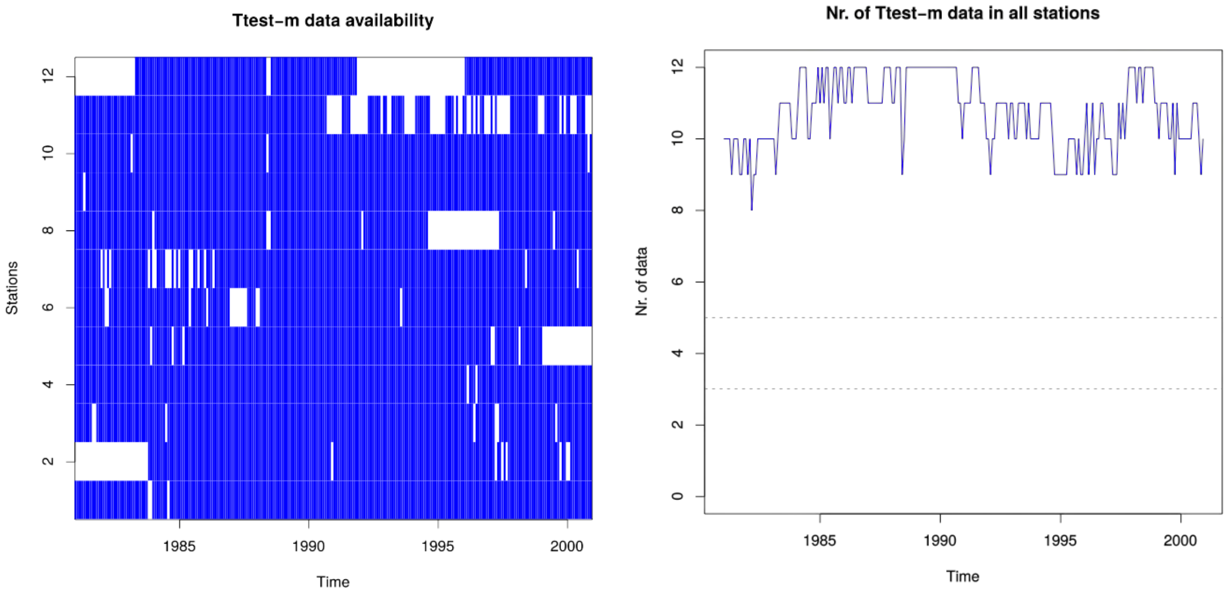


Fig. 2. Detailed (left) and global (right) data availability. Dashed green and red horizontal lines mark the desired (5) and critical (3) minimum number of data available at every time step.

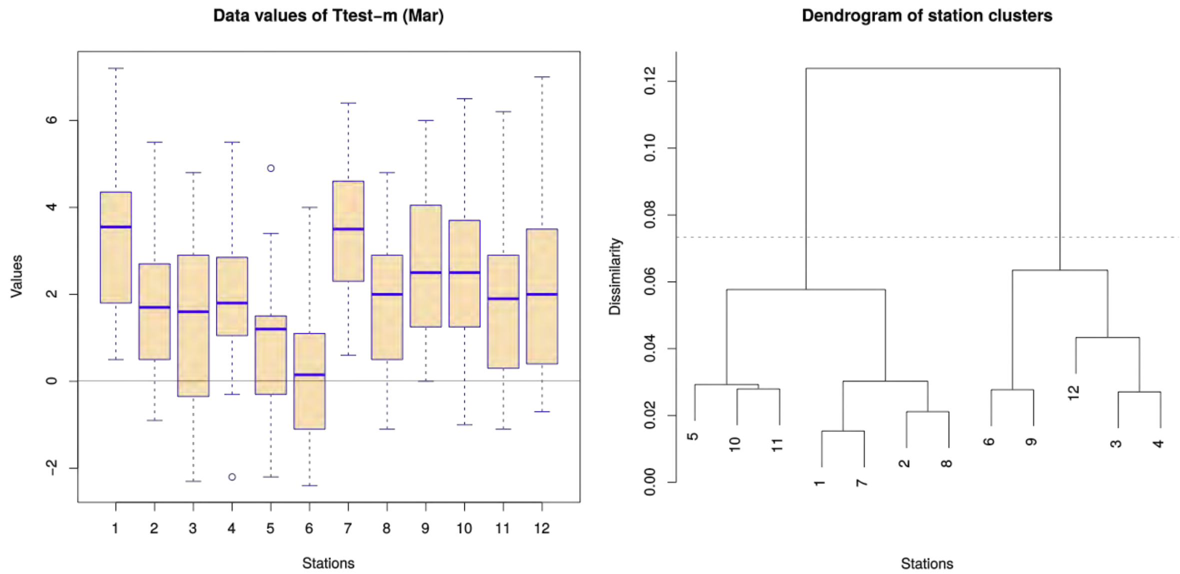


Fig. 3. Example of boxplots of the data values (left) that help to visually inspect biases and outliers in the data, and a dendrogram of station clusters that, with the help of a provided map of the stations (not shown) can be useful to detect climatic discontinuities that might suggest a separate homogenization for the areas with different climate regimes.

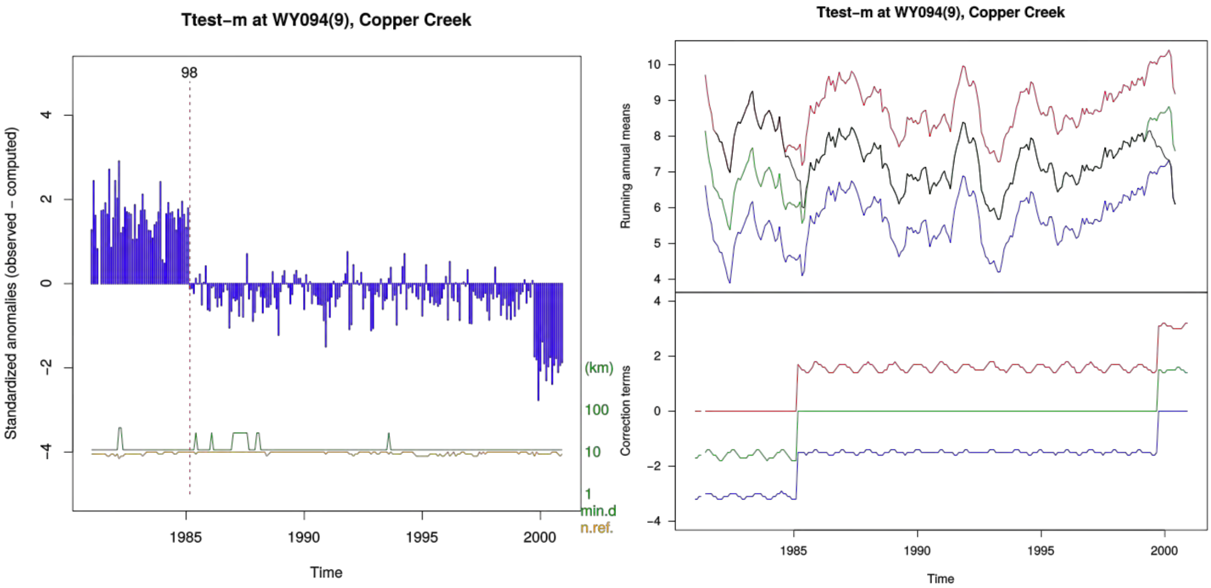


Fig. 4. The left side shows an example of detection of a first break-point on the series of anomalies (observed minus estimated data) with a SNHT value of 98. The green line shows the distance to the closest reference data at each time step, and the orange line displays the number of references. A second break-point will be detected in another iteration, and therefore three homogeneous sub-periods will be fully reconstructed by the missing data filling algorithm at the end of the process. These reconstructions and the applied corrections can be seen in the right side of the figure, where original data are plotted in black, and the series reconstructions appear in different colors.

PROGRAMME

Budapest, Hungary
03-07 April 2017

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

Homogenization sessions: 03 April Monday-05 April Wednesday
Interpolation session: 05 April Wednesday-06 April Thursday
Software session: 07 April Friday

MONDAY, 03 APRIL

8:30 – 9:00 Registration

9:00 – 12:00

Opening addresses by

President of HMS

Delegate of WMO

Organizers

Introductory Presentations

Hechler, P., Baddour, O.: Underpinning Climate Services

Coffee break

Szentimrey, T.: Introduction on homogenization, quality control, spatial interpolation, gridding

Venema, V., Lindau, R.: Global temperature trend biases and statistical homogenization methods

Lunch break

14:00 – 17:00 Homogenization and quality control

Aguilar, E., Asín, J., Azorín, C., Gilabert, A., Guijarro, J. A., Lopez, J. A., Prohom, M., Rasilla, M., Solé, G.: Evaluation of the Impact over temperature series of the transitions between observation systems (IMPACTRON)

Guijarro, J.A., Cuxart, J., Simó, G., Martínez-Villagrasa, D., López, A., Picos, R., Martí, B., Jiménez, M.A., Gilabert A., Aguilar, E.: Small scale surface heterogeneities and impact on station relocations

Hannak, L.: Analysis of the impacts of the automatization of measurement systems using parallel measurements from German Climate Reference Stations

Coffee break

Domonkos, P.: Time series homogenisation with optimal segmentation and ANOVA correction: past, present and future

Guijarro, J. A., Lopez, J. A., Aguilar, E., Domonkos, P., Venema, V.K.C., Sigró, J., Brunet, M.: Comparison of homogenization packages applied to monthly series of temperature and precipitation: the MULTITEST project

18:00 – Welcome party (Hungarian Meteorological Service, 1 Kitaibel Pál street, Budapest)

TUESDAY, 04 APRIL

9:00 – 12:00 Homogenization and quality control

9:00 – 10:30 New WMO Guidance on homogenisation: Introductory presentation by Victor Venema, followed by an open discussion seeking feedback from seminar participants

Coffee break

Szentimrey, T.: Some theoretical questions and development of MASH for homogenization of standard deviation

Lundstad, E., Stepanek, P., Zahradníček, P.: Long-term homogenised precipitation data sets for Norway

Motrøen Gjelten, H., Lundstad, E., Tveito, O. E.: Homogeneity testing of seasonal precipitation series in Norway

Lunch break

14:00 – 17:00 Homogenization and quality control

Chimani, B., Ganekind, M.: Differences in Climate Evolution Analyses depending on the choice o by homogenization method and time span

Shelton, K., Warren, S., Davis, R., Faulkner, D.: Homogenisation of daily station data in England and Wales

Squintu, A.: Homogenization of ECA&D temperature series

Coffee break

Vint, K., Keevallik, S., Meitern, H.: Inhomogenieties in Estonian air temperature series with CLIMATOL and HOMER

Yosef, Y., Aguilar, E., Alpert, P.: Comparison of different daily adjustment methods for the maximum and minimum temperature in Israel

Mateus, C., Potito, A., Curley, M.: Digitisation and homogenisation of the long term daily (max/min) summer and winter air temperature records in Ireland

WEDNESDAY, 05 APRIL

9:00 – 12:00 Homogenization and quality control

Rustemeier, E., Kapala, A., Meyer-Christoffer, A., Finger, P., Schneider, U., Venema, V., Ziese, M., Simmer, C., Becker, A.: HOMPRA Europe - A gridded precipitation data set from European homogenized time series

Zahradníček, P., Petr, S.: Homogenization of the wind speed time series in Czech Republic

Rasol, D.: Modernisation of Croatian Meteorological Measurements Network

Coffee break

Bertrand, C., Journée, M.: Data QC within the Belgian synoptic and climatological networks: an overview

Michel A, P.: A machine learning perspective towards fully automated quality control in daily weather time series

Delvaux, C.: Quality Control and Homogenization of the Belgian Historical Weather Data

Lunch break

14:00 – 15:20 Homogenization and quality control

Graw, K., Schröder, M., Andersson, A., Kinzel, J., Fennig, K., Hollmann, R.: Stability of satellite based climate data records (CDRs) retrieved by CM SAF

Van Malderen, R., Pottiaux, E., Klos, A., Bock, O., Bogusz, J., Chimani, B., Elias, M., Gruszczynska, M., Guijarro, J. A., Zengin Kazancı, S., Ning, T.: The homogenization of GPS Integrated Water Vapour time series: methodology and benchmarking the algorithms on synthetic datasets

Schröder, M., Lockhoff, M., Shi, L., Graw, K.: The GEWEX water vapor assessment (G-VAP) – results from inter-comparisons and stability analysis

Coffee break

15:40-17:00 Spatial interpolation, gridded datasets

Szentimrey, T.: New developments of interpolation method MISH: modelling of interpolation error RMSE, automated real time Quality Control

Peña-Angulo, D., Celia, S. S., Rodrigues, M., Azucena, J. C., Brunetti, M., González-Hidalgo, J. C.: Comparison of different interpolation methods for the generation of a climatology with maximum and minimum monthly temperatures

19:00 – Seminar banquet

THURSDAY, 06 APRIL

9:00 – 12:00 Spatial Interpolation, gridded datasets

Isotta, F., Begert, M., Frei, C.: Temperature and precipitation grid datasets for climate monitoring based on homogeneous time series in Switzerland

Frei, C., Isotta, F.: Uncertainty in the interpolation of daily precipitation – Insights from an ensemble analysis for the Alps

Lussana, C., Tveito, O. E.: Nordic gridded climate data set, status and plans

Coffee break

Tveito, O. E, Lussana, C.: Ensemble approaches to assess uncertainties in observation gridding

Rehfeldt, K., Kolbe, C., Ziese, M., Rustemeier, E., Krähenmann, S., Becker, A.: Comparison of Three Interpolation Schemes for Six Parameters

Kveton, V: Experiences with snow level estimation for spatial analyse of new snow depth based on precipitation data

Lunch break

14:00 – 17:00 Spatial interpolation, gridded datasets

Mamara, A., Anadranistakis, M., Argiriou. A. A.: Homogenization and gridding of the Greek time series

Petrović, P.: Comparison of Gridded and Observed Temperature and Precipitation Episode Series: A Case Study

Bihari, Z., Szentimrey, T., Kircsi, A.: Some details about the theoretical background of CarpatClim – DanubeClim gridded databases and their practical consequences

Coffee break

Lakatos, M., Szentimrey, T., Izsák, B., Hoffmann, L.: Comparison of E-OBS and CARPATCLIM gridded datasets of minimum temperatures, maximum temperatures and precipitation by Analysis of Variance (ANOVA)

Dobi, I.: Comparison of monthly satellite, modelled and in situ surface radiation data over Hungary

Höpp, S., Rauthe, M., Deutschländer, T., Krähenmann, S., Hänsel S.: Developing a gridded data set of global radiation covering Germany and its neighbouring river catchment areas

FRIDAY, 07 APRIL

9:00 – 12:00 Software Presentations

Domonkos, P.: Software ACMANT3

Szentimrey, T.: Software MASH (Multiple Analysis of Series for Homogenization)

10:30 – 11:00 coffee break

Stepanek, P.: Software AnClim for tutorial of statistical methods in climatology (including homogenization) and ProClimDB for processing of climatological datasets

Szentimrey, T.: Software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis)

This session is still open please to inform us in case of intention to present any homogenization or QC or interpolation software!

LIST OF PARTICIPANTS 2017

AUSTRIA

BARBARA CHIMANI
Central Institute for Meteorology and Geodynamics
barbara.chimani@zamg.ac.at

BELGIUM

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

CHARLES DELVAUX
Royal Meteorological Institute of Belgium
charles.delvaux@meteo.be

PEREZ MICHEL A.
Royal Meteorological Institute of Belgium& ETRO-VUB
maperezg@etrovub.be

ROELAND VAN MALDEREN
Royal Meteorological Institute of Belgium
roeland.vanmalderen@meteo.be

ROMAIN INGELS
Royal Meteorological Institute of Belgium
romain.ingels@meteo.be

CROATIA

ANA WEISSENBERGER
Meteorological and Hydrological Service, Croatia
ana.weissenberger@cirus.dhz.hr

DUBRAVKA RASOL
Meteorological and Hydrological Service, Croatia
rasol@cirus.dhz.hr

KUZMIC JOSIPA
Meteorological and Hydrological Service, Croatia
kuzmic@cirus.dhz.hr

SLAVICEK LASTA
Meteorological and Hydrological Service, Croatia
lasta.slavicek@cirus.dhz.hr

CZECH REPUBLIC

PETR STEPANEK
Global Change Research Centre AS CR, v. v. i.
stepanek.p@czechglobe.cz

PAVEL ZAHRADNÍCEK
Global Change Research Institute, AS CR, v.v.i, Brno, Czech Republic/ Czech Hydrometeorological Institute, Brno, Czech Republic
zahradnicek.p@czechglobe.cz

VÍT KVĚTOŇ
Czech Hydrometeorological Institute
vit.kveton@chmi.cz

ESTONIA

KAIRI VINT
Estonian Environment Agency
kairi.vint@envir.ee

FRANCE

OLIVIER BOCK
IGN (French National Geographic Institute)
Olivier.Bock@ign.fr

GERMANY

ELKE RUSTEMEIER
Global Precipitation Climatology Centre / Deutscher Wetterdienst
elke.rustemeier@dwd.de

KARSTEN FRIEDRICH
Deutscher Wetterdienst
karsten.friedrich@dwd.de

KATHRIN GRAW
Deutscher Wetterdienst
Kathrin.Graw@dwd.de

KIRA REHFELDT
Deutscher Wetterdienst
Kira.Rehfeldt@dwd.de

LISA HANNAK
Deutscher Wetterdienst
lisa.hannak@dwd.de

SIMONA HÖPP
Deutscher Wetterdienst
simona-andrea.hoepp@dwd.de

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

GREECE

ATHANASSIOS ARGIRIOU
Laboratory of Atmospheric Physics, University of Patras
athanarg@upatras.gr

HUNGARY

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

ZITA BIHARI
Hungarian Meteorological Service
bihari.z@met.hu

MÓNIKA LAKATOS
Hungarian Meteorological Service
lakatos.m@met.hu

ILDIKÓ DOBI
Hungarian Meteorological Service
dobi.i@met.hu

SÁNDOR SZALAI
Szent István University
Szalai.Sandor@mkk.szie.hu

LILLA HOFFMANN
Hungarian Meteorological Service
hoffmann.l@met.hu

ANDREA KIRCSI
Hungarian Meteorological Service
kircsi.a@met.hu

BEATRIX IZSÁK
Hungarian Meteorological Service
izsak.b@met.hu

IRELAND

CARLA PATRÍCIA PEDROSO MATEUS
National University of Ireland Galway
C.PEDROSOMATEUS2@nuigalway.ie

MARY CURLEY
Met Eireann
mary.curley@met.ie

ISRAEL

YIZHAK YOSEF
Israel Meteorological Service and Tel-Aviv University
yosefy@ims.gov.il

ITALY

GUIDO FIORAVANTI
ISPRA
guido.fioravanti@isprambiente.it

LATVIA

SVETLANA ANISKEVICH
Latvian Environment, Geology and Meteorology Centre
svetlana.aniskevica@lvgmc.lv

MONTENEGRO

VANJA RAJOVIC
Institute of Hydrometeorology and Seismology of Montenegro
vanja.rajovic@meteo.co.me

NETHERLAND

ANTONELLO SQUINTU
KNMI - The Royal Netherlands Meteorological Institute
antonello.squintu@gmail.com

NORWAY

ELIN LUNDSTAD
Norwegian Meteorological Institute
elinl@met.no

HERDIS M. GJELTEN
Norwegian Meteorological Institute
herdismg@met.no

OLE EINAR TVEITO
Norwegian Meteorological Institute
ole.einar.tveito@met.no

POLAND

AGNIESZKA WYPYCH
Jagiellonian University & Institute of Meteorology and Water Management – National Research Institute
agnieszka.wypych@uj.edu.pl

ZBIGNIEW USTRNUL
Institute of Meteorology and Water Management – National Research Institute
zbigniew.ustrnul@uj.edu.pl

SERBIA

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

SLOVAKIA

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

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

SPAIN

CELIA SALINAS-SOLE
Department of Geography, University of Zaragoza
cs@unizar.es

ENRIC AGUILAR
Center for Climate Change, C3, Universitat Rovira i Virgili, Tarragona
enric.aguilar@urv.cat

JIMÉNEZ CASTANEDA AZUCENA
University of Zaragoza
geoazu.flysch@gmail.com

JOSÉ A. GUIJARRO
AEMET (Spanish State Meteorological Agency)
jguijarrop@aemet.es

PÉTER DOMONKOS
dpeterfree@gmail.com

SWITZERLAND

CHRISTOPH FREI
Federal Office of Meteorology and Climatology MeteoSwiss
christoph.frei@meteoswiss.ch

FRANCESCO ISOTTA
Federal Office of Meteorology and Climatology MeteoSwiss
francesco.isotta@meteoswiss.ch

RETO STÖCKLI
MeteoSwiss
reto.stoeckli@meteoswiss.ch

UNITED KINGDOM

KAY SHELTON
JBA Consulting
kay.shelton@jbaconsulting.com

WMO

PEER HECHLER
phechler@wmo.int

For more information, please contact:

World Meteorological Organization

Observing and Information Systems Department

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

Tel.: +41 (0) 22 730 82 68 – Fax: +41 (0) 22 730 80 21

Email: wcdmp@wmo.int

public.wmo.int

JN 171647