DOCTORAL PROGRAMME

Information Management

Specialization in Geographic Information Systems

A Geostatistical Simulation Algorithm for the Homogenisation of Climatic Time Series: a Contribution to the Homogenisation of Monthly Precipitation Series

Sara Cristina da Silva Ribeiro

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor in Information Management

November 2016

NOVA Information Management School

Universidade Nova de Lisboa

NOVA IMS Information Management School

Doctoral Programme in **Information Management**

Professora Doutora Ana Cristina Costa, Supervisor

ABSTRACT

As defined by the Intergovernmental Panel on Climate Change (IPCC), climate change refers to a change in the state of the climate that can be identified by changes in the statistical characteristics of its properties and that persists for an extended period, typically decades or longer. In order to assess climate change and to develop impact studies, it is imperative that climate signals are clean from any external factors. However, non-natural irregularities are an inevitable part of long-time climate records. They are introduced during the process of measuring and collecting data from weather stations. Accordingly, it is essential to detect and correct those irregularities a priori, through a process called homogenisation. This process became a hot topic in the last decades and many researchers have focused on developing efficient methods. Still, some climatic variables are lacking homogenisation procedures due to their high variability and temporal resolution (e.g., monthly precipitation).

We propose the gsimcli (Geostatistical SIMulation for the homogenisation of CLImate data) homogenisation method, which is based on a geostatistical simulation method, namely the direct sequential simulation. The proposed approach considers simulated values of the candidate station's neighbouring area, defined by the local radius parameter, aiming to account for local characteristics of its climatic zone. gsimcli has other modelling parameters, such as the candidates order in the homogenisation process, the detection parameter, and the correction parameter (also used to fill in missing data). A semi-automatic version of gsimcli is also proposed, where the homogenisation adjustments can be estimated from a comparison series. The efficiency of the gsimcli method is evaluated in the homogenisation of precipitation data. Several homogenisation exercises are presented in a sensitivity analysis of the parameters for two different data sets: real and artificial precipitation data. The assessment of the detection part of gsimcli is based on the comparison with other detection techniques using real data, and extends a previous study for the south of Portugal. Artificial monthly and annual data from a benchmark data set of the HOME project (ACTION COST-ES0601) is used to assess the performance of gsimcli. These results allow the comparison between gsimcli and state-of-the-art methods through the calculation of performance metrics.

This research allowed identifying gsimcli parameters that have a high influence in the homogenisation results: correction parameter, grid cell size and local radius parameter. The set of parameters providing the best values of performance metrics are recommended as the most suitable set of homogenisation parameters for monthly precipitation data. Results show gsimcli as a favourable homogenisation method for monthly precipitation data that outperformed a few well established procedures. The filling in of missing data is an advantage when compared to other methods. Taking advantage of its capability of filtering irregularities and providing

comparison series, gsimcli can also be used as a pre-homogenisation tool followed by the use of a traditional homogenisation method (semi-automatic approach).

As future work, it is recommended the performance assessment of the gsimcli method with denser monitoring networks, and the inclusion of a multivariate geostatistical simulation algorithm in the homogenisation procedure.

RESUMO

As alterações climáticas, tal como definidas pelo Painel Intergovernamental para as Alterações Climáticas das Nações Unidas, referem-se a uma modificação no estado do clima que pode ser identificada através de alterações nas suas propriedades estatísticas e que perdura por um largo período de tempo, tipicamente décadas ou períodos mais longos. Para a avaliação das alterações climáticas, e para o desenvolvimento de estudos de impacte, é imperativo que os sinais climáticos estejam isentos de quaisquer fatores externos. Inevitavelmente, as séries temporais de dados climáticos contêm irregularidades não-naturais. Tais irregularidades são introduzidas durante o processo de medição e recolha de dados nas estações meteorológicas. Assim, é essencial a prévia deteção e correção dessas irregularidades, através de um processo chamado homogeneização. Nas últimas décadas, este processo tornou-se um tópico relevante e muitos investigadores procuraram desenvolver métodos de homogeneização eficientes. Contudo, existe um número reduzido de métodos para algumas variáveis climáticas devido à sua elevada variabilidade e resolução temporal (e.g., precipitação mensal).

Neste trabalho propomos o método de homogeneização gsimcli (Geostatistical SIMulation for the homogenisation of CLImate data), o qual se baseia num método de simulação geoestatística, a simulação sequencial direta. A abordagem proposta tem em consideração valores simulados na vizinhança da estação candidata, definida pelo parâmetro raio local, com o objetivo de incorporar características locais da sua zona climática. O gsimcli tem outros parâmetros de modelação, tais como a ordem das estações candidatas no processo de homogeneização, o parâmetro de deteção e o parâmetro de correção (também usado na substituição de observações omissas). Propõe-se também uma abordagem semi-automática do gsimcli onde os ajustamentos para a correção de irregularidades podem ser estimados a partir de uma série de comparação. A eficiência do método gsimcli é avaliada na homogeneização de dados de precipitação. São apresentados vários exercícios de homogeneização numa análise de sensibilidade dos parâmetros para dois conjuntos de dados: dados reais e artificiais de precipitação. A avaliação da componente de deteção do gsimcli baseia-se na comparação com outras técnicas de deteção de irregularidades utilizando dados reais, e constitui uma extensão de um estudo anterior para o sul de Portugal. O desempenho do método gsimcli é avaliado a partir de dados artificiais (mensais e anuais) de um conjunto de dados de referência (benchmark) do projeto HOME (ACTION COST-ES0601). Estes resultados permitem a comparação do gsimcli com métodos que se constituem como o estado-da-arte neste domínio, a partir do cálculo de métricas de desempenho.

Este estudo permitiu identificar os parâmetros do gsimcli que mais influenciam os resultados da homogeneização: parâmetro de correção, o tamanho da célula e o raio local. O conjunto de parâmetros com os melhores resultados das métricas de desempenho é recomendado como sendo

o mais adequado à homogeneização da precipitação mensal. Os resultados mostram que o gsimcli tem um contributo positivo na homogeneização da precipitação mensal, tendo superado o desempenho de alguns métodos de homogeneização bem estabelecidos. A sua capacidade para substituir valores omissos é uma vantagem em relação a outros métodos. Tirando partido da sua capacidade para filtrar irregularidades e para disponibilizar séries de comparação, o gsimcli também pode ser usado como uma ferramenta de pré-homogeneização, seguindo-se a aplicação de um método tradicional de homogeneização (abordagem semi-automática).

Como trabalhos futuros, recomenda-se a avaliação de desempenho do método gsimcli com redes meteorológicas mais densas, e a inclusão de um algoritmo de simulação geoestatística multivariada no procedimento de homogeneização.

AGRADECIMENTOS

Agradeço à Professora Doutora Ana Cristina Costa pelo incansável apoio, motivação, disponibilidade, ideias, correções e planeamento, que permitiram concluir este trabalho no tempo previsto. Por ter acreditado em mim, pelo rigor científico que sempre imprimiu ao trabalho produzido e do qual tanto desejo aproximar-me. Pela oportunidade de trabalho, que me enriqueceu academica, e acima de tudo, pessoalmente. O que me ensinou servirá sempre de base de tudo quanto de melhor empreender realizar.

Aos professores que me acompanharam durante a fase curricular deste projeto e que me ajudaram a melhorar o trabalho, com as suas sábias recomendações, nomeadamente Professor Doutor Marco Painho, Professor Doutor Fernando Bação e Professor Doutor Roberto Henriques.

Ao Júlio, pela ajuda enorme durante a realização deste projeto, pela paciência em me ensinar, tirar dúvidas e guiar no mundo da geoestatística. Pelas discussões que levaram sempre à melhoria dos resultados, e pela inclusão no software gsimcli de todas as funcionalidades necessárias a tornar o processo de homogeneização mais rápido e fácil.

Aos colaboradores da NOVA IMS pela gentileza e preciosa ajuda em todas as etapas deste processo.

Ao Ilyas, pela carta de recomendação escrita há três anos, que creio ter feito mudar o rumo da minha vida académica.

Ao meus amigos e colegas, pelo incentivo e apoio ao prosseguimento deste projeto. Teria sido certamente muito mais dificil este caminho sem o vosso entusiasmo e boa disposição.

À minha família, por estar sempre aonde eu preciso, mesmo quando eu não sei. Pelo apoio e força incondicionais para ultrapassar todos os desafios, e por me ter ensinado que tudo o que fazemos na nossa vida necessita de esforço e dedicação, mas mais importante, de amor.

INSTITUTIONAL ACKNOWLEDGEMENTS

The author gratefully acknowledges the financial support of "Fundação para a Ciência e Tecnologia" (FCT), Portugal, through the research project PTDC/GEO-MET/4026/2012 ("GSIMCLI - Geostatistical simulation with local distributions for the homogenization and interpolation of climate data").

FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

2PS – Two-part search

ACMANT – Adapted Caussinus-Mestre Algorithm for Homogenizing Networks of Temperature series

ACMANT2 - New unit of ACMANT for the homogenization of monthly or daily precipitation series

BAMS – Bayesian multiple change-point detection in multiple linear regression (model)

BARE – Bayesian change-point in multiple linear regression (model)

BNHT – Bayesian normal homogeneity test

CRMSE – Centred Root Mean Square Error

DN – Data nodes

DSS – Direct Sequential Simulation

ECA&D – European Climate Assessment & Dataset

ETCCDI – Expert Team on Climate Change Detection and Indices

gsimcli – Geostatistical SIMulation for the homogenisation of CLImate data

HOCLIS – Software package for homogenization of climatological time series

HOME – Advances in Homogenization Methods of Climate Series: and Integrated Approach (COST Action ES 0601)

HOMER - HOMogenization software in R

IPCC – Intergovernmental Panel for Climate Change

MASH – Multiple Analysis of Series for Homogenization

MeteoSwiss – Federal Office of Meteorology and Climatology, Switzerland

MLR – Multiple linear regression

MPR – multi-phase regression

OHOMs – Objective homogenisation methods

pdf(s) – Probability density function(s)

PMF - Penalised Maximal F-test

PMT – Penalised Maximal T-test

RMSE – Root Mean Square Error

SNHT – Standard Normality Homogeneity Test

SNIRH – Sistema Nacional de Informação de Recursos Hídricos

SUR – Seemingly Unrelated Regression equations (SUR)

THOMAS – Tool for homogenization of monthly data series

TPR - Two-phase regression

UBRIS – Urban Bias Remaining in Series

USHCN – United States Historical Climatology Network

WMO – World Meteorological Organization

ZAMG – Central Institute for Meteorology and Geodynamics, Austria

1 Introduction

1.1 Problem Statement

Few long-term climate data series are free from irregularities (Auer et al., 2005). Those irregularities comprise two categories: natural and non-natural. Natural irregularities are caused by natural phenomena, such as ashes and gases of an erupting volcano that would prevent solar radiation from reaching the earth's surface, introducing a decrease in temperature, or the effect of the North Atlantic Oscillation in extreme events across Europe (Maugeri et al., 2004). Non-natural irregularities are caused by non-natural factors, like changes in instrumentation, observing practices, relocation of the weather stations. Some changes cause sharp discontinuities (Puglisi et al., 2010) while other changes, particularly changes in the environment around the station, can cause gradual biases in the data (Peterson et al., 1998). A high number of non-natural irregularities are introduced during the process of collecting, calculating, digitizing, processing, transferring, storing and transmitting climate data series (Brunet and Jones, 2011). Also, the magnitude of inhomogeneities may differ with varying weather situations (Nemec et al., 2013).

Most long-term climatological time series have been affected by a number of non-natural factors that make these data unrepresentative of the actual climate variation occurring over time (Aguilar et al., 2003). Those non-natural irregularities, also named inhomogeneities, must be removed prior to the use of the climate data series in studies like climate change monitoring, weather forecasting or other hydrological and environmental projects (Domonkos, 2013a). Reliable results cannot be expected from those projects if the climate data series used as input contain inhomogeneities. In that sense, it is extremely important to homogenise those series, which means detecting and correcting the non-natural irregularities.

Moreover, due to the increase of storage capacity, the recent gathering of massive amounts of weather data implies also a toilsome effort to guarantee its quality. Effective and agile homogenisation procedures should be undertaken to ensure that big data, regarding weather variables, can also be used as a valuable source.

1.2 Scientific Background

A homogeneous climate time series is defined as one where variations are caused only by variations in climate (Aguilar et al., 2003). Several homogenisation methods have been proposed in the last decades (Domonkos et al., 2012; Peterson et al., 1998; Ribeiro et al., 2016a). They were developed using classical statistical tests, such as the SNHT - Standard Normality Homogeneity Test (Alexandersson, 1986; Alexandersson and Moberg, 1997), the Buishand test (1982), the Pettitt test (1979), or using regression models (e.g., Easterling and Peterson, 1995; Reeves et al.,

2007) and Bayesian approaches (e.g., Perreault et al., 2000). Most modern procedures concentrate on methods specifically designed to detect and correct multiple inhomogeneities, such as MASH – Multiple Analysis of Series for Homogenization (Szentimrey, 1999, 2006b, 2011), ACMANT - Applied Caussinus-Mestre Algorithm for Homogenizing Networks of Temperature series (Domonkos, 2011c, 2015), PRODIGE (Caussinus and Mestre, 2004) and HOMER (Mestre et al., 2013). Those homogenisation techniques typically depend on the type of climate variable (temperature, precipitation, wind speed and direction), the temporal resolution of the observations (annual, seasonal, monthly or sub-monthly), the availability of metadata (station history information) and also the weather station network density or spatial resolution (Costa and Soares, 2009a).

Homogenisation methods can be distinguished in two groups: absolute and relative methods. Absolute methods imply the application of the tests to each station data individually. In case of relative methods, the procedures use records from the neighbouring stations (also named reference stations) to assess the homogeneity of the studied station (named candidate station), presuming neighbouring stations as homogeneous.

Only a limited set of studies (e.g., Ducré-Robitaille et al., 2003; Beaulieu et al., 2008; Domonkos, 2011c; Guijarro, 2013; Yozgatligiland Yazici, 2016) provided comparison exercises between methods to identify the most successful homogenisation procedure. In 2008, the European initiative COST Action ES0601: Advances in Homogenization Methods of Climate Series: an Integrated Approach (HOME), was released "in order to produce standard methods designed to facilitate such comparisons and promote the most efficient methods of homogenisation" (HOME, 2006). This project included a benchmark dataset, comprising monthly datasets of temperature and precipitation values with inserted, and known, inhomogeneities. These inhomogeneities include outliers, breaks, local and global trends and missing data periods. In order to assess the best techniques, the methods were compared and evaluated using performance metrics (Venema et al., 2012).

Following some of the recommendations of the HOME project (HOME, 2011), homogenisation software packages were developed, such as Climatol (Guijarro, 2011) and HOMER (Mestre et al., 2013). Some of the previously developed homogenisation methods were also improved and converted to automatic software packages, becoming updated versions. Along with HOMER, ACMANT, MASH and PRODIGE were considered the best performing homogenisation methods, due to their capabilities of detecting and correcting multiple breakpoints and working with inhomogeneous references.

The HOME project recommended that further research should give priority to the homogenisation of precipitation, given the low number of contributions for precipitation data and their results within the project (HOME, 2011). This recommendation also meets the consideration provided by Auer et al. (2005), referring that precipitation data require much greater effort, as their variability is more spatially complex. In other words, the spatial and temporal correlation between neighbouring stations should be considered when performing homogenisation of precipitation data series.

A geostatistical stochastic approach showed promising results in homogenising precipitation data (Costa et al., 2008a). This work used the direct sequential simulation (DSS) algorithm (Soares, 2001) to calculate the local probability density function (pdf) at a candidate's station location. The algorithm generates realisations of the climate variable through the resampling of the global pdf using the local mean and variance of the candidate station, which are estimated through a spatiotemporal model. The local pdf from each instant in time is used to verify the existence of irregularities: a breakpoint is identified whenever the interval of a specified probability p centred in the local pdf, does not contain the observed (real) value of the candidate station. When an irregularity is identified, Costa and Soares (2009a) proposed to adjust the candidate time series by replacing the inhomogeneous records by the mean (or median) of the pdfs calculated at the candidate's station location for the inhomogeneous periods.

The use of geostatistical models based on stochastic simulation is a reliable option for addressing problems in environmental and earth sciences, if the purpose is to assess the spatial distribution of a certain attribute as well as spatial uncertainty. With respect to the homogenisation of climate data, Costa and Soares (2009a) enumerate the potential advantages of geostatistical simulation over traditional approaches as follows:

- Considers the temporal and spatial correlation between different weather stations;
- Avoids the iterative construction of composite reference series, increasing the contribution of records from closer stations, both in spatial and correlation terms, by accounting for the spatial and temporal dependence between observations;
- Deals with the problem of missing values and varying the availability of stations through time, by using different sets of neighbouring stations at different periods and by including shorter and non-complete records;
- Seems to be able to simultaneously detect multiple breaks;
- Is able to identify breakpoints near the start and end of the time series while traditional approaches have less power in detecting them.

This geostatistical stochastic approach was only applied to 4 candidate stations and compared with popular detection techniques by Costa et al. (2008a). The climate variable considered in these studies was the annual number of wet days (threshold of 1 mm). Hence, the method's capability to homogenise climate data (detect and remove inhomogeneities) requires further research. Moreover, the method assumes that the global pdf is representative of the reference and candidate stations. However, this assumption may not be realistic in many situations, such as when the study area is extensive and includes different climatic zones, or when the local pdf of the candidate station is different from the global pdf due to local circumstances. In order to mitigate this fact, a new version of the geostatistical homogenisation method, which considers the local characteristics of the candidate, should be investigated.

1.3 Relevance

As discussed above, due to the spatial and temporal variability of precipitation, well-established methods for homogeneity testing monthly and sub-monthly precipitation data are lacking (e.g., Auer et al., 2005; Venema et al., 2012). The geostatistical stochastic approach proposed by Costa et al. (2008a), even though promising, as never been comprehensively evaluated. In particular, the detection part of the procedure requires further research, and its homogenisation efficiency has never been assessed. Furthermore, a new homogenisation method based on the geostatistical stochastic approach could be a valuable contribution for the homogenisation of monthly precipitation series, since it could consider the local characteristics of the variable at the temporal and spatial resolution scale.

Regarding one of the recommendations provided by Venema et al. (2012), it is also important to prepare the homogenisation method in order to deal with large data sets, in an easy and seamless manner. Such study would involve the creation of a computer application, and the performance of sensitivity analyses that contribute to the improvement of the homogenisation efficiency.

Furthermore, the geostatistical stochastic approach is a ground breaking interpolation method. Like other interpolation methods, it could be used for the construction of a data series within the range of a discrete and georreferenced set of known data points, for all types of attributes. It could also be used as a homogenisation method for other climate variables, at different temporal resolutions.

1.4 Research questions

According to the previous discussion, fundamental research questions are:

- Is the geostatistical simulation approach more efficient than some of the existent methods in the homogenisation of precipitation data?
- Can the geostatistical simulation approach be improved to account for specific characteristics of the local climatic zone of the candidate station?

1.5 Objectives

Taking into consideration the research questions previously stated, this research has two main objectives:

1. To evaluate the efficiency of the geostatistical simulation approach in the homogenisation of precipitation data;
2. To investigate an extension of the geostatistical stochastic approach for the homogenisation of climate data. In this new method, the local pdf of the candidate station should better estimate the climatic signal of the surrounding area of the candidate station's location.

The specific objectives of the research are:

1. To perform a thorough literature review;
2. To assist in the development of a homogenisation software that comprises the geostatistical approach and the proposed method;
3. To extend the study of Costa et al. (2008a);
4. To investigate the mathematical formulation of the new homogenisation method;
5. To gather and analyse the precipitation data of the HOME benchmark data set;
6. To assess the performance of the geostatistical simulation approach and of the proposed method, considering different parameterization strategies.

1.6 Expected contribution

Expected results of the research encompass an innovative homogenisation algorithm. The local pdf could be characterised at the candidate station’s location in the space-time reference system. This approach could also help dealing with the problem of sparse monitoring networks. If the new method shows to be effective, it will open new perspectives for research on the homogenisation of high temporal resolution data.

1.7 Thesis outline

The following outline describes the content of each of the five sections that are part of the present research.

The current Section 1 stands as the introductory section, including the problem statement, scientific background, relevance, research questions, objectives, and the expected contribution of this research, as well as the outline of the document.

Section 2 depicts the literature review, introducing the main characteristics and a comparison between the existing homogenisation methods. It also includes a list of studies where those homogenisation methods were used, including the study area, the characteristics of the studied variables, and its main conclusions. The text from this section has been published by Ribeiro et al. (2016a).

Section 3 presents the geostatistical approach proposed as the homogenisation method, and provides the results of the homogenisation exercise that was carried out with real data of an annual precipitation index (wet day count) measured in the south of Portugal. This research is an extension of a previous study by Costa et al. (2008a). The text from this section corresponds to the article published by Ribeiro et al. (2016b).

Section 4 describes the homogenisation exercises undertaken with the benchmark data set (prepared by the HOME project), which comprises annual and monthly precipitation time series and the corresponding performance assessment. It also presents the mathematical formulation of the proposed homogenisation method, named gsimcli. The two research questions, previously stated in Section 1.4, are addressed in Section 4. The text from this section has been published by Ribeiro et al. (2016d).

Finally, Section 5 portrays the main conclusions and recommendations for future research.

1.7.1 Publications

As stated before, sections 2, 3, and 4 correspond to three research articles that have been published by international scientific journals. They are presented as published by the journals with the exception of some layout changes (e.g., the bibliographic references have been harmonised in the References section). Table 1 lists the full references of the articles and the corresponding sections where they are presented.

Table 1 – Thesis sections and corresponding publications in international scientific journals.

Section	Reference
2	Ribeiro S, Caineta J, Costa AC. 2016. Review and discussion of homogenisation methods for climate data. *Journal of Physics and Chemistry of the Earth* **94**: 167–179. doi: 10.1016/j.pce.2015.08.007.
3	Ribeiro S, Caineta J, Costa AC, Henriques R. 2016. Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation, *Atmospheric Research* **171**: 147–158. doi: 10.1016/j.atmosres.2015.11.014.
4	Ribeiro S, Caineta J, Costa AC, Henriques H. 2016. gsimcli: a geostatistical procedure for the homogenisation of climatic time series, *International Journal of Climatology*, in press. doi: 10.1002/joc.4929.

In the scientific paper corresponding to Section 2, the author prepared the manuscript of the literature review, whereas the English corrections and suggestions for improvement were provided by the remaining authors and two anonymous reviewers.

In the scientific paper presented in Section 3, the author prepared the original manuscript text and data analyses, under the supervision of Ana Cristina Costa. Júlio Caineta developed the software, supervised by Roberto Henriques, and collaborated in the data analysis. Three anonymous reviewers provided valuable recommendations that lead to an improved manuscript.

In the scientific paper corresponding to Section 4, the author prepared the data sets and gsimcli analyses, as well as the original manuscript text. Júlio Caineta extended the gsimcli software for the implementation of the proposed approach, supervised by Roberto Henriques. Ana Cristina Costa prepared the calculations of the semi-automatic procedure. All authors contributed to the final version of the manuscript text, which was also improved with the recommendations of two anonymous reviewers.

2 Review and discussion of homogenisation methods for climate data[1]

Abstract

The quality of climate data is of extreme relevance, since these data are used in many different contexts. However, few climate time series are free from non-natural irregularities. These inhomogeneities are related to the process of collecting, digitising, processing, transferring, storing and transmitting climate data series. For instance, they can be caused by changes of measuring instrumentation, observing practices or relocation of weather stations. In order to avoid errors and bias in the results of analysis that use those data, it is particularly important to detect and remove those non-natural irregularities prior to their use. Moreover, due to the increase of storage capacity, the recent gathering of massive amounts of weather data implies also a toilsome effort to guarantee its quality. The process of detection and correction of irregularities is named homogenisation. A comprehensive summary and description of the available homogenisation methods is critical to climatologists and other experts, who are looking for a homogenisation method wholly considered as the best. The effectiveness of homogenisation methods depends on the type, temporal resolution and spatial variability of the climatic variable. Several comparison studies have been published so far. However, due to the absence of time series where irregularities are known, only a few of those comparisons indicate the level of success of the homogenisation methods. This article reviews the characteristics of the most important procedures used in the homogenisation of climatic variables based on a thorough literature research. It also summarises many methods applications in order to illustrate their applicability, which may help climatologists and other experts to identify adequate method(s) for their particular needs. This review study also describes comparison studies, which evaluated the efficiency of homogenisation methods, and provides a summary of conclusions and lessons learned regarding good practices for the use of homogenisation methods.

2.1 Introduction

Success in atmospheric modelling, weather forecasting or climate change monitoring depends on the quality of climate data used as input. Long time series without artificial discontinuities in their statistical characteristics are rare (Alexandersson and Moberg, 1997). Those irregularities can be due to climatic factors, or can be related to facts that happened during the process of collecting or recording climate data. Examples of climatic factors are the eruption of a volcano and the emission of its gases and ashes to the atmosphere contributing to the decrease of solar radiation,

[1] Ribeiro S, Caineta J, Costa AC. 2016. Review and discussion of homogenisation methods for climate data. *Journal of Physics and Chemistry of the Earth* **94**: 167–179.doi: 10.1016/j.pce.2015.08.007.

or the effect of the North Atlantic Oscillation in extreme temperature and precipitation records across Europe (Gaffen et al., 2000).

Non-climatic factors may introduce abrupt or gradual changes in the time series (Alexandersson and Moberg, 1997). Examples of the former are changes in the method of measuring and calculating climate values, such as the use of different daily times in the calculation of daily mean temperature (Peterson et al., 1998), change of measurement units (K, º C and º F for temperature) without any notice (Aguilar et al., 2003), changes in the formula for calculation of the variable's average (Puglisi et al., 2010), relocation of a station (Venema et al., 2013), or its repositioning to a different height (Auer et al., 2005). Gradual and soft changes can be exemplified by the presence of a tree or bush growing nearby the weather station, or the development of an urban area on its surroundings – the increasing of nocturnal temperature called the "Urban Heat Island Effect" (Brunet et al., 2006; Li et al., 2004; Sahin and Cigizoglu, 2010). A high number of non-natural irregularities are also introduced during the process of collecting, digitising, processing, transferring, storing and transmitting climate data series (Brunet and Jones, 2011).

These non-climatic factors may introduce artificial discontinuities, or inhomogeneities, in the time series. Such discontinuities can lead to misinterpretations of the studied climate. In order to avoid errors and obtain homogeneous climate time series, non-natural irregularities in climate data series must be detected and removed prior to its use.

Three main types of inhomogeneities can be distinguished: point errors (coming from the observation to transmission and mechanisation processes); breakpoints corresponding to change-points or shifts in the mean (changes of location, instrumentation, observing practices or land use of the surroundings); and trends (sensor decalibration or urban growth) (Guijarro, 2006). Breakpoints are the most frequent form of inhomogeneities, since most technical changes happen abruptly (Domonkos, 2011a). Trend inhomogeneities are generally more difficult to detect, because they may be superimposed on a true climate trend (Easterling and Peterson, 1995).

Homogenisation is known as the process of detecting and correcting inhomogeneities (Aguilar et al., 2003). Another definition is provided by Štěpánek et al. (2006), where homogenisation includes the following steps: detection, verification and possible correction of outliers, creation of reference series, homogeneity testing (various homogeneity tests), determination of inhomogeneities in the light of test results and metadata, adjustment of inhomogeneities and filling in missing values. Mathematics, software and metadata are referred by Szentimrey (2011) as indispensable for homogenisation of climate data.

Recently, the importance of studying extremes of weather and climate required the development of homogenisation methods for climate data series with higher temporal resolution (e.g., daily data) (Brunetti et al., 2012). In case of precipitation, this task became a challenge due to its great variability (Rustemeier et al., 2011). This variability also results in great uncertainty in homogenisation. True climatic fluctuations in daily precipitation may be interpreted as change-points and removed from time series as inhomogeneities. Moreover, the magnitude of inhomogeneities may differ with varying weather situations (Nemec et al., 2013). Another problem is associated with errors linked to the measuring process, particularly during extreme weather events. For example, larger adjustments are likely to be required for precipitation as its recording is strongly affected by wind strength (Auer et al., 2005). Systematic underestimation of snowfall is also a serious problem in areas where a substantial part of precipitation is collected by rain gauges as snow (Auer et al., 2005; Eccel et al., 2012). To overcome these issues, daily homogenisation methods require complex techniques or the improvement of homogenisation methods previously used for monthly and annual climate series. Those homogenisation methods are of paramount importance as those series are the basis for political decisions with socio-economic consequences (Venema et al., 2013).

The present section provides a description and discussion of homogenisation methods for climate data series, and summarises the conclusions of some comparison studies undertaken to assess their efficiency. Section 2.2 addresses the classification of homogenisation methods, Section 2.3 comprises a review of the available homogenisation methods, and Section 2.4 presents several homogenisation software packages. Comparison studies are briefly described in Section 2.5, where it is also given focus to the HOME project (COST Action ES0601). Finally, some conclusions are drawn in Section 2.6.

2.2 Approaches for detecting and correcting inhomogeneities

Homogenisation methods may have different characteristics, depending on the use of metadata, the subjectivity involved, the use of additional climate time series, the capability of detecting multiple breakpoints, etc. Those characteristics are discussed in the following subsections.

2.2.1 Direct and indirect homogenisation methods

Some authors define direct methods as those that are only based on metadata and subjective judgements (e.g., Li-Juan and Zhong-Wei, 2012). Direct methods have also been defined as mathematical algorithms that are able to detect multiple breakpoints in a direct way (e.g., Domonkos, 2011a), or that are able to deal with inhomogeneous reference time series (e.g., Venema et al., 2012). In the following, we will consider the definitions of direct and indirect methods provided by Aguilar et al. (2003) and Peterson et al. (1998). For these authors, direct methods include the use of metadata, the analysis of parallel measurements, and statistical studies of instrument changes. The indirect methodologies consider the use of single station data (absolute approaches), the development of reference time series (relative approaches), and include both subjective and objective methods.

2.2.1.1 Direct methods

Direct methods aim to keep the climate time series homogeneous by anticipating changes in and around a meteorological station and limit their impact on data homogeneity (Aguilar et al., 2003, pp. 30-31). Direct methods rely on registering in the station history a metadata entry describing any change, and on collecting parallel measurements for a long enough period of time or by reproducing the old conditions (Aguilar et al., 2003, pp. 30-31; Peterson et al., 1998). Metadata information can provide precise knowledge of when the discontinuity occurred and what caused it, but correction factors can only be objectively derived from the records of the “new” and “old” conditions or from a plausible correction model.

2.2.1.2 Indirect methods

Indirect methods use a variety of statistical and graphical techniques to test the homogeneity and adjust the data series (Peterson et al., 1998; Szentimrey, 2006a). Many of these procedures use metadata for identifying or validating the discontinuities found in a time series, as recommended by Aguilar et al. (2003, pp. 33). Among the indirect methods, Peterson et al. (1998) also distinguish between subjective and objective approaches. Subjective methods rely mostly on experts’ judgments. Subjective judgement can be useful in the exploratory analysis stage to identify discontinuities, for example by plotting the stations’ data, by using the Double-mass analysis (Kohler, 1949), or by assessing the reliability of metadata.

Domonkos and Štěpánek (2009) define objective detection methods as those that can be applied in automatic way, without any subjective step. Objective homogenisation methods (OHOMs) have become increasingly more complex (e.g., Domonkos, 2006, 2011b). Domonkos (2006) discusses the conditions, advantages, and limitations related to the practical application of many of these methods.

OHOMs search and correct significant inhomogeneities of time series. Their procedures are applied in a fully computerised way, so no subjective decision is needed during the application. These methods are appropriated for the homogenisation of large data sets, and their efficiency can be quantitatively determined. The statistical methods applied in recent OHOMs are as follows (Domonkos, 2011b): calculation of extremes of accumulated anomalies; non-parametric methods relying on rank-order of sample elements; comparison of averages for adjacent sub-periods; regression functions and the calculation of residual sum of squares; maximum likelihood methods; and tendency of separation of sample elements into different clusters around change-points.

Aguilar et al. (2003, pp. 32-40) and Peterson et al. (1998) include in the set of objective methods the group of absolute and relative approaches, which will be detailed in the following sections.

2.2.2 Absolute and relative homogenisation methods

Considering the use of additional climate data series, homogenisation methods can be distinguished in two classes: absolute and relative methods. Absolute methods consider only the time series of a single station to identify and adjust inhomogeneities (candidate station). Relative methods use data from the surrounding stations (reference stations) to homogenise the candidate station. Some relative approaches are based on a pairwise comparison of the candidate time series with the reference stations data, while other methods are based on composite reference series of differences (for temperature or pressure) or ratios (for precipitation) between candidate and reference stations. According to Domonkos (2013a), there are three main approaches for time series comparisons: building one reference series from composite series for each candidate series; using multiple reference comparisons for each candidate series; and using multiple comparisons without defining which are the candidate and the reference series.

When detecting a discontinuity, an absolute method cannot distinguish if it is natural or artificial without the support of the station's history records. Begert et al. (2005) referred a clear limitation in the absolute methods' capacity to separate discontinuities from true climate signals. Same opinion is shared by Guijarro (2011), advising that absolute homogenisation methods are to be avoided in favour of relative methodologies.

Surrounding stations are exposed to almost the same climate signal. Relative homogenisation is favoured when the spatial density and coherence of the climate data series allows it, because the climatic variation that is common for the study region does not appear in the differences between the candidate and nearby stations (Domonkos, 2013a). The difference time series can be used to detect inhomogeneities, but if a break is detected it may be not clear to which of the stations it belongs to. Furthermore, time series typically have more than just one break. These are two of the problems that homogenisation techniques try to solve. Moreover, the difference time series is useless when the whole network has been simultaneously affected by changes. However, such collective changes are usually well documented, otherwise changes can be detected by comparing multiple networks, and thus this situation is not so problematic.

Most of the relative methods can only be effective if the surrounding weather stations are homogeneous, i.e. if they include natural discontinuities only. This fact raises another question: how to select surrounding stations that are free from artificial discontinuities? According to Reeves et al. (2007), a good reference series should be homogeneous and highly correlated with the candidate series. The use of a reference series that is not homogeneous and/or has different climate signals (trends and periodicities) complicates the problem of change-point detection/adjustment. Peterson et al. (1998) mention the use of metadata to determine which nearby stations would not be expected to have inhomogeneities during specific time periods. Another possible solution is to combine data from different reference stations into a composite reference series assumed as homogenised. Szentimrey (2006a) refers that the spatial covariance structure of data series is very important to develop efficient methods addressing reference series creation, difference series constitution or multiple comparisons of series.

Menne and Williams Jr. (2009) discuss the limitations and challenges of many relative homogeneity testing methods, and propose an algorithm that is able to deal with inhomogeneous neighbouring series. Other methods currently address the presence of change points within the reference series (e.g., Caussinus and Mestre, 2004; Domonkos, 2011c, 2015; Mestre et al., 2013; Szentimrey, 1999, 2006b, 2011).

2.2.3 Multiple breakpoint techniques

One of the fundamental problems of homogenisation is that usually more than one breakpoint is present in the candidate time series (Lindau and Venema, 2013). The majority of the statistical homogenisation methods deals with this problem by applying single-breakpoint techniques multiple times. Typically, when a breakpoint is detected, the time series is divided in two subsets of observations at the identified break and the single-breakpoint algorithm is applied separately to each subset of data. This process is repeated until no more breaks are found or the number of observations becomes too small. The disadvantage of this segmentation process is that the same test applied several times on the same observations can increase the risk of false detection (Beaulieu et al., 2009). The most efficient single-breakpoint technique is known as cutting algorithm (Domonkos et al., 2012), which is a hierarchic method for identifying multiple breakpoints proposed by Easterling and Peterson (1995).

Multiple breakpoint methods are those that detect and correct multiple change-points jointly, and not step-by-step. Recent studies indicate that these are the most effective detection procedures (e.g., Domonkos, 2011b; Venema et al., 2012). Multiple breakpoint algorithms use as detection criterion the maximum external variance between the means of constant time segments in between multiple breakpoints (Lindau and Venema, 2016). These methods apply a relatively simple model (step-function) and select the most probable parameters of this model by the examination of all possible combinations of breakpoint positions (Domonkos, 2013a).

2.3 Statistical homogenisation methods and homogenisation procedures

There are many homogenisation methods described in the literature. A chronological review of the development of homogenisation methods for temperature series is provided by Domonkos et al. (2012). This section highlights the most used approaches, as well as the state-of-the-art homogenisation algorithms that are able to handle inhomogeneous reference series and multiple structures of inhomogeneities. The homogenisation techniques are classified by type of approach (Table A.1 of the Appendix A). Statistical techniques were classified based on their characteristics: non-parametric tests, classical tests (traditional techniques), regression models and Bayesian approaches. Techniques that were directly proposed as methods for the homogenisation of climate data series are named “homogenisation procedures”. These procedures may include more than one statistical technique. Moreover, considering the discussion in Section 2.2, the procedures listed in Table A.1 (Appendix A) are classified as objective bearing in mind the definition provided by Domonkos and Štěpánek (2009). Several techniques are used in the detection stage only (qualifying tests), thus they are useful for homogeneity diagnosis. A sample of studies where the referred methods were applied is provided in Table A.2 (Appendix A), to illustrate their applicability regarding the study region, climate variable and temporal resolution.

2.3.1 Non-parametric tests

The most common non-parametric tests used for homogeneity testing are: Von Neumann ratio test (Von Neumann, 1941), Wald-Wolfowitz runs test (Wald and Wolfowitz, 1943), Mann-Kendall test (Mann, 1945; Kendall, 1975), Wilcoxon-Mann-Whitney (Wilcoxon, 1945; Mann and Whitney, 1947), Kruskall-Wallis test (Kruskal, 1952; Kruskal and Wallis, 1952) and Pettitt’s test (Pettitt, 1979).

The Von Neumann ratio test (Von Neumann, 1941) calculates a ratio of the mean square between successive (year-to-year) differences to the variance, which is closely related to the first-order serial correlation coefficient (Talaee et al., 2014). The calculated value of this ratio is an indicator of the presence of irregularities in the series. This test does not provide the information regarding the date of the discontinuity (Costa and Soares, 2009a) and usually it is used together with other homogeneity tests.

The Wald-Wolfowitz runs test (Wald and Wolfowitz, 1943) is a well-known non-parametric test for randomness. It calculates a statistic based on the sum of the number of changes, by comparing every datum from the time series with the median, over time. This test is sensitive to shifts and trends, but gives little information about the probable dates for breaks. This method is not powerful enough to be used individually in the relative homogeneity analysis and must be supported by graphical analysis so to increase the power of overall analysis, and to obtain the probable date and magnitude of the inhomogeneity, as stated by Tayanç et al. (1998).

The Mann-Kendall test (Mann, 1945; Kendall, 1975) has been popularly used for assessing the significance of trend in hydrological time series, such as stream flow and precipitation. This test has proved to be a valuable tool on trend detection, since it provides useful information on the possibility of change tendency of the variables in the future (Yue and Wang, 2004). It has the advantage of not assuming any special form for the data distribution function, while having a power nearly as high as their parametric competitors. For this reason, it is highly recommended by the World Meteorological Organization (WMO) (Mourato et al., 2010).

The Wilcoxon-Mann-Whitney test (Mann and Whitney, 1947; Wilcoxon, 1945) is based in the use of rank order change-point detection (Aguilar et al., 2003). This approach is advisable when the normality of data is in doubt, such as precipitation data. For this variable, normality is easier to achieve in yearly averaged or in accumulated quantities than in monthly data.

The Kruskal-Wallis test (Kruskal, 1952; Kruskal and Wallis, 1952) is used to compare two or more independent groups of data. The Kruskal-Wallis test allows determining if the difference in the average ranks of three or more independent samples is significant. This test verifies if the hypothesis that all the samples came from the same parent population can be safely rejected.

Pettitt's test (Pettitt, 1979) is a non-parametric rank test that detects single break points. The calculated statistic, derived from the Mann-Whitney, achieves the maximum value for the year with the most likely break point. The test is capable of locating the period where a break may occur, but is more sensitive to breaks in the middle of the time series (Wijngaard et al., 2003).

2.3.2 Classical tests

Double mass analysis (Kohler, 1949), Craddock's test (Craddock, 1979), Bivariate test (Potter, 1981), and Buishand Range test (Buishand, 1982) are classified as (statistical) classical tests as they correspond to traditional homogenisation techniques.

The Double-mass analysis (Kohler, 1949) was one of the first techniques specifically proposed for homogeneity testing. The double-mass curve method is performed by plotting the cumulative amounts of the station under consideration against the cumulative amounts of a set of neighbouring stations. The plotted points tend to fall along a straight line under conditions of homogeneity. Cumulative deviations from some average value can alternatively be plotted to verify the time series homogeneity. It is only used during the exploratory analysis of the time series (Costa and Soares, 2009a). For precipitation time series, cumulative deviations are preferred, since changes in the mean amount are easier to be recognised (Buishand, 1982).

The Craddock's test (Craddock, 1979) is a simple statistical method developed to compare annual precipitation records. This test requires a homogeneous reference series or, in some cases, long enough homogeneous sub-periods. It accumulates the normalised differences between the test series and the homogeneous reference series to determine the inhomogeneities (Aguilar et al., 2003). Craddock's test is recommended by Venema et al. (2012). This test was included in two homogenisation packages: HOCLIS (software package for homogenisation of climatological time series) and THOMAS (tool for homogenisation of monthly data series) from ZAMG (Central

Institute for Meteorology and Geodynamics, Austria) and MeteoSwiss (Federal Office of Meteorology and Climatology, Switzerland), respectively (Auer et al., 2005; Begert et al., 2005).

Potter (1981) applied the bivariate test, developed by Maronna and Yohai in 1978, to precipitation annual series. This is a test for detecting a single systematic change in the mean of an independent time series, based on a second correlated series which is assumed as unchanged (Aguilar et al., 2003). Potter's method generates a test statistic for each data value and an estimate of the maximum probable offset, or adjustment, for that year (Plummer et al., 1995). It closely resembles the double mass curve analysis (Aguilar et al., 2003).

Buishand (1982) used the cumulative deviations to perform some statistical tests, which were compared with the Von Neumann ratio test. This author concluded that both methods give nearly the same results. The Buishand Range test is more sensitive to breaks in the middle of the time series (Wijngaard et al., 2003).

2.3.3 Regression methods

Three methods using regression models are described: Two-phase regression (Easterling and Peterson, 1995), Multiple linear regression (Vincent, 1998), and the Method of cumulative residuals (Allen et al., 1998).

Easterling and Peterson (1995) developed the Two-phase regression (TPR) model, following the work of Solow (1987) who has constrained two regression functions to meet at the point of the inhomogeneity. These authors modified the previous technique so that the two regression lines do not need to meet at the discontinuity. For a given year (or time unit), one regression line is fitted to the reference series for the previous time interval of that year, and the second regression line is fitted to the second part of the time series. This process is repeated for all the years of the time series. The lowest residual sum of squares between the two regression functions will determine the point of discontinuity.

Vincent (1998) proposed the Multiple linear regression (MLR) homogenisation procedure. This technique consists of four linear regression models, applied in a sequence. The first model determines if the candidate series is homogeneous for the tested time interval. If it is homogeneous, the test will end and the remaining models are not used. If inhomogeneities are found, a second model is estimated to ascertain the existence of an overall trend in the candidate series. If the inhomogeneity found in the first model is not an overall trend, the third model is applied to identify the single step change. The fourth model will define the existence of trends before and after that step. If the four models are applied, it indicates that the candidate series have multiple inhomogeneities. In this case, the candidate time series will be divided at the position of the identified step and each segment will be tested separately, starting from the first model. Ducré-

Robitaille et al. (2003) classified MLR as one of the most robust homogenisation methods. More recently, efficiency tests have shown that its detection skills are often lower than other objective methods (Domonkos, 2011b).

The Method of cumulative residuals (Allen et al., 1998) provides a way to relate data sets from two weather reference stations. For a given weather station with a homogeneous time series (independent variable), the records of a second station (dependent variable) can be considered to be homogeneous if the cumulative residuals from their simple linear regression model are not biased. This is tested by verifying if the residuals are contained within an ellipsis, which depends on the size of the data set, the standard deviation of the tested sample and the probability used to test the hypothesis (80% is commonly used). Costa and Soares (2006) proposed an extension of the cumulative residuals method that takes into consideration the concurrent relationship between several candidate series from the same climatic region. This technique uses the residuals from a Seemingly Unrelated Regression equations (SUR) model instead of the residuals from a simple linear regression model.

2.3.4 Bayesian approaches

Bayesian methods have a different approach from classical techniques. Through a prior distribution, the Bayesian approach acquires some knowledge about the climate variable being studied. That information and the observations are combined in a posterior information, which is used to make inference about the parameters. Their advantage is the formal use of non-experimental sources of information to complement the posterior probability distribution function for the studied variable, comprising the position of the shifts, which can be multimodal or skewed. After specifying a loss function, an estimate of the shift's position can be obtained. Several Bayesian techniques were already used for the homogenisation of climate data series, which are described in this section: Bayesian multiple change-point detection in multiple linear regression (Seidou and Ouarda, 2007), Bayesian change-point in multiple linear regression (Seidou et al., 2007), Bayesian change-point algorithm (Ruggieri, 2013), Bayesian multiple change-points and segmentation algorithm (Hannart and Naveau, 2009), Change-point detection algorithm (Gallagher et al., 2012), and Bayesian Normal Homogeneity Test (Beaulieu et al., 2010).

The Bayesian multiple change-point detection in multiple linear regression (BAMS) (Seidou and Ouarda, 2007) follows a Bayesian linear regression model designed to detect multiple change-points. Its main characteristic is the identification of an unknown number of shifts. This procedure requires two training data sets and a prior distribution on the distance between adjacent change-points, which reveals the assumption of the number of existing change-points (Ruggieri, 2013). Beaulieu et al. (2009) considered this approach effective as it often detects the exact number of shifts in an artificial data set.

The Bayesian change-point in multiple linear regression (BARE) model (Seidou et al., 2007) was designed to infer the position of a single change-point in the parameters of a multiple linear regression equation. Seidou et al. (2007) considered non informative prior distributions for the regression parameters and the variance. The prior for the change-point position is a uniform distribution. The method can also be applied for multiple change-points using a segmentation approach. Beaulieu et al. (2009) compared BAMS and BARE using synthetic series of total annual precipitation data series from Canada. Both techniques had similar detection skills, but BAMS performed better for the series with multiple shifts.

Ruggieri (2013) introduced a Bayesian Change-point Algorithm, which provides uncertainty estimates both in the number and location of change-points through a probabilistic solution to the multiple change-point problem. Two main differences should be referred, when comparing this method to BAMS: the nature of recursion and the prior distributions on the model parameters. This algorithm follows three steps: calculation of the probability density of the data; forward recursion (dynamic programming) and stochastic back-trace via Bayes rule (by sampling the number of change-points, the locations of change-points and the regression parameters for the interval between adjacent change-points). Ruggieri (2013) studied the performance of this method by analysing the irregularities in annual global surface temperature.

Hannart and Naveau (2009) used Bayesian Decision Theory to minimise a cost function for the detection of multiple change-points, the Bayesian multiple change-point and segmentation algorithm. The method identifies subsequences of the time series that isolate a unique change-point. These authors studied the performance of this method, by comparison with other methods using simulated series, and they also applied the method to annual temperature data from 16 weather stations located in France (1882-2007).

Gallagher et al. (2012) proposed a Bayesian homogenisation method, the Change-point detection algorithm, for daily precipitation series. The model can be described as a two-state Markov chain with periodic dynamics. The chain serves to induce dependence in the daily (precipitation) amounts, having two different states (dry or wet). If the state considered for a specific day is wet, the amount of the precipitation is modelled as a positive random variable with a seasonally dependent mean (amounts are distribution-equivalent, but the distribution is not necessarily the same). This method was used to homogenise daily precipitation data from Alaska and Massachusetts.

The Bayesian normal homogeneity test (BNHT) enables the detection of a change in the mean of a single normally distributed time series (Beaulieu et al., 2010). It is applied to a reference series, similarly to SNHT. This test also allows the integration of prior information on the date of the change-point (metadata or expert knowledge). Beaulieu et al. (2010) applied this test to synthetic series of total annual precipitation in Canada.

2.3.5 Homogenisation procedures

Techniques that were directly proposed as methods for the homogenisation of climate data series are summarised in this section: SNHT – Standard Normality Homogeneity Test (Alexandersson, 1986), SNHT with trend (Alexandersson and Moberg, 1997), MASH – Multiple Analysis of Series for Homogenisation (Szentimrey, 1999), PRODIGE (Caussinus and Mestre, 1996, 2004), Geostatistical simulation approach (Costa et al., 2008a), ACMANT – Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series (Domonkos, 2011c), and ACMANT2 for homogenising daily and monthly precipitation series (Domonkos, 2015).

The Standard Normal Homogeneity test (SNHT) (Alexandersson, 1986) is one of the most popular and robust homogenisation methods for climatic variables (Ducré-Robitaille et al., 2003). The application of SNHT begins with the creation of a composite (ratio or difference) series between the station values and some regional reference values assumed homogeneous. This composite series is then standardised. At a given moment, averages are calculated for the previous and the following period of that composite series. If the difference between those averages meets a critical value, a shift is inferred to exist at that moment, and the series is said to be inhomogeneous (Ducré-Robitaille et al., 2003).

Later, Alexandersson and Moberg (1997) improved the SNHT method to extend its detection to trends as well. In this innovative SNHT with trend, the alternative hypothesis is that the change of the mean level is gradual, starting and ending at arbitrary points of time, a and b. A test value is computed for all combinations of a and b. The pair that maximises this value has the highest likelihood for being the starting and ending of the trend section. When an inhomogeneity occurs as a sudden shift, such inhomogeneity will be determined by the trend test to be an abrupt change. SNHT with trend is suitable for gradual trends in climate time series, like the increasing of the urban heat island effect (Moberg and Alexandersson, 1997).

The Multiple Analysis of Series for Homogenisation (MASH) (Szentimrey, 1999, 2006b, 2011) was one of the first multiple breakpoint techniques. Currently, it is based on mutual comparisons of series within the same climatic area, and does not assume a homogenised reference series. Breakpoints commonly identified in the difference series (or ratio series for multiplicative variables) are attributed to the candidate series, since it is the only series presented in all. It is a step by step procedure: the role of the series (candidate or reference) changes gradually in the course of the procedure. MASH can be applied to yearly, seasonal and monthly time series. In the new multiple breakpoint procedure, significance and efficiency are formulated according to the conventional statistics related to types I and II errors, respectively. Additionally to the breakpoints and shifts, confidence intervals are also determined. MASH has turned into a software, where metadata can be used automatically to detect inhomogeneities. This method is included in the HOCLIS-system (Auer et al., 2005). Since MASH v3.01, it is possible to homogenise daily datasets (Szentimrey, 2006b).

Caussinus and Mestre (1996, 2004) proposed a new multiple breakpoint technique named PRODIGE, which is based on penalised likelihood methods. The methodology uses a pairwise comparison for preselecting a set of accidents, which are considered within the framework of a multidimensional approach. This method is based on the principle that the series is reliable between two change-points. Those sections will be used as reference series. Instead of comparing a given series with a reference series whose definition is problematic, the comparisons are performed with all other series, by a series of differences. The series of differences is tested against discontinuities through the Caussinus and Lyazrhi (1997) technique. If a change-point (or an outlier) is constantly detected in all the difference series, it can be attributed to the candidate station. The second step of this method is an overall detection and correction. Those two steps are performed by using moving neighbourhoods. The size and the shape of these neighbourhoods are a compromise between the knowledge of the climatologist about the regional climate and the necessity to have enough data, in order to ensure good estimation. Another technique was later developed on basis of PRODIGE, named ACMANT.

The Geostatistical simulation approach proposed by Costa et al. (2008a) can be summarised as follows (Costa and Soares, 2009a). The Direct Sequential Simulation (DSS) algorithm (Soares, 2001) generates realisations of the climate variable through the resampling of the global probability density function (pdf), using the local mean and variance of the candidate station, which are estimated through a spatiotemporal model. The local pdf for each time instant is used to verify the existence of irregularities: a breakpoint is identified whenever the interval of a specified probability p centred in the local pdf, does not contain the observed (real) value of the candidate station. When an irregularity is identified, the time series can be adjusted by replacing the inhomogeneous record by the mean (or the median) of the pdfs calculated at the candidate station location for the inhomogeneous periods.

Domonkos (2011c) proposed an Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series (ACMANT), which is a relative homogenisation technique applicable to monthly temperature series (Domonkos, 2011d). ACMANT is a fully automatic homogenisation method, and its most relevant characteristics are: (i) harmonisation of examinations in different time-scales (annual or monthly); (ii) use of optimal segmentation and the criterion proposed by Caussinus and Lyazrhi (1997) in the detection of inhomogeneities; and (iii) use of ANOVA for the final corrections of inhomogeneities. ACMANT comprises four main steps: preparation; pre-homogenisation; homogenisation and final adjustments (Domonkos, 2011d). Recently, Domonkos (2015) proposed a new unit for the homogenisation of monthly or daily precipitation series, ACMANT2. This new version takes into consideration the climatic regions of snowy winters, by making a distinction between rainy season and snowy season and by searching the seasonal inhomogeneities with bivariate detection. Another main difference from the previous version of ACMANT is that outlier filtering and detection of short-term inhomogeneities are not included in the homogenisation of precipitation series because, in this case, due to the lack of spatial consistency at short-time scale, a possible identified break is very likely to be a true local extreme and not an erroneous precipitation record. Currently, ACMANT and its unit ACMANT2 are a homogenisation software package.

2.4 Homogenisation software packages

Lately, some of the homogenisation methods already described in the previous sections were developed into software, in order to diminish the time consumed during the homogenisation process and to minimise the interaction of users. The examples described are: Climatol (Guijarro, 2006), RHTest (Wang, 2008), AnClim and ProClimDB (Štěpánek, 2008a, 2008b), USHCN (Menne and Williams Jr., 2009), and HOMER (Mestre et al., 2013).

Climatol (Guijarro, 2006) is a set of routines for climatological applications than run under the cross-platform statistical programming language R. Although it may be applied to daily data, it is generally used in the homogenisation of monthly series. This computational application compares each candidate series with a reference series. Once the reference series has been computed, it can be used to determine which variations in the candidate series are due to the climate variability and which are real inhomogeneities that should be corrected. Climatol avoids the use of regression techniques and enables the use of data from surrounding stations when there is no common period of observation. The comparison between the candidate series and their estimated references allows the detection of point errors, shifts and trends through standard statistical tests. The graphical representations of the results can also be shown. Missing values from the candidate series can be directly replaced by the computed reference values. The application of the method to a dense monthly database indicates the importance of using an iterative strategy, thereby detecting and correcting only the coarser errors in the first place, and leaving the less prominent ones to the following iterations. Literature refers this method as robust and simple. However, the final decision on which inhomogeneities to correct must be complemented with visual inspection of the graphical representations.

The RHTest software package (Wang, 2008) is designed to detect multiple step change-points that might exist in a time series. Its recent version, RHTestV3, includes a fully automatic package. This package comprises two penalised maximal tests, PMF (Penalised Maximal F-test) and PMT (Penalised Maximal T-test). The PMF test allows the tested time series to have a linear trend throughout the whole period of the data record, with the annual cycle, linear trend, and autocorrelation of the base series, being estimated one after the other through iterative procedures, while accounting for all the identified mean shifts (Wang, 2008). No reference series is used in any of these functions. The PMT test assumes the tested time series with zero-trend and Gaussian errors. In this case, a reference series is needed. The base-minus-reference series is tested to identify the position(s) and significance of change-point(s), but a multi-phase regression (MPR) model with a common trend is also fitted to the anomalies of the base series in the end to obtain the final estimates of the magnitude of shifts (Wang, 2008). In the MPR fit, the annual cycle, linear trend, and autocorrelation are estimated sequentially through iterative procedures, while accounting for all the identified mean-shifts.

AnClim (Štěpánek, 2008a) and ProClimDB (Štěpánek, 2008b) were developed as a combination of several features from methods mentioned above. ProClimDB is used for processing whole datasets (finding outliers, combining series, creating reference series, preparing data for homogeneity testing, etc.). AnClim works with one station at a time for homogeneity testing, but automated processing of many stations is enabled as well. Results from homogeneity testing produced by AnClim are imported back to ProClimDB and further processed. Two main steps are carried out (Štěpánek et al., 2009): data quality control and homogenisation. The first step is performed by several methods: (i) analysing difference series between candidate and neighbouring stations through pairwise comparisons; (ii) applying limits derived from interquartile ranges; and (iii) comparing the series values tested with "technical" series created by means of statistical methods for spatial data. In the homogenisation step, SNHT, Bivariate and Two-Phase Regression tests are applied to the series. The criterion for identifying a year of inhomogeneity is the probability of detection of a given year, calculated by the ratio between the number of detections for a given year from all tests results for a given station and the total of all theoretically possible detections. The correction of the inhomogeneity is given by the value of the instant before the detected break plus a calculated correction factor, which is determined by the reference series. Štěpánek et al. (2009) applied AnClim and ProClimDB to daily temperature and precipitation data sets.

Menne and Williams Jr. (2009) developed an automated homogenisation algorithm for monthly data that builds on efficient change-point detection techniques, named USHCN (United States Historical Climatology Network). The pairwise algorithm proposed by those authors is able to detect undocumented breakpoints and to deal with inhomogeneous neighbouring series. The algorithm conducts a pairwise comparison in order to first identify all evidences of change-points, combining those evidences with information about documented changes. The algorithm relies upon a pairwise comparison of series in order to reliably distinguish artificial changes from true climate variability, even when the changes are undocumented. In addition, the algorithm employs a recursive testing strategy to resolve multiple undocumented change-points within a single time series. Lastly, the procedure explicitly looks for abrupt “jumps” as well as local and unrepresentative trends in the series.

HOMER, HOMogenization softwarE in R, is an interactive semi-automatic procedure that explores the best characteristics of other state-of-the-art homogenisation methods (PRODIGE and ACMANT), as well as from Climatol and the cghseg joint-segmentation method (Mestre et al., 2013). Basic quality control and network analysis are adapted from Climatol. Detection can be performed using a partly subjective pairwise comparison technique (adapted from PRODIGE) or, alternatively, by applying the full automatic cghseg detection. HOMER includes the ACMANT capability to coordinate the operations on different time scales (from multiannual to monthly). HOMER also includes the UBRIS (Urban Bias Remaining in Series) procedure, which allows characterising artificial climatic trends, in most cases related to urbanisation.

2.5 Comparison of homogenisation methods

A homogenisation method is considered efficient when is able to overcome two problems: the fact that nearby stations are also inhomogeneous, and the existence of more than one irregularity within the time series (Lindau and Venema, 2013). Depending on the used techniques, some homogenisation methods can be more appropriate for a specific climate variable (e.g., first version of ACMANT for temperature), while others can only be used at a given time scale resolution, providing less efficiency for high temporal resolution data series (e.g., daily observations). In order to assess their efficiency, numerous comparison exercises are described in the literature. This section summarises comparison studies undertaken for homogenisation methods, emphasising the HOME project (COST Action ES0601) in the second sub-section.

2.5.1 Comparison tests

In the past two decades several comparison studies have been published in order to determine the most efficient homogenisation method. A synopsis of those comparison tests is disclosed as Table A.3 (Appendix A), and describes the location, variable and periodicity of the climate time series, the compared tests, and some of the achieved conclusions. Those comparison tests are described by chronological order.

Comparison studies also proved the difficulty of indicating which method is the most efficient. Some of the studies were performed using a set of common homogenisation methods, achieving different conclusions. Climate variables also have influence on the efficiency of the method, due to their variability and temporal resolution. Venema et al. (2012) provide a valuable discussion on many of these comparison tests. Problems related to the choice of efficiency measures and the creation of appropriate test-datasets are discussed by Domonkos (2011b, 2013a).

2.5.2 HOME project (Advances in Homogenisation Methods of Climate Series: An Integrated Approach)

In 2008, a European Cooperation in the field of Scientific and Technical Research, HOME – Advances in Homogenisation Methods of Climate Series: An Integrated Approach (COST Action ES0601), was released to compare, evaluate and develop homogenisation methods (HOME, 2011). New (or extensions of earlier) methods were proposed as homogenisation techniques to test a benchmark data set comprising temperature and precipitation data. HOME's main objective was to achieve a general method for homogenising climate and environmental data sets.

The benchmark data set contains real inhomogeneous data as well as simulated data with inserted inhomogeneities, which comprise outliers, break points and local trends. Missing data was also simulated (on those generated data sets) and a global trend was added. This benchmark was composed of three distinct data sets: inhomogeneous (real) climate networks, surrogated and synthetic data sets. The real data set allows comparisons between the different homogenisation methods, since it is comprised of the most realistic type of data and inhomogeneities. Surrogate data was prepared to reproduce the structure of real data in an accurate way so that it could be used as its substitute. Synthetic data is based on surrogate networks. However, the differences between the stations have been modelled as uncorrelated Gaussian white noise. Later, it was concluded that synthetic data is easier to homogenise than the more realistic surrogate data (Venema et al., 2012).

Twenty-five contributions based on 13 algorithms (including MASH, PRODIGE, USHCN, AnClim, Craddock, RH Test V2, SNHT, ACMANT and Climatol) were submitted before the release of the list of known/inserted inhomogeneities in data sets (blind contributions). Different performance metrics and detection skill scores were calculated for monthly, yearly and decadal scales. The blind contributions ((1) for Temperature, (2) for Precipitation) that had the best metrics considered by HOME are as follows:

- MASH: station and network Centered Root Mean Squared Error (CRMSE) (1), trends (1);
- PRODIGE: station and network CRMSE (1); CRMSE anomalies (2) and trends (2);
- USHCN: station and network CRMSE (1), probability of false detection (1), Heidke skill score (1);
- Craddock test: CMRSE anomalies (1), network CRMSE (1), probability of detection (1), Heidke skill score (1);
- Climatol: Heidke special skill score (2).

From the climatologists' point of view, the most important factor to account for in homogenisation is the methods capability to improve the temporal consistency of the climate time series. In this sense, the CRMSE and the trend error metric are more relevant than detection scores such as the Heidke skill score. On the other hand, results also depend on the averaging scale at which the CRMSE is computed and the period under consideration. Domonkos (2013a) provides a comprehensive discussion on the problems related to the choice of efficiency measures, and summarises the results of the blind test experiment of the HOME project. For a more thorough discussion on the assessment of the contributions performance see Venema et al. (2012). There was only one contribution (PMFred abs) that performed absolute homogenisation, and it produced much more inhomogeneous data.

After the truth was revealed to the participants, some of the blind contributions were improved in order to address problems revealed by the results. The all-over best blind contributions were MASH and PRODIGE. Although more limited regarding some tasks, Craddock also had an excellent performance. The USHCN contribution had the lowest probability of false detection and its general performance was only slightly lower than the other best methods. Hence, besides MASH and PRODIGE, Craddock and USHCN were also recommended for practical use (Domonkos, 2013a; Venema et al., 2012). However, the updated ACMANT late contribution suggested that ACMANT was the most accurate method for temperature (Venema et al., 2012). Improved homogenisation methods were included in software packages and are available at http://www.climatol.eu/DARE (accessed April, 2014).

Some of the conclusions agreed by the participants at the end of the project can be described as follows (HOME, 2011; Venema et al., 2012):

- There is not one ideal metric for homogenisation, but the use of detection scores as sole performance criterion should be discouraged;
- More homogenisation algorithms should implement the automatic use of metadata;
- Within the same climatic area, series share a common climate signal;
- Additive structure of the models seems fairly reasonable: temporal and spatial behaviours are separable;
- At monthly to annual time scales, models focus on correction of the means only;
- Covariance is time independent; residuals are not serially correlated;
- Spatial covariance can play a role. Techniques for estimation of spatial covariance are still to be compared. Based on 1st order differentiation of the series (MASH approach), this simple technique relies on a “smooth climate” assumption. Many parameters have to be estimated, or based on the variography analysis of residuals (PRODIGE approach). This technique relies on the variogram of the residuals. It requires the estimation of few parameters at the cost of modelling the spatial structure, which may be more complex.

2.6 Concluding remarks

The importance of having accurate and precise climate records is the main reason for the development of homogenisation methods. Many techniques proposed in the literature aim to detect artificial discontinuities. However, the correction of time series is a very delicate task, and the availability of stations' history information is extremely important to assist the homogenisation process. Furthermore, the number of procedures to correct the artificial discontinuities is limited. In fact, some researchers choose to exclude from further analysis the inhomogeneous series and those with no metadata available, or only consider the longest homogeneous period in the analysis (e.g., Buishand et al., 2013; Costa and Soares, 2009b; de Lima et al., 2013; Santos and Fragoso, 2013).

An up-to-date list of the most important homogenisation methods for climate data series has been discussed in the previous sections, as well as several homogenisation software packages. A classification of the methods has also been proposed. An extensive review of applications is disclosed in the Appendix A (Table A.1, Table A.2, Table A.1), which may also provide guidance to climatologists and other experts to choose the most appropriate method(s) for a particular climatic region, climate variable and temporal resolution.

Based on the analysis from the comparison studies and on a thorough literature review, it is possible to enunciate the following conclusions:

- Techniques that detect and correct multiple breakpoints and work with inhomogeneous references generally perform better than other methods, namely ACMANT, MASH, PRODIGE and HOMER;
- Relative homogenisation algorithms improve the homogeneity of data;
- Absolute homogenisation methods have the potential of making the data even more inhomogeneous;
- Training of the operator when performing homogenisation is very important;
- Homogenisation algorithms developers should invest more effort into making their software easy to use and to include relevant warnings;
- Currently, automatic and semi-automatic algorithms can perform as well as manual ones;
- The use of metadata and the climatological knowledge of the operator are advantages of manual methods;
- Strengths of automatic methods are their objectivity, reproducibility, and easiness to be applied in large data sets;
- Efficiency tests need the use of simulated test datasets with similar properties to real observational datasets;

- Annual climate data sets achieve better homogenisation results than monthly data sets, which may be due to the increase of variability of data series, when the temporal resolution also increases;
- Given the low number of homogenisation studies for precipitation data and their results, the homogenisation of precipitation should be a priority.

The latter conclusion also meets the consideration provided by Auer et al. (2005), referring that precipitation data requires much greater effort, as their variability is more spatially complex. In other words, the spatial and temporal correlation between neighbouring stations should be included when performing homogenisation (Costa and Soares, 2009a; Eccel et al., 2012), particularly for precipitation.

3 Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation[2]

Abstract

Climate data homogenisation is of major importance in climate change monitoring, validation of weather forecasting, general circulation and regional atmospheric models, modelling of erosion, drought monitoring, among other studies of hydrological and environmental impacts. The reason is that non-climate factors can cause time series discontinuities which may hide the true climatic signal and patterns, thus potentially bias the conclusions of those studies. In the last two decades, many methods have been developed to identify and remove these inhomogeneities. One of those is based on a geostatistical simulation technique (DSS – direct sequential simulation), where local probability density functions (pdf) are calculated at candidate monitoring stations using spatial and temporal neighbouring observations, which then are used for the detection of inhomogeneities. Such approach has been previously applied to detect inhomogeneities in four precipitation series (wet day count) from a network with 66 monitoring stations located in the southern region of Portugal (1980–2001). That study revealed promising results and the potential advantages of geostatistical techniques for inhomogeneities detection in climate time series. This work extends the case study presented before and investigates the application of the geostatistical stochastic approach to ten precipitation series that were previously classified as inhomogeneous by one of six absolute homogeneity tests (Mann–Kendall, Wald–Wolfowitz runs, Von Neumann ratio, Pettit, Buishand range test, and Standard normal homogeneity test (SNHT) for a single break). Moreover, a sensitivity analysis is performed to investigate the number of simulated realisations which should be used to infer the local pdfs with more accuracy. Accordingly, the number of simulations per iteration was increased from 50 to 500, which resulted in a more representative local pdf. As in the previous study, the results are compared with those from the SNHT, Pettitt and Buishand range tests, which were applied to composite (ratio) reference series. The geostatistical procedure also allowed to fill in missing values in the climate data series. Finally, based on several experiments aimed at providing a sensitivity analysis of the procedure, a set of default and recommended settings is provided, which will help other users to apply this method.

[2] Ribeiro S, Caineta J, Costa AC, Henriques R. 2016. Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation, *Atmospheric Research* **171**: 147–158. doi: 10.1016/j.atmosres.2015.11.014.

3.1 Introduction

Several environmental and atmospheric studies depend on climate data, in which precipitation data assume a vital role. However, its measurement and recording is prone to systematic and random errors (Sevruk et al., 2009; Teegavarapu and Chandramouli, 2005). Systematic errors may occur due to the growth of trees or urbanisation around the location of the weather station or to precipitation gauge malfunctions, such as water loss during measurement, adhesion loss on the surface of the gauge and raindrop splash from the collector. Random errors include sporadic faults which happen during the process of collecting, recording and transmitting precipitation data records (Brunet and Jones, 2011). These non-natural errors are critical as they affect the continuity of precipitation data and ultimately influence the results of models that use precipitation as input. Indices calculated from daily precipitation data, such as the number of wet days per year (wet day count), are also influenced by the errors in the measurement. Spurious shifts often have the same magnitude as the climate signal, such as long-term variations, trends or cycles, and might lead to wrong considerations about the results of the studies (Caussinus and Mestre, 2004).

In order to obtain trustful results, climate data should be free from non-climatic irregularities. Hence, the detection and the correction of these errors are absolutely necessary before any reliable climate study is based on instrumental series (Auer et al., 2005; Brunetti et al., 2012; Domonkos, 2013a; Tuomenvirta, 2001). Moreover, the World Meteorological Organization (WMO) emphasises the importance of homogenisation in one of the ten climate monitoring principles: "The quality and homogeneity of data should be regularly assessed as a part of routine operations." (World Meteorological Organization, 2010). Homogenisation includes the following steps (Štěpánek et al., 2006): detection, verification and possible correction of outliers, creation of reference series, homogeneity testing (through various homogeneity tests), determination of inhomogeneities in the light of test results and metadata, adjustment of inhomogeneities and filling in missing values. Various methods have been used in the homogenisation of climate data (Aguilar et al., 2003; Beaulieu et al., 2008; Domonkos et al., 2012; Peterson et al., 1998), and their efficiency is dependent on the climate variable, analysed time period, availability of data or other stations located in the same climatic region which may be used as reference series (Costa and Soares, 2009a). Homogenisation methods can be classified into different groups, depending on their characteristics (Aguilar et al., 2003): objective/subjective, direct/indirect and absolute/relative. Relative methods make use of data from neighbouring stations (called reference stations) for comparison with data series from the candidate station (the station to be homogenised). Absolute methods only consider the data from the candidate station in the detection of inhomogeneities.

Recently, the European initiative (COST Action ES0601) ‘HOME’ (Advances in homogenisation methods of climate series: an integrated approach), evaluated the performance of a set of statistical homogenisation methods, using a benchmark data set of temperature and precipitation. Due to their excellent performance, the algorithms ACMANT, Craddock, MASH, PRODIGE and USHCN are strongly recommended by Venema et al. (2012). These authors also refer the need to give priority to the homogenisation of precipitation, due to the less good results presented by the contributions for precipitation. Moreover, Domonkos et al. (2012) mention the need of further tests to better understand the performance of homogenisation methods. Due to the diversity of the characteristics of climatic time series, it is essential to perform more tests with different data set properties. These authors provide a thorough literature review on the methodological evolution of the homogenisation methods for temperature. Ribeiro et al. (2016a) compare homogenisation methods based on literature reviews and discuss their advantages and disadvantages.

Craddock test (Craddock, 1979) accumulates the normalised differences between the test series and the homogeneous reference series in order to find inhomogeneities. This author applied the method to precipitation time series and concluded that best results were obtained by the use of station pairs with the minimum coefficient of variation of the ratio of the two series. This test is part of the homogenisation package THOMAS, from the Federal Office of Meteorology and Climatology in Switzerland (Begert et al., 2005; Michael Begert, 2015, personal communication).

MASH, Multiple Analysis of Series for Homogenisation (Szentimrey, 1999; 2006b, 2007) is a homogenisation method originally developed for monthly series. This relative method does not assume reference series as homogeneous. It is a multiple breakpoint detection algorithm that increases its performance taking the problem of significance and efficiency in account. Metadata is used automatically, in particular the possible dates of breakpoints. The algorithm also includes a procedure for the evaluation of the homogenisation results. In the version of the MASH algorithm for daily data, the estimation of daily inhomogeneities is based on the monthly inhomogeneities calculated (Lakatos et al., 2008).

Caussinus and Mestre (2004) introduced a new methodology for the detection of inhomogeneities, which included pairwise comparison, step function fitting, the Caussinus and Lyazhri (1997) algorithm, and variance optimisation. This method, later named PRODIGE, is based on the idea that a series is homogeneous between two change points. Pairwise comparisons are then obtained between the candidate series and the other reference series, creating a series of differences. These series are tested against the Caussinus and Lyazrhi technique. If a common breakpoint is detected in all the difference series, it is attributed to the candidate station. The overall detection and correction are performed by moving neighbourhoods. The correction estimation is based on ANOVA.

ACMANT, Adapted Caussinus-Mestre Algorithm for Networks of Temperature Series (Domonkos, 2011c; Domonkos et al., 2011a), is a fully automated and relative homogenisation method, which uses the core of the detection and adjustment methods of the PRODIGE (step function fitting and ANOVA correction segments). It applies a bivariate-test for detecting change points that uses the annual mean and the summer-winter difference.

The USHCN homogenisation method is another automatic homogenisation method applied to the United States Historical Climatology Network (Menne and Williams Jr., 2009). The detection part of this method is composed by an early version of SNHT, the cutting algorithm, a Bayesian-based decision about the form of the inhomogeneities (trend-like inhomogeneities can be detected), and a special purpose significance test. Pairwise comparisons are made in an automated way, and metadata can also be used automatically.

The present study provides a follow-up of a previous study (Costa and Soares, 2009a), where a new detection methodology based on direct sequential simulation (DSS) was tested with very auspicious results. However, due to technology and time limitations, a small number of simulations were performed at that time and the number of candidate series was limited to four. In this study, the number of simulations is increased, some sensitivity experiments are performed, and some conclusions are drawn regarding those analyses. For comparison purposes, the same data set was used, which is composed of 66 stations located in the south of Portugal. The analysed climate variable is the annual number of wet days (threshold of 1 mm), calculated from the measured daily value of precipitation, at each weather station, per year. Two sets of candidate stations are used in different stages of the study: the first set, composed of 4 stations, is used for the sensitivity analysis of the DSS parameters; the second set, comprising 10 stations, is used for the sensitivity analysis of the number of neighbour nodes used in the simulation of each node.

The results of the analysis of both sets of candidate stations are compared with the results achieved by Costa and Soares (2009a) through the Standard normality homogenisation test (SNHT, Alexandersson, 1986), the Buishand range test (Buishand, 1982) and the Pettitt test (Pettitt, 1979). These techniques are commonly used and generally accepted for the detection of inhomogeneities (e.g., Sahin and Cigizoglu (2010); Santos and Fragoso, 2013; Wijngaard et al., 2003). Pandžić and Likso (2010) indicate SNHT as one of the most popular methods. Wijngaard et al. (2003) make a brief description of the advantages and disadvantages of those three tests.

Section 3.2 details the network used in this study. Section 0 briefly describes the methodological framework, particularly the DSS process and the sensitivity analysis methodology. Results are presented in Section 3.4. Finally, some conclusions and future work are stated in Section 3.5.

3.2 Data and study background

The inhomogeneities detection methods were applied to precipitation data from 66 monitoring stations located in the south of Portugal (Figure 1). The annual number of wet days between 1980 and 2001 was used as the studied variable, which was calculated from the daily values of precipitation measured at each station, with a threshold of 1 mm defining a wet day. The annual wet day count was used because it is expected to be representative of important characteristics of variation at the daily scale (Wijngaard et al., 2003). This is one of the extreme climate indices defined by the joint CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI), which may contribute to gain a uniform perspective on observed changes in climate extremes (e.g., Klein Tank et al., 2009). The analysis of changes in climate extremes usually requires daily resolution data, but well-established statistical methods for homogeneity testing daily precipitation data are lacking. According to Wijngaard et al. (2003), this variable generally has a lower variability than the annual amounts, particularly in areas with a large contribution from convective precipitation. These authors also referred the easiness of inhomogeneities detection in this climate index, when compared with annual amounts.

The daily precipitation series were compiled from the European Climate Assessment (ECA) data set and the National System of Water Resources Information (Sistema Nacional de Informação de Recursos Hídricos (SNIRH), currently managed by the Portuguese Environment Agency) database. Data are available through free downloads from the ECA&D project website (http://eca.knmi.nl) and the SNIRH website (http://snirh.apambiente.pt, previously http://snirh.inag.pt), respectively (for more information please refer to Costa and Soares, 2009a).

A complete data set of 96 series was initially subjected to an absolute approach of six statistical tests (Costa and Soares, 2009a, 2009b): Mann-Kendall (Mann, 1945; Kendall, 1975), Wald-Wolfowitz runs test (Wald and Wolfowitz, 1943), Von Neumann ratio test (Von Neumann, 1941), SNHT (Alexandersson, 1986), Pettitt test (Pettitt, 1979), and Buishand range test (Buishand, 1982). Thirty stations whose data series were rejected by at least two of the referred absolute tests were discarded from the network. The remaining 66 stations, which are used in this study, are located in the river basins of Arade, Guadiana, Mira, Ribeiras do Algarve and Sado. A list of codes, names and role (candidate or reference) for the 66 monitoring stations used in the study is presented in the Table B.1 (Appendix B).

The analysis of precipitation time series is of particular importance in areas such as the south of Portugal due to its susceptibility to the desertification phenomenon (Costa and Soares, 2012; Pereira et al., 2006). Being located at the Mediterranean climate region, the south of Portugal is exposed to long periods of drought, causing land degradation through soil erosion, reduction of vegetation cover and water resources, increase of vulnerability to salinisation and exhaustion, and degradation of agricultural lands. Analysing the quality of precipitation time series contributes to the improvement of the input data that can be used in climate studies such as those related to desertification processes (Costa and Soares, 2009b).

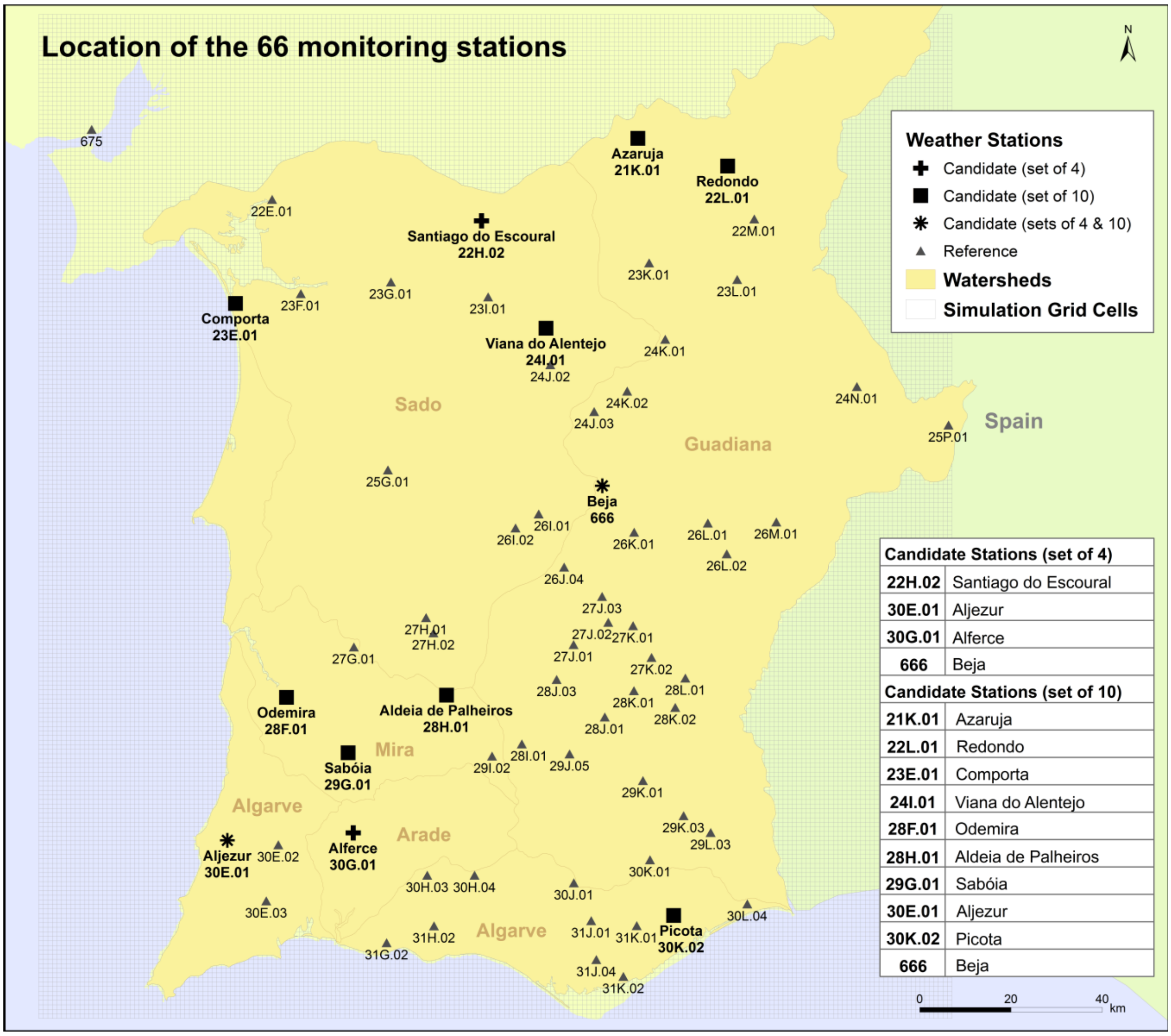

Figure 1 - Location of the 66 monitoring stations in the south of Portugal.

Two sets of candidate stations were defined, containing 4 and 10 stations each (Figure 1). The first set, comprising the stations of Santiago do Escoural (SNIRH 22H.02), Aljezur (SNIRH 30E.01), Alferce (SNIRH 30G.01) and Beja (ECA 666), was used to undertake a sensitivity analysis regarding the number of simulations and other parameters of the DSS method. Those four candidate stations were chosen by Costa et al. (2008a) to illustrate the proposed methodology. The four candidate stations have a long term time series with a common period of 20 years, from 1980 to 1999, with the exception of the Santiago do Escoural station in which the value for the year of 1998 is missing. Those four candidate stations are well spatially distributed in the study area, and they also are representative of the differences from elevation in the study area: 48 m (Aljezur), 243 m (Santiago do Escoural), 246 m (Beja) and 328 m (Alferce).

Table 2 - Length of annual time series for wet day count, per candidate station (dark grey - presence of value, light grey - missing value).

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Azaruja (SNIRH 21K.01)																					
Redondo (SNIRH 22L.01)																					
Comporta (SNIRH 23E.01)																					
Viana do Alentejo (SNIRH 24I.01)																					
Odemira (SNIRH 28F.01)																					
Aldeia dos Palheiros (SNIRH 28H.01)																					
Sabóia (SNIRH 29G.01)																					
Alzejur (SNIRH 30E.01)																					
Picota (SNIRH 30K.02)																					
Beja (ECA 666)																					

The second set was used for the sensitivity analysis of the number of neighbour nodes, and included the following stations: Azaruja (SNIRH 21K.01), Redondo (SNIRH 22L.01), Comporta (SNIRH 23E.01), Viana do Alentejo (SNIRH 24I.01), Odemira (SNIRH 28F.01), Aldeia de Palheiros (SNIRH 28H.01), Sabóia (SNIRH 29G.01), Aljezur (30E.01), Picota (SNIRH 30K.02) and Beja (ECA 666). These ten candidate stations were selected since their data sets were rejected by one of the six above-mentioned absolute tests for homogeneity. Their time series have different lengths (Table 2). The time series from Azaruja and Redondo weather stations comprise three values only of the wet day count index (between 1980 and 1982). For these two stations, the major effect of the geostatistical analysis is expected to be the completion of the time series rather than the detection of inhomogeneities. It is also noteworthy that only two weather stations present wet day count values for the year of 2000: Comporta and Viana do Alentejo stations. Data completion, during this procedure, did not include assigning values for that year.

3.3 Methodological framework

3.3.1 Homogeneity tests

The two sets of candidate stations were analysed using the SNHT, the Buishand range test, and the Pettitt test. The null hypothesis for the three tests is that data are independent, identically distributed random quantities, and the alternative is that a step-wise shift in the mean (a break) is present. If such step cannot be determined in the time series data, the null hypothesis of homogeneity is not rejected.

The application of the SNHT begins with the creation of a ratio (or difference for temperature data) series between the candidate station values and some regional reference values. This composite series is then standardised. At a given moment ν, averages are calculated for the previous and the following period of that composite series. If the difference between those averages meets a critical value, a step is inferred to exist at ν, and the series is said to be inhomogeneous. Two of the most mentioned characteristics of this method are its capability to detect the time period where the breakpoint is likely (month or year) and the skill to easily identify an irregularity at the beginning or at the end of the time series (Ducré-Robitaille et al., 2003).

The application of the Buishand range test starts with the calculation of the sum from the differences between each value of the time series and the mean, at a given time period k. The time series will be considered homogeneous if the sum calculated for each k fluctuates around zero, since no systematic deviations will appear. If the time series is inhomogeneous around k, the sum of the differences will reach a maximum (for a negative shift) or a minimum (positive shift). Buishand (1982) provides critical values to evaluate the significance of the test.

Pettitt (1979) proposed a non-parametric test based on the ranks of the observations, which follows the calculation of test statistics proposed by Mann-Whitney. The test statistic will indicate the presence of a change point when its value is maximal or minimal at a given time period. Pettitt (1979) also provides the significance tables for this test.

The Pettitt test is distribution-free, thus it is applicable to variables with a measurement scale that it is, at least, ordinal. Therefore, applying it to testing variable series of the annual number of wet days is not problematic. However, the SNHT and the Buishand range test assume that data are independent, identically normally distributed random quantities. The wet day count is a discrete variable but, providing that the sample size is large enough, its probability distribution can be approximated by the normal distribution. Costa and Soares (2009b) applied four normality tests (Shapiro-Wilk, Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling) to the testing variable series at 107 monitoring stations. This set of stations comprises the initial set of 96 stations considered in this study. Those authors concluded that the normal distribution fits well the testing variable data, thus the SNHT and the Buishand range test can be applied to the wet day count series. Furthermore, Wijngaard et al. (2003) also detected inhomogeneities in European daily precipitation series by testing series of the number of wet days (threshold 1 mm) using the SNHT for a single break, the Buishand range test, the Pettitt test, and the Von Neumann ratio test (Von Neumann, 1941).

3.3.2 Direct sequential simulation algorithm

In geostatistics, it is common to refer to simulation as a stochastic process, opposed to estimation which is regarded as a deterministic process. Besides correlating the values of different samples of a given variable, geostatistical interpolation adds their spatial structure to the equation. Interpolation usually leads to a smoothing effect of the distribution inferred by the observations and thus to a loss of variance. For example, it is well known that kriging is locally accurate in the minimum error variance sense, but does not provide representations of spatial variability given the smoothing effect of kriging (Yamamoto, 2005). To overcome this limitation, geostatistical stochastic simulation has become a widely accepted procedure to reproduce the spatial variability and uncertainty of highly variable phenomena in geosciences (e.g., Bourennane et al., 2007; Franco et al., 2006; Robertson et al., 2006).

While using the same sequential procedure, some versions of the sequential simulation require different transformations of variables and different approaches to estimate local distribution functions. Examples of those methods are the sequential Gaussian simulation and the sequential indicator simulation (Deutsch and Journel, 1998; Emery, 2004). Following the work of Journel (1994) and Caers (2000), Soares (2001) proposed the direct sequential simulation (DSS) method to reproduce the covariance and the histogram of the variable, a drawback initially found for sequential simulation algorithms without any variable transformation. DSS is also one of the geostatistical simulation methods that has been widely used in different contexts, such as air and water pollutants (e.g., Ribeiro et al., 2014), health (e.g., Oliveira et al., 2013), and climate (e.g., Costa and Soares, 2012; Durão et al., 2010).

Kriging methods used in the simulation process require a stationarity assumption, expressed in two parts. First, the mean of the process is assumed constant and invariant with spatial location (first order stationarity). Second, the variance of the difference between two values is assumed to depend only on the distance between the two points, and not on their location (second order stationarity). Stationarity assumptions on kriging are traditionally accounted for by using local search neighbourhoods so that the dependence on stationarity becomes local (Goovaerts, 1997).

3.3.3 Homogenisation with a geostatistical approach

As previously stated, this work extends the study by Costa and Soares (2009a), where a new method for the homogenisation of climate data was proposed, and the detection phase was illustrated with the data used in the current study. This method integrates the DSS in its algorithm, which serves the purpose of computing the local probability density functions (pdfs) at every candidate station's location, using the spatial and temporal observations of the surrounding reference stations, and excluding the observations of the candidate station itself. Those pdfs can later be used to identify the presence of irregularities at the candidate time series. An observation will be indicated as an inhomogeneity whenever the interval of a specified probability p (e.g. 0.95), centred in the estimated local pdf, does not contain the corresponding real value of the candidate station (Figure 2). Local pdfs are computed by the aggregation of the simulated maps. The method allow the correction of each irregularity (inhomogeneity or outlier) with the replacement of that value by one of the following options: mean, median, or other statistic calculated from the estimated pdf calculated at the candidate station's location for the inhomogeneous period(s). Similarly to Costa and Soares (2009a), irregular and missing values were replaced by the mean of the estimated pdf. Once a candidate station is tested, the corrected time series is included in the detection process of the next candidate station as a reference time series for the calculation of the local pdf. Hence, inhomogeneities detection in the second candidate station benefits from the corrections applied to the first candidate station, the third one will benefit from the previous two, and so on and so forth. These corrections are expected to be especially important for trend-type inhomogeneities.

The DSS algorithm guarantees that the spatial covariance and the global sample mean and variance of the original variable are reproduced, as well as the histogram (Soares, 2001). Hence, the statistical characteristics of the time series are accounted for, even though only individual annual values are examined for inhomogeneities detection purposes. The variance and the spatial correlation of the time series are considered in the semivariogram model used in the ordinary kriging applied during the simulation process. For long-term time series, it is advisable to split the series in smaller sections, in order to guarantee that the statistical properties are consistent within these sections, as recommended by Durão et al. (2010).

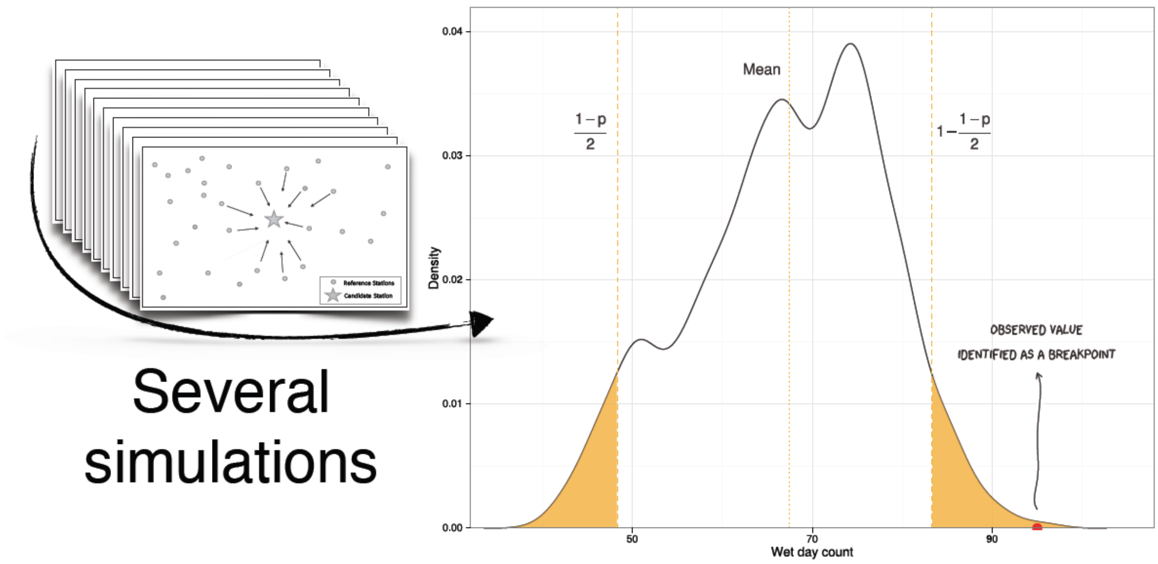

Figure 2 - DSS procedure schema and local pdf for a candidate station.

Some of the potential advantages of this method were mentioned in Costa and Soares (2009a): (i) avoids the iterative construction of composite reference series, increasing the contribution of records from closer stations, both in spatial and correlation terms, by accounting for the joint spatial and temporal dependence between observations; (ii) deals with the problem of missing values and varying availability of stations through time, by using different sets of neighbouring stations at different periods, and by including shorter and non-complete records; (iii) seems to be able to detect multiple breaks; and (iv) is able to identify breakpoints near the start and end of the time series, while traditional approaches have less power in detecting them.

Two stochastic sequential simulation runs were undertaken for each of the candidate stations sets. Both stochastic simulations used the same semivariogram model from the previous study (Costa et al., 2008a): a spherical semivariogram modelled from the complete set of 66 monitoring stations. The spatial dimension was modelled using an isotropic semivariogram model with a range of 72 km, and the temporal dimension was modelled with a range of 1.8 years. Simulations ran in three dimensions (x, y, z), considering time (years) as the z dimension.

For a given candidate station, within the first or second candidate data set, time series from the remaining 65 stations were used. Candidate stations are also used as reference stations in the simulations where they are not being tested, since they are also included in the calculation of the pdfs for the other candidate stations. It is also possible to choose the sequence in which the candidate stations are tested. In the case of the present study, the sequence of candidate stations to be tested was set to the descending order of variance. Assuming that large variance of a time series is an indicator of the presence of inhomogeneities, correcting and completing the data of candidate stations with high variance in the first place is expected to enhance the detection of irregularities in the following candidate stations.

3.3.4 Search parameters and sensitivity analysis

The DSS algorithm generates a set of equally probable realisations for each candidate station, using a set of reference time series, for every unit of time (e.g., every year). Each equally probable realisation is a regular grid of nodes with calculated values. It is possible to manage the set of parameters in the calculation of those realisations, in order to adjust the sequential simulation. Some of those parameters are related to the search of existing values (samples from reference stations and nodes previously calculated in the simulation maps). Search parameters that can be set are described as follows (Deutsch and Journel, 1998):

- Minimum number of data – the minimum number of data (samples or simulated nodes) used in the simulation of each node (minimum value of 1);
- Maximum number of samples – the maximum number of samples used in the simulation of each node (maximum of 64 samples);
- Number of nodes – the maximum number of nodes previously calculated to be considered for the simulation of each node;
- Search radius – maximum distance from the node to be estimated to the samples that may be considered for the calculation of each node; the search radius should cover the entire sampled area in the three directions (x, y, z);
- Search method – two different methods to select the data to be considered for the estimation of the grid nodes: “two part search” searches for samples and estimated grid nodes separately; “data nodes” searches for estimated grid nodes and samples concurrently.

To study the influence of the number of simulations in the detection of the irregularities, different experiments are executed based on the number of undertaken simulations (per candidate station): 50 and 500 simulations. Additionally, two search parameters are tested: search radius and search method. Hence, two sets of tests, comprising four tests each, are established. The first set aims to test the importance of the search radius and the number of simulations, with the “data nodes” search method. The second set tests the number of simulations and the importance of the search radius using the “two part search” method. The provided ranges for the search radius are named as follows: “wide” tests include the entire study area as search radius (220000, 200000, 20 for each of the main directions); and in the “narrow” tests the search radius consists in the variogram ranges (72000, 72000, 1). The minimum and maximum numbers of samples are kept constant in all the tests (1 and 16, respectively), as well as the maximum number of nodes (16). The maximum number of values included in the simulation of a new grid node is 16 for the first set of tests (search method as “data nodes”). In the second set of tests, with the “two part search” method, that maximum number increases to 32 (16 samples plus 16 nodes). In total, eight sensitivity experiments are undertaken (Table 3). All eight experiments are named with the following syntax: “DN”/”2PS” are the acronyms to identify the applied search method (DN – data nodes; 2PS – two part search), the values 50/500 describe the number of simulations computed, and the “narrow”/ ”wide” expressions identify the search radius used in the test (Table 3). It is important to note that if the minimum number of nodes is not found within the search radius, the radius will be ignored and the search will continue until the minimum number of nodes is reached.

The second set of ten candidate stations are later tested for the number of nodes included in the simulation of new grid nodes. The same search parameters as the “DN 500 wide” test are used, except for the number of nodes: 8, 16 and 32 nodes are tested (Table 3).

Table 3 - Search parameters used in the different sensitivity experiments.

Search Parameters	**DN 50 wide**	**DN 50 narrow**	**DN 500 wide**	**DN 500 narrow**	**2PS 50 wide**	**2PS 50 narrow**	**2PS 500 wide**	**2PS 500 narrow**
Number of simulations	50	50	500	500	50	50	500	500
Minimum number of samples	1	1	1	1	1	1	1	1
Maximum number of samples	16	16	16	16	16	16	16	16
Number of nodes	16	16	16	16	16	16	16	16
Search radius (x, y, z)	220000, 200000, 20	72000, 72000, 1	220000, 200000, 20	72000, 72000, 1	220000, 200000, 20	72000, 72000, 1	220000, 200000, 20	72000, 72000, 1
Search method	Data nodes	Data nodes	Data nodes	Data nodes	Two part search	Two part search	Two part search	Two part search

3.4 Results and discussion

3.4.1 Homogenisation of the first set: four candidate stations

The first set of four candidate stations is used to analyse the search parameters. Experiments named DN 50 wide, DN 50 narrow, DN 500 wide, DN 500 narrow, 2PS 50 wide, 2PS 50 narrow, 2PS 500 wide, and 2PS 500 narrow (Table 3) are performed aiming the detection of inhomogeneities for the candidate stations of Aljezur, Alferce, Santiago do Escoural and Beja. The results of these experiments are compared with SNHT, Pettitt and Buishand range tests, which were applied to a composite (ratio) reference series by Costa and Soares (2009a), named hereafter OTHER tests. The results are also compared with the geostatistical approach conducted by Costa and Soares (2009a).

The four candidate stations are considered inhomogeneous by all of the sensitivity tests (Table 4 and Table 5). Comparing the number of performed simulations, the results show that a low number of simulations generally present a high number of detected inhomogeneities. This fact may be explained by the irregularity of the local pdf due to the low number of simulated values used in the pdf calculation (Figure 3).

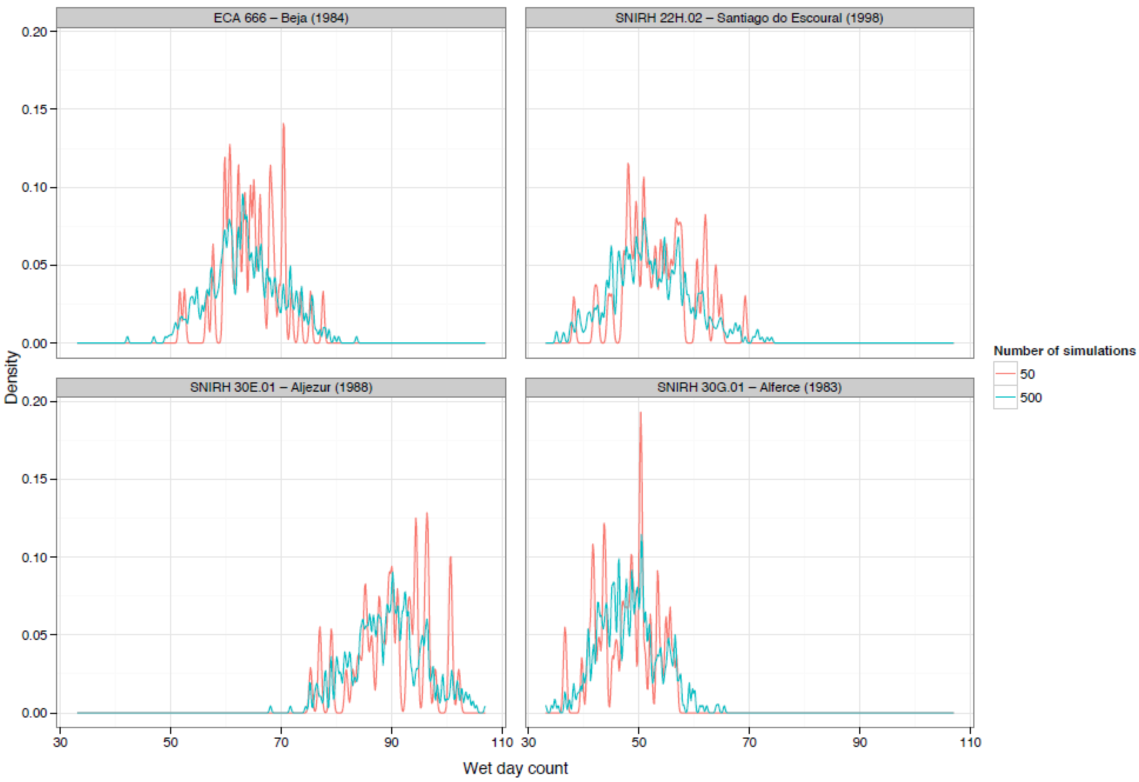

Figure 3 - Local pdfs of four candidate stations computed with 50 and 500 simulations (DN 50 wide and DN 500 wide sensitivity experiments).

Analysing the results between the "wide" and "narrow" experiments, the former presents a low number of detections when compared to the latter (Table 4 and Table 5). In the case of the "wide" tests, the simulated local pdf of the candidate station is characterised by a higher variance due to the use of values that are more distant from the candidate stations, and therefore tend to be more different (Figure 4).

Therefore, the percentile for inhomogeneities detection is also more distant from the mean of the distribution, i.e. the rejection interval is smaller and a lower number of detections is identified. The fact that the "narrow" version is detecting a sequence of years as inhomogeneous might be due to the capability to detect trends. However, a high number of identified irregularities may also correspond to the detection of false positives (i.e., correct values identified as inhomogeneous), which could not be verified because historical metadata was not available

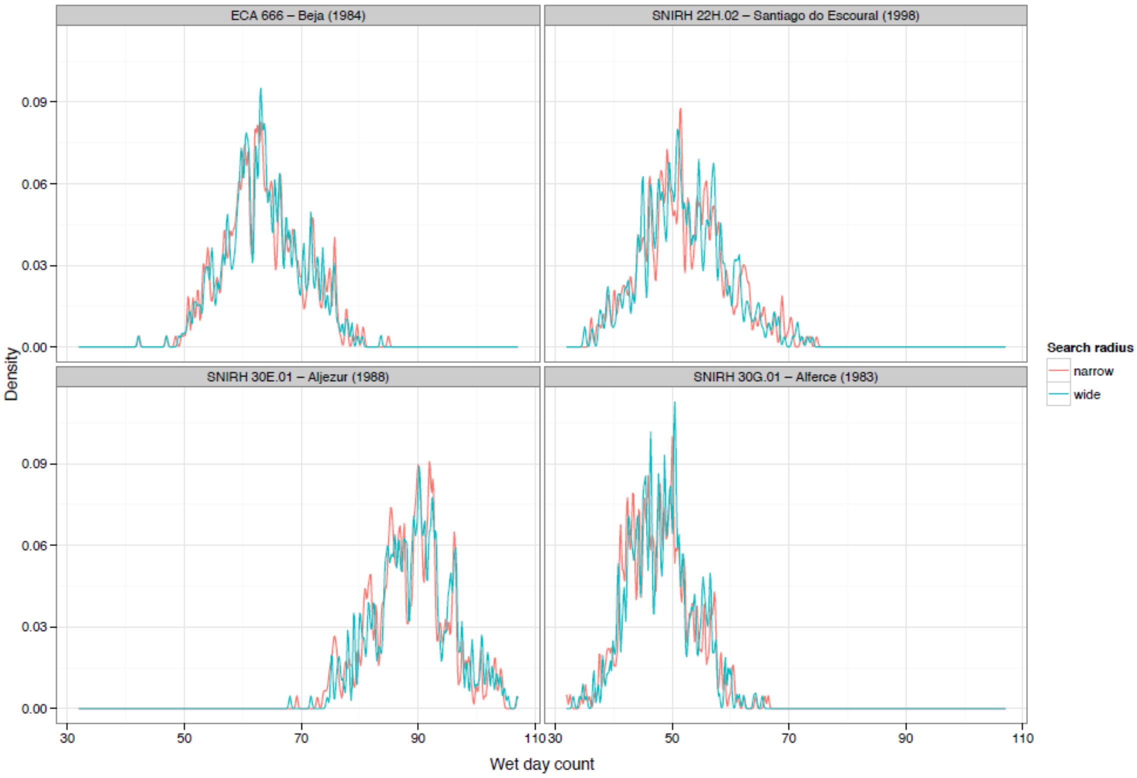

Figure 4 - Local pdfs of four candidate stations computed with "narrow" and "wide" search methods (DN 500 wide and DN 500 narrow sensitivity experiments).

Comparing the "data nodes" and "two part search" experiments, it is only possible to identify a slight increase of detections when the tests are performed with 500 simulations, for the latter (Table 4 and Table 5). However, the tests performed with "two part search" are quite longstanding when compared with the "data nodes" search method. For that reason, and since there are no significant advantages in the use of the "two part search" method it can be concluded that the "data nodes" search method should be the preferred search method.

Table 4 - Inhomogeneities detected for each of the sensitivity experiments (four candidate stations) using the "data nodes" search method.

Stations	DN 50 wide	DN 50 narrow	DN 500 wide	DN 500 narrow	OTHER tests (Costa and Soares, 2009a)	Geostatistical approach (Costa and Soares, 2009a)
Santiago do Escoural ***SNIRH 22H.02***	1984 1987 1989	1984 1987 1988 1989 1996	1989	1987 1989 1996	1988 1989	1987 1988 1996
Aljezur ***SNIRH 30E.01***	1988 1998	1988	1988 1998	1988	Homogeneous	Homogeneous
Alferce ***SNIRH 30G.01***	1983	1983 1999	1983	1983	1984	1983
Beja ***ECA 666***	1988 1992 1996 1997	1987 1996 1997	1996	1996	Homogeneous	1991

Comparing the results per candidate station between the sensitivity experiments and the OTHER tests, some considerations must be stated. In the Santiago do Escoural station, the wet day count value for the year of 1989 is considered inhomogeneous by almost all of the sensitivity experiments and by the OTHER tests. The OTHER tests also detect the year of 1988 as irregular; however, the majority of the sensitivity experiments considered the year of 1987. Regarding Alferce, the year classified as a breakpoint by the sensitivity tests is 1983, while the OTHER tests detected the year of 1984. Those detections corresponding to one-year difference may be considered as the same breakpoint detection (Hannart and Naveau, 2009). For the Aljezur station, the year of 1988 is considered inhomogeneous by the eight sensitivity experiments, while the OTHER tests consider Aljezur as homogeneous. The year of 1996 is commonly detected by the sensitivity experiments in the Beja station, while the OTHER tests consider the station as homogeneous.

The organisation responsible for the monitoring network, SNIRH, has been contacted to provide some historical information (metadata) regarding the detected inhomogeneities. SNIRH communicated the absence of information regarding those irregular years.

Table 5 - Inhomogeneities detected for each of the sensitivity experiments (four candidate stations) using the "two-part search" method.

Stations	2PS 50 wide	2PS 50 narrow	2PS 500 wide	2PS 500 narrow	OTHER tests (Costa and Soares, 2009a)	Geostatistical approach (Costa and Soares, 2009a)
Santiago do Escoural ***SNIRH 22H.02***	1984 1987 1988	1984 1987 1988 1989 1996 1997	1987 1989 1996	1984 1987 1988 1989 1996	1988 1989	1987 1988 1996
Aljezur ***SNIRH 30E.01***	1988 1990 1996 1998	1988	1988 1998	1988	Homogeneous	Homogeneous
Alferce ***SNIRH 30G.01***	1983	1983 1999	1983	1983	1984	1983
Beja ***ECA 666***	1991 1996	1987 1996	1991 1996	1987 1996	Homogeneous	1991

3.4.2 Homogenisation of the second set: ten candidate stations

For the second set with ten candidate stations, three experiments with 500 simulations are carried out with different maximum numbers of nodes (8, 16 and 32). The remaining search parameters are: minimum number of data (1), search radius (220000, 200000, 1 for each search direction), and "data nodes" search method. These settings are assumed to be optimal, based on the results achieved in the previous set of tests: a higher number of simulations leads to a more representative pdf; a low minimum number of data contributes to the absence of non-simulated nodes; a wider search radius broadens the possible range of simulated values, while the spatial correlation is guaranteed by the variogram, which may be preferable when the relation between the pdfs of the candidate station and its neighbours is unknown; and, lastly, the "data nodes" search method is much faster than the "two-part search" method, albeit it provides similar results.

The detected inhomogeneous years for that second set are presented in Table 6. Azaruja, Redondo, Viana do Alentejo, Odemira and Aldeia de Palheiros stations are considered homogeneous by all the DN 500 wide and the OTHER tests. Comporta station is classified as inhomogeneous by the OTHER tests in 1986, but it is considered as a homogeneous time series by the sensitivity experiments. Alzejur and Beja stations are classified as homogeneous by the OTHER tests, whereas all the DN 500 wide tests consider them as inhomogeneous in the years of 1988 and 1996. Sabóia and Picota are considered inhomogeneous by all the tests. In the case of the Sabóia weather station, the inhomogeneous period comprises the years between 1981 and 1986: the DN 500 wide tests consider it irregular in the years of 1981, 1982, 1983 and 1986, while the OTHER tests classify it as inhomogeneous in 1984 and 1985. This fact may indicate the presence of a trend in the beginning of this time series. It may also be due to non-natural changes at that weather station (e.g., change of instrumentation, relocation of the time station, or change in the data collection procedure). In this case, metadata would be an essential auxiliary for the understanding of this inhomogeneous period detected (Trewin, 2013). Regarding the Picota weather station, the year of 1988 is commonly identified as inhomogeneous by all the tests. The DN 500 wide experiments also identified the years of 1993, 1995 and 1998.

Table 6 - Inhomogeneities detected for the second set of ten candidate stations.

<table>
<tr><th>Candidate Stations</th><th colspan="2">8 nodes</th><th colspan="2">16 nodes</th><th colspan="2">32 nodes</th><th>OTHER tests</th></tr>
<tr><td>Azaruja
SNIRH 21K.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>Homogeneous</td></tr>
<tr><td>Redondo
SNIRH 22L.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>Homogeneous</td></tr>
<tr><td>Comporta
SNIRH 23E.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>1986</td></tr>
<tr><td>Viana do Alentejo
SNIRH 24I.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>Homogeneous</td></tr>
<tr><td>Odemira
SNIRH 28F.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>Homogeneous</td></tr>
<tr><td>Aldeia de Palheiros
SNIRH 28H.01</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td colspan="2">Homogeneous</td><td>Homogeneous</td></tr>
<tr><td>Sabóia
SNIRH 29G.01</td><td>1981
1983</td><td>1986</td><td>1981
1982</td><td>1983
1986</td><td>1981
1982</td><td>1983
1986</td><td>1984
1985</td></tr>
<tr><td>Aljezur
SNIRH 30E.01</td><td colspan="2">1988</td><td colspan="2">1988</td><td colspan="2">1988</td><td>Homogeneous</td></tr>
<tr><td>Picota
SNIRH 30K.02</td><td>1988
1993</td><td>1995
1998</td><td>1988
1993</td><td>1995
1998</td><td>1988
1993</td><td>1995
1998</td><td>1988</td></tr>
<tr><td>Beja
ECA 666</td><td colspan="2">1996</td><td colspan="2">1996</td><td colspan="2">1996</td><td>Homogeneous</td></tr>
</table>

Concerning the two stations included in both the first and second test sets, Aljezur and Beja, and in particular Aljezur, it is important to note that the detected inhomogeneities are different. For this station, two years are detected in the DN 500 wide experiment, when tested as part of the set containing four candidate stations (the years of 1988 and 1998). In the second set, the Aljezur station only has a breakpoint in 1988. This may be explained by the fact that the second test set uses references with different data, as some of them were tested and corrected when they previously assumed the role of candidates. These three tests, for the sensitivity of the maximum number of nodes included in the simulation, prove that increasing the number of nodes does not provide a substantial additional proficiency in the detection of inhomogeneities, as the detected irregularities are almost the same. Moreover, increasing the maximum number of nodes significantly extends the required processing time.

Figure 5 presents the wet day count values per year of corrected and original series for the second set of candidate series. The values of the wet day count for the year 2000 are not calculated. Although the original time series present high variability, the corrected series capture their temporal pattern appropriately in most cases.

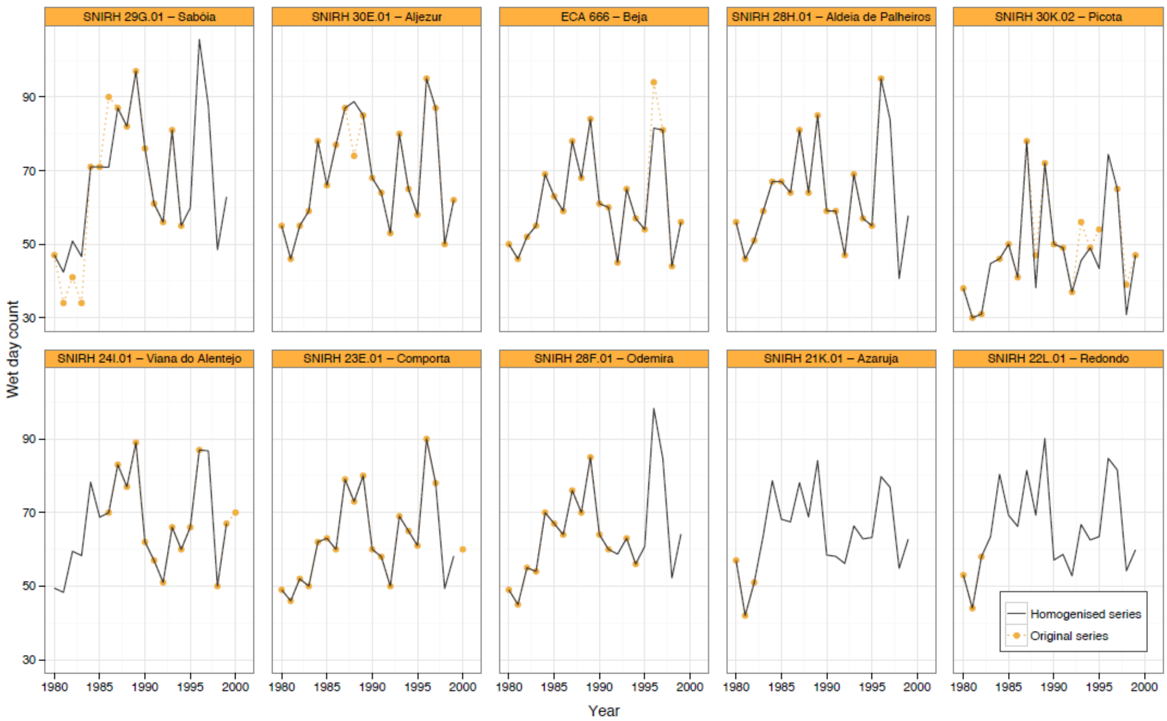

Figure 5 - Corrected versus original time series per candidate station.

3.5 Concluding remarks

Several sensitivity experiments were conducted in order to evaluate the performance of a method based on DSS for the detection of inhomogeneities in climate data series, continuing a previous study undertaken by Costa et al. (2008a). In this sense, the geostatistical approach was used as a qualifying method for quality control, being compared with other detection methods. The inhomogeneities detected cannot be considered outliers, but breakpoints, because Costa and Soares (2009b) exhaustively scrutinised the same data set in order to remove the outliers present in the data.

A data set comprised of 66 monitoring weather stations located in the south of Portugal was compiled and the wet day count precipitation index was used as the climate variable. From the initial data set, two smaller sets, comprising four and ten candidate stations each, were selected in order to test some parameters used by the DSS algorithm. The evaluated parameters included the number of simulations and the search neighbourhood specification, thus determining the number of nodes to be included in each simulation of a grid node. It was concluded that this method succeeds in the detection of inhomogeneities for climate data series, since it provides similar results to other popular detection techniques (Costa and Soares, 2009a). Hence, the geostatistical approach has only been evaluated as an inhomogeneities detection technique, so it has not been sufficiently assessed to be considered a homogenisation procedure. Accordingly, the geostatistical approach should be further investigated.

It was also possible to conclude that a higher number of simulations lead to better detection results, since allows estimating the local distribution with higher precision. However, increasing the number of nodes included in the simulations did not bring enough benefits to justify the increasing computing time. Another advantage of the geostatistical approach is the filling in of missing values in the climate data series. The estimation of missing data is one of the most important tasks required in many hydrological modelling studies (Teegavarapu and Chandramouli, 2005). Moreover, the inclusion of new values to replace missing data may similarly contribute to the improvement of the testing of the following candidate stations, since these new data values will also be considered in the process.

It should also be emphasised the importance of metadata to confirm inhomogeneities detection, regarding artificial discontinuities inserted to data series due to changes in the measurement procedure, as also referred in the third monitoring climate principle provided by the WMO: “The details and history of local conditions, instruments, operating procedures, data processing algorithms, and other factors pertinent to interpreting data (metadata) should be documented and treated with the same care as the data themselves.” (World Meteorological Organization, 2010).

Costa and Soares (2009a) considered the geostatistical approach as slow and laborious, since it required a considerable amount of user interaction in the creation of data files and parameters settings prior to its initialisation. For that reason, it was not practical to assess a large number of candidate stations. Nonetheless, that study revealed promising results and proved the potential advantages of geostatistical techniques for inhomogeneities detection in climate time series. The present study brought new developments to the geostatistical approach. The process was enhanced in terms of computational efficiency and ease of application, enabling the increase of the number of candidate stations and the number of simulations.

The performed analyses are very important for the construction of a new software package that uses the DSS in the homogenisation algorithm that should be further investigated. All the steps carried out in the procedure were completed with the assistance of computer scripts which will lead to the development of a new software package. This new package, called gsimcli, is a work in progress project aiming to make the inhomogeneities detection and homogenisation of climate data series easier and more straightforward, with less user interaction, by also including the management and automatic creation of input data files. The set of parameters that provided the best results in the sensitivity analysis (DN 500 wide test with 16 nodes) will be included in gsimcli as the default values.

4 gsimcli: a geostatistical procedure for the homogenisation of climatic time series[3]

Abstract

Climate data homogenisation is of major importance in monitoring climate change and in validating weather forecasts, general circulation and regional atmospheric models, modelling of erosion and drought monitoring, among other impact studies. Discontinuities in the time series, also named inhomogeneities, may lead to biased conclusions in such studies, so they should be detected and corrected. Previous studies have suggested a geostatistical stochastic approach, which uses Direct Sequential Simulation (DSS), as a promising methodology for the homogenisation of precipitation data series. Based on the spatial and temporal correlation between the neighbouring stations, DSS calculates local probability density functions at a candidate station to detect inhomogeneities. Here, we present a new method named gsimcli (Geostatistical SIMulation for the homogenisation of CLImate data), which is an improved and extended version of that approach. This technique is novel in its incorporation of spatial correlation metrics for the homogenisation of climate time series. The method's performance is assessed with annual and monthly precipitation, and monthly temperature data from two regions of the COST-HOME benchmark data set, and the results are compared using performance metrics. We also evaluate a semi-automatic version of the gsimcli method, which performs additional adjustments for sudden shifts. Both gsimcli versions provided similar results in the homogenisation of annual series. The gsimcli method was more efficient in the homogenisation of the benchmark's precipitation series than the original geostatistical approach. The gsimcli approach performed more closely to state-of-the-art procedures in the homogenisation of monthly data than in the homogenisation of annual data. We expect that the proposed procedure will open new perspectives for the development of techniques that detect and correct inhomogeneities in climate data with monthly and sub-monthly resolution.

4.1 Introduction

Climatic time series may be affected by non-natural irregularities caused by sudden or gradual changes on the surrounding environment of the weather station, or changes in the process of measurement and recording of the climate variable (e.g., Aguilar et al., 2003; Brunet and Jones,

[3] Ribeiro S, Caineta J, Costa AC, Henriques H. 2016, gsimcli: a geostatistical procedure for the homogenisation of climatic time series, *International Journal of Climatology*, in press. doi: 10.1002/joc.4929

2011; Trewin, 2010). Station relocations, repositioning at different heights and changes in the instrumentation are examples of the former. Gradual changes may be exemplified by slowly urban development around a weather station, contributing to the phenomenon known as urban heat island effect (Sahin and Cigizoglu, 2010). The presence of inhomogeneities can distort or even hide the true climatic signal, and thus bias the results of studies (e.g., Domonkos, 2013a; Yozgatligil and Yazici, 2016). Several homogenisation methods have been developed in the last decades to detect inhomogeneities and to adjust the climatic time series in order to improve their temporal consistency (Domonkos et al., 2012; Ribeiro et al., 2016a). Homogenisation methods depend on the climate variable (temperature, precipitation, pressure, evaporation), on the temporal resolution of the observations (annual, seasonal, monthly or daily), on the availability of information on the history of the weather station, and on the spatial density of monitoring stations within the study area (Costa and Soares, 2009a). Ribeiro et al. (2016a) classified the homogenisation methods according to their characteristics: non-parametric tests, classical tests, regression methods, Bayesian approaches, and procedures specifically proposed for the homogenisation of climate data series. Those authors also describe comparison studies that evaluated the efficiency of homogenisation methods, and summarise many methods applications. Domonkos et al. (2012) present a chronological review of the theoretical properties of the most relevant statistical tools that have been developed for the homogenisation of temperature series. Aguilar et al. (2003) and the World Meteorological Organization (2010) emphasise the importance of metadata in the homogenisation of climate time series. By using all the available metadata and stations' history, it is possible to anticipate and preview the type of problems that climate data may have and when they should appear. Since this is often unattainable, it is advisable to compare the stations' history with the data analysis, in a double check process.

Homogenisation approaches can be classified as absolute and relative. Absolute methods only consider the climatic time series of the station to be homogenised (candidate station), while relative homogenisation uses time series from neighbouring stations. Absolute homogenisation may be problematic, because it is difficult to determine if changes, or lack of changes, result from non-climatic or climatic influences without the support of the station's history information (Peterson et al., 1998). Absolute approaches are not recommended as they can even introduce more errors into the climate series (Begert et al., 2005; Guijarro, 2011; Venema et al., 2012). Relative homogenisation is preferred when the spatial density and coherence of the observed data allows it (Costa and Soares, 2009a; Domonkos, 2013a; Ribeiro et al., 2016a). Relative homogenisation relies on comparing the candidate time series to multiple reference series from surrounding stations in a pairwise fashion, or to a single composite reference series computed from multiple neighbouring stations (Venema et al., 2012). More specifically, time series comparisons can rely either on building one composite reference series for each candidate series, on using multiple reference comparisons for each candidate series, or on using multiple comparisons without defining which are the candidate and the reference series (Domonkos, 2013a). Composite reference series are usually built as a weighted average of data from surrounding stations by using some measure of statistical similarity between them (Aguilar et al., 2003). The comparison series are computed as the difference (in case of temperature, pressure, etc.) or ratio (precipitation, wind, etc.) between the candidate and the reference. The comparison series are statistically tested, or a penalised likelihood criteria is used, to assess the significance of changes. Homogenisation corrections may be estimated directly from the comparison series as follows (Aguilar et al., 2003). If a series must be adjusted for a sudden shift, a common approach is to calculate separate averages on the comparison series for the two sections defined by the breakpoint. Then, the obtained means are compared by calculating their ratio or their difference, depending on the variable, and the resulting factor is then applied to the inhomogeneous part. When gradual inhomogeneities are detected, the usual approach is to de-trend the inhomogeneous section using the slope calculated on the ratio time series. When multiple references or pairwise estimates are available, a combination of those estimates is used (e.g., a mean or median). A different approach based on multiple reference series is used by MASH ¬– Multiple Analysis of Series for Homogenisation (Szentimrey, 1999), which considers the adjustment-factors as the lower limits of confidence intervals to keep a low false alarm detection rate (Domonkos, 2013a). Once a first correction has been performed, most methods perform a review (Venema et al., 2012).

Aguilar et al. (2003) recommend the adoption of a reverse chronological approach to adjust annual (monthly) series experiencing more than one discontinuity, in which the most recent homogeneous period is used as a standard and earlier periods are adjusted to reflect these current conditions. By doing so, incoming data in the future will still be homogeneous unless further changes occur in the monitoring station. Moreover, even if additional changes take place, another advantage of this strategy is that it allows for easier updating (Auer et al., 2005). Allen and DeGaetano (2000) argue that it is also reasonable to base adjustments on the longest stationary homogeneous period within a station’s record, and then proceed chronologically, but with the decision to adjust earlier or more recent periods again based on the series length. One advantage of this approach is that the quantity of data that is subject to adjustment is minimised.

The selection of the homogenisation procedure is an effortful task. Domonkos (2015) refers three reasons for the complexity of the selection of the homogenisation procedure: first, the applicability of the method highly depends on the properties and the spatial and temporal structure of the climatic records to be homogenised; second, the efficiency of the homogenisation can be measured empirically only with synthetic test data sets, even though the observed efficiency might differ from the true efficiency due to the deviations in the test data set from the real data; and, third, metadata sometimes provide more reliable information than statistical tests. In 2008, the HOME project (COST Action ES0601) gathered a group of climate experts in order to compare, evaluate and develop homogenisation methods using a benchmark dataset of temperature and precipitation series (Venema et al., 2012). To create the COST-HOME benchmark datasets, known inhomogeneities and other data disturbances were inserted. Under this project, 25 contributions based on 13 statistical homogenisation algorithms were submitted before the release of the list of known/inserted inhomogeneities (the "truth") in the data sets (blind contributions), and their results were evaluated with performance metrics. Later, some of the blind contributions were improved to address problems revealed by the results. One of the main conclusions of the HOME project is that the most efficient methods are those that deal with inhomogeneous neighbouring series, as well as with the interactions of multiple breakpoints and their effects on the calculation of correction terms, namely ACMANT (Domonkos et al., 2011a), MASH (Szentimrey, 1999, 2006b, 2007,), PRODIGE (Caussinus and Mestre, 1996, 2004) and USHCN (Menne and Williams Jr., 2009; Menne et al., 2009). According to Domonkos et al. (2012), these procedures provide the reconstruction and preservation of true climatic variability in observational time series with the highest reliability. Although more limited regarding some tasks, the Craddock method (Brunetti et al., 2006; Craddock, 1979) also had an excellent performance and it is recommended for practical use (Domonkos, 2013a; Venema et al., 2012).

Several methods proposed in the literature have been developed as software packages, which intend to reduce the time consumed during the homogenisation process and to minimise the users' interaction. Ribeiro et al. (2016a) describe their main characteristics, namely of ACMANT and its units ACMANT2 (Domonkos, 2015), Climatol (Guijarro, 2006), RHTest (Wang, 2008), AnClim and ProClimDB (Štěpánek, 2008a, 2008b), and HOMER (Mestre et al., 2013). More recently, the ACMANT3 unit has been released (Domonkos and Coll, 2016). Some of the methods recommended by the HOME project are available in HOMER for monthly data, and HOM/SPLIDHOM for daily data (Mestre et al., 2011).

This article presents the gsimcli method, which is an extension of the geostatistical approach proposed by Costa and Soares (2009a) and Costa et al. (2008a). Costa et al. (2008a) proposed to use the DSS – Direct Sequential Simulation algorithm (Soares, 2001) to calculate the local probability density function (pdf) at a candidate station's location. The DSS algorithm generates realisations of the climate variable through the resampling of the global pdf using the local mean and variance of the candidate station, which are estimated through a spatiotemporal model using Ordinary Kriging. The local pdf from each instant in time is then used to verify the existence of irregularities in the candidate station's series. Costa and Soares (2009a) proposed to adjust the candidate series by replacing the inhomogeneous records with the mean (or median) of the pdfs calculated at the candidate station's location for the inhomogeneous periods. The capability of the geostatistical approach to detect inhomogeneities in real precipitation data was tested with very auspicious results by Costa and Soares (2009a) and Ribeiro et al. (2016b). However, the original geostatistical approach was considered slow, laborious and very computationally intensive.

The gsimcli method aims to provide more local information to the calculation of the local pdf of the candidate station, in order to better estimate the climatic signal of its surrounding area. Furthermore, we propose a different approach to adjust for sudden shifts in the inhomogeneous series, which is based on composite reference series derived from the estimated local pdf. Along with the implementation of the new methodology, a software package was developed, also named gsimcli, with the purpose of making its application easier and more direct. The gsimcli software and its source code are freely available on the internet (http://iled.github.io/gsimcli).

The gsimcli method's efficiency was assessed through the homogenisation of annual and monthly precipitation data from surrogate networks of the COST-HOME benchmark. This was also the main type of artificial data considered by researchers under the HOME project, because the surrogate data provide an estimate of the accuracy of the homogenisation algorithms. Unlike most of those researchers, we evaluated the gsimcli method's performance using precipitation data, which is more difficult to homogenise than temperature.

This article is organised as follows. Section 4.2 describes the methodology, including the gsimcli method formulation and the considered performance metrics. The study area and the surrogate precipitation data are addressed in Section 4.3. Several homogenisation exercises have been performed using the (original) geostatistical approach and different implementation strategies of the gsimcli method, as detailed in Section 4.3. The results of the different homogenisation exercises are presented and discussed in Section 4.4. Finally, the conclusion and future work are presented in Section 4.5.

4.2 Methodology

4.2.1 gsimcli method

Climate observations correspond to realisations (outcome values) of a spatiotemporal random variable $Z(u,t)$ that can take a series of values at any location in space u and instant in time t according to a probability distribution. The set of climate data measured at n locations u_α and in t_i time instants is

$$\{\mathrm{z}(\mathrm{u}_\alpha, \mathrm{t}_\mathrm{i}): \alpha = 0, 1, \dots, \mathrm{n}-1; \mathrm{i} = 1, \dots, \mathrm{T}\}, \tag{1}$$

where $\{z(u_0, t_i): i = 1, \dots, T\}$ denotes the set of values of the candidate station, and $\{z(u_\alpha, t_i): \alpha = 1, \dots, n-1;\ i = 1, \dots, T\}$ denotes the set of values of the reference stations. For each instant in time t_i, the DSS algorithm is applied in order to obtain a set of m equally probable realisations of $Z(u, t_i)$ using the whole set of climate data except the $z(u_0, t_i)$ value. In practice, m equally probable surfaces are simulated on a grid without taking into account the candidate's data for the period being tested.

The DSS algorithm simulates directly in the original data space and does not rely on multi-Gaussian assumptions. The simulated surfaces have the same statistical characteristics (auto-covariance, global sample mean and variance, and histogram) of the original variable (Soares, 2001). Because kriging interpolation requires a positive definite model of spatial variability, a variogram model must be specified. For long-term time series, it is advisable to split the series in smaller sections, in order to guarantee that the statistical properties are consistent within these sections, as recommended by Costa et al. (2008b) and Durão et al. (2010). Accordingly, the DSS algorithm should be applied independently on those smaller sections (e.g., by decade).

In the gsimcli method, the local pdf of the candidate station, for each instant in time (t_i), is defined by the set of spatiotemporal random variables that belong to a circular local neighbourhood centred at the candidate station's location:

$$\{Z^k(u_\alpha, t_i): r = 0, \dots, R;\ \alpha = 0, \dots, W_r;\ i = 1, \dots, T;\ k = 1, \dots, m\}, \tag{2}$$

where W_r denotes the number of locations within a circle of radius r (*local radius parameter*) centred at the candidate station location (u_0). Accordingly, the estimated local pdf of the candidate station for a given instant in time t_0 is the set of simulated values:

$$\{z^k(u_\alpha, t_0): r = 0, \dots, R;\ \alpha = 0, \dots, W_r;\ k = 1, \dots, m\} \tag{3}$$

When r = 0, it is implied that the local pdf of the candidate station will only depend on the simulated values at its exact location. This parameter allows estimating the local pdfe of the candidate station with data that contribute to better describe the climatic signal of the area on which the candidate is located. The corresponding empirical cumulative distribution function gives the estimated probability that the variable Z at location u_0 in space and instant t_0 in time is no greater than any given threshold z: $F^*(u_0, t_0; z) = Prob^*\{Z(u_0, t_0) \leq z\}$.

For the detection of irregularities (breakpoints, trend-type inhomogeneities and outliers), the method proceeds as proposed by Costa and Soares (2009a). An irregular record $z(u_0, t_0)$ is identified if the interval of a specified probability p (*detection parameter*, e.g., 0.95), centred in the estimated local pdf of the candidate station for the instant t_0, does not contain the observed $z(u_0, t_0)$ value:

$$Prob^*\{Z(u_0, t_0) \leq z(u_0, t_0)\} < \frac{1-p}{2} \quad \text{or} \quad Prob^*\{Z(u_0, t_0) \leq z(u_0, t_0)\} < 1 - \frac{1-p}{2} \tag{4}$$

The detection and correction of irregularities, as well as missing values filling, are automatic procedures in the original geostatistical approach (Costa and Soares, 2009a). Missing values are replaced by the mean of the local histogram of the candidate station for the corresponding time instant. Irregular values can be replaced by the mean, median, or other statistic (*correction parameter*) of the estimated local pdf for the inhomogeneous period(s). If the correction parameter is set to a percentile value equal to *p* (e.g., 0.95), irregularities are replaced with the percentile (1–p)/2 or 1–(1–p)/2, depending on the irregularities being located in the lower or upper tail of the pdf, respectively. In such case, the values of the percentiles (*p*) used for detection and correction do not have to be the same. The geostatistical approach deals with trend-type inhomogeneities by correcting multiple irregularities within inhomogeneous periods. No further corrections and adjustments have been proposed by Costa and Soares (2009a) and Costa et al. (2008a).

Once a candidate station is tested, the corrected time series is included in the detection process of the next candidate station as a reference time series for the calculation of the local pdf. Therefore, the detection of inhomogeneities in the second candidate station benefits from the corrections applied to the first candidate station, the third one will benefit from the previous two, and so on and so forth. Accordingly, it would be desirable to homogenise the most inhomogeneous series first, but those are unknown when homogenising real data. To overcome this limitation, the homogenisation sequence may be determined by an indicator of the level of the series inhomogeneity, such as the descending order of variance or the decreasing value of the difference between the station average and the network average (network deviation). The gsimcli software includes several alternative options to determine the order in which stations are tested: ID order, network deviation, random, variance (greater or lower), and the sequence specified by the user (e.g., to start with the series with more missing values in order to fill them in).

The automatic gsimcli method, previously described, can be extended to adjust for sudden shifts using a semi-automatic approach. Adjustments should be done cautiously and station history information should be used to support decisions, since corrections may introduce higher errors than the irregularities they try to remove. Moreover, Domonkos et al. (2011b) state that “not correcting some detected breaks may well sometimes lead to more accurate data”.

The homogenisation adjustments are estimated from a comparison series, which is computed as the ratio (in case of precipitation) between the automatically corrected candidate series and the corresponding composite reference series. This reference series is defined by the time series of the means $\bar{z}(u_\alpha, t_i)$ calculated from the local pdfs of the candidate station for each instant in time t_i:

$$\bar{z}(u_\alpha, t_i) = \frac{1}{m+W_r} \sum_{k=1}^{m} \sum_{\alpha=1}^{W_r} z^k(u_\alpha, t_i)\,, \;\; i = 1, \dots, T \tag{5}$$

The reason for considering the time series of the means, instead of another statistic, was that the mean and the median have very similar time series if the number of simulations is high enough (Ribeiro et al., 2016b). Besides, the detection percentile should not be used because its time series has high variability, since in some instances in time it takes values corresponding to the lower tail of the local pdf, and for other instances it corresponds to the upper tail value.

The dates of the detected irregularities, together with the inspection of both the comparison series and the candidate series, serve to judge each detected inhomogeneity as a potential sudden shift, an outlier or a trend-type inhomogeneity. When decisions cannot be supported by stations' metadata, the comparison series can be statistically tested to assess the significance of such changes. In this study, we used the Buishand range test (Buishand, 1982) with a 5% significance level for this purpose.

Outlier and trend-type inhomogeneities are adjusted using the correction parameter, as suggested by Costa and Soares (2009a), before applying any adjustments for sudden-shifts. The dates of sudden shifts are used to divide the comparison series into segments, and separate averages are calculated on each segment. Then, the obtained means are compared by calculating their ratio (in case of precipitation) with the mean of the most recent period. The resulting factors are then applied to the corresponding segments of the automatically corrected candidate series.

4.2.2 Performance metrics

From the climatologists' point of view, efficiency metrics are more appropriate to evaluate the homogenisation methods capability to improve the temporal consistency of the climatic time series than detection scores (Domonkos et al., 2011b). Domonkos (2013a) discusses the problems that arise from the application of the hit rate and detection skill, which are the most traditional efficiency measures used by developers of homogenisation methods. In this study, we used the efficiency metrics proposed by Venema et al. (2012) to assess the homogenisation methods' performance.

A well-known statistical metric for measuring model performance is the root mean square error (RMSE):

$$\mathrm{RMSE(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - t_i)^2}\,, \tag{6}$$

where the x_i are the homogenised values, the t_i are the true (fully homogeneous) values, and n is the sample size. The RMSE can be calculated for various time units of the observed series (e.g., month, year, and decade time units).

Venema et al. (2012) introduced a modified version of RMSE, the Centred RMSE (CRMSE), which is used as a basic accuracy metric of the data at the highest available resolution. The motivation of using CRMSE instead of RMSE in the HOME project was to eliminate the effect of unknown mean station effects (Domonkos, 2013a). The Station CRMSE is defined as the RMSE of the anomalies relative to the mean bias, and it is computed on single station data directly:

$$\mathrm{CRMSE(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - t_i - \overline{\mathrm{X} - \mathrm{T}})^2}\ , \tag{7}$$

where the upper stroke means arithmetical average, and X and T stand for homogenised and true (fully homogeneous) time series, respectively. This metric is similar to the standard deviation of the time series of the difference between the homogenised data and the truth.

The Station CRMSE quantifies the homogenisation efficiency for each station individually. The Network CRMSE measures the efficiency of the homogenisation of the network, as a whole. It is calculated using the mean CRMSE, by network. The Station [Network] Improvement evaluates the enhancement over the inhomogeneous data, and it is computed as the ratio of the Station [Network] CRMSE of the homogenised networks with the Station [Network] CRMSE of the same inhomogeneous networks. As in Venema et al. (2012), data corresponding to missing data or outliers were not taken into account in the above computations.

The performance metrics were also computed for the blind submissions to the HOME project using the homogenised series available at the HOME project's website (http://www.homogenisation.org; accessed May 2016).

4.3 Climate data and homogenisation framework

The HOME project (COST Action ES0601) included the creation of a benchmark data set containing real inhomogeneous data, as well as simulated data with inserted inhomogeneities. Venema et al. (2011) discuss the generation of this benchmark data set, the climate variables considered, which types of data are considered, how they have been produced, the ways to introduce artificial inhomogeneities, and additional specifications such as length, missing data and trends. The benchmark has different types of monthly datasets (temperature and precipitation) organised in three sections: real, surrogate, and synthetic data. Real inhomogeneous data is composed of temperature and precipitation monthly data series from a set of weather stations located in Europe, because of their importance for climate studies, and because they represent two important types of statistics (additive and multiplicative, respectively). These real data sets allow the comparison between different homogenisation methods with the most realistic type of data and inhomogeneities (Venema et al., 2011). The objective of the surrogate data set is to reproduce the structure of measured data accurately enough that it can be used as substitute for measurements. Surrogate climate networks reproduce the temporal cross-correlation structure of existing homogenised networks, as well as the temporal auto-correlation functions of the stations (Venema et al., 2012). Inhomogeneities were random and independently inserted in the surrogate data sets, with a normal distribution of the breakpoint sizes, and they were simultaneously introduced in multiple station series within a simulated network (Venema et al., 2012). These inhomogeneous surrogate data sets also include outliers, missing data periods and local station trends. Additionally, a stochastic nonlinear network-wide trend was added. The synthetic data sets are based on the surrogate networks, though the differences between the stations have been modelled as uncorrelated Gaussian white noise. The statistical properties of the synthetic data are those assumed by most statistical tests used for homogenisation. This data is easier to homogenise than the more realistic surrogate data (Venema et al., 2012).

4.3.1 Study area and data

Only surrogate series from the COST-HOME benchmark were subject to homogenisation. The following describes the precipitation data from networks 5 and 9 that have been homogenised. These networks have nine and five weather stations, respectively, and are both located in France (Figure 6). Network 9 includes five of the nine weather stations from network 5, but the time series are different in the two networks. The benchmark data set comprises precipitation monthly data for a period of 100 years (1900 – 1999). It also contains temporal intervals with missing data, which occur in the first decades (1900 – 1930) and in the beginning of the fifth decade (1940 – 1945). The lack of data intends to mimic the absence of weather stations in the beginning of the century, and the absence of measurements during the Second World War, respectively. Networks 5 and 9 cover a rectangular area of approximately 4000 km^2 (50 km x 80 km). These two networks were selected because they correspond to the precipitation networks homogenised by the MASH Marinova submission to the HOME project (MASH method operated by a first-time user named Marinova) described by Venema et al. (2012).

In this study, the monthly and annual precipitation data from those networks were subject to exploratory data analysis and homogenisation. The annual precipitation series were derived from the monthly series. As expected, the annual and monthly series from all stations have high variability and several potential outliers. Regarding network 5, station 21142001 has the highest precipitation values in the first decades. Considering the data from all nine stations, there are 102 years with missing precipitation data. The correlation coefficients of the stations' annual series vary between 0.496 and 0.847. The lowest correlation corresponds to two stations located at the centre of the network (21142001 and 21425001). The highest correlation corresponds to the stations 21142001 and 21386001. Considering the annual series from network 9, all stations have similar distributions, except station 21584001 that has higher values. The correlation coefficients of the stations' annual series vary between 0.498 (stations 21454001 and 21425001) and 0.883 (stations 21454001 and 21109001).

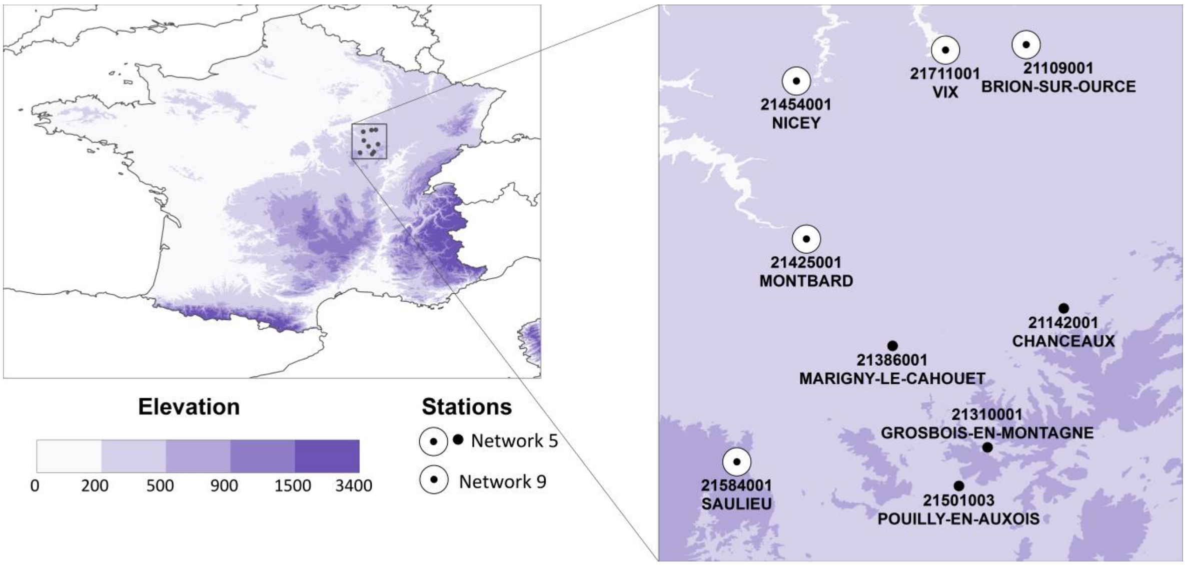

Figure 6 - Location of stations from networks 5 and 9 in France (Digital Elevation Model source: Jarvis et al., 2008).

The main spatial patterns of the precipitation data were also investigated, particularly the presence of anisotropy. An attribute has an anisotropic behaviour when it exhibits a different spatial auto-correlation structure for different directions. The possible existence of anisotropy was analysed by producing an interpolated surface of the precipitation data for a sample of years using the Inverse Distance Weighting (IDW) interpolator. Although time consuming, these analyses were important, because if an attribute shows different auto-correlation structures in different directions, then an anisotropic variogram model should be developed to reflect these differences. The most commonly employed model for anisotropy is the geometric anisotropy, with the variogram reaching the same sill in all directions, but at different ranges. The interpolated surfaces obtained using IDW neither revealed an overall trend, nor an overall anisotropic pattern in any of the networks.

Considering that the variogram modelling is a very important stage of geostatistical methods, a thorough variography analysis was undertaken. Due to the variability of precipitation data, the lack of data in several decades, and, mainly, the reduced size of the monitoring networks, that analysis revealed to be a challenging task. Many experimental variograms exhibit a variability pattern such that the correlation between stations' data seems to be lost at short distances. Accordingly, the spatial features of precipitation occur at scales smaller than the distance between monitoring stations. For this reason, modelling the experimental variograms, and the nugget effect in particular, was not a straightforward task. Moreover, with very widely spaced data, a realistic estimate of the range parameter was also sometimes difficult to obtain. A way to overcome these drawbacks is the use of additional data provided by other weather stations located in the surrounding study area. However, such task could not be performed, since only the provided data sets by the HOME project could be used in the process.

In previous exploratory homogenisation activities with the original geostatistical approach, different variogram models for the spatial continuity structure of the data were assessed (Costa et al., 2015). The variogram models that lead to the best performance metrics are considered in this study. As recommended by Costa et al. (2008b) and Durão et al. (2010), variograms were first estimated by decade in order to account for possible long-term trends, or fluctuations, in the precipitation auto-correlation structure. This approach was followed in the case of annual data from network 5 (Table 7). Due to the small number of stations in network 9, a single variogram model for the whole 1900–1999 period was estimated for the annual data (Table 7). Previous exploratory homogenisation activities indicated that using all yearly data to infer a single variogram model for network 9 provided similar results to using the same decadal variogram models inferred for network 5 (Costa et al., 2015). Hence, estimating a single variogram model for the whole period is the recommended solution in case of small networks.

Table 7 - Variogram models of the annual precipitation series from networks 5 and 9.

Decade	Model	Nugget	Range	Partial Sill
Network 5				
1900 - 1909	Exponential	11000	26000	55000
1910 - 1919	Exponential	2500	24000	34000
1920 - 1929	Exponential	2000	19000	52000
1930 - 1939	Exponential	6500	20000	47500
1940 - 1949	Exponential	0	22000	43000
1950 - 1959	Exponential	4500	23000	26000
1960 - 1969	Exponential	10000	20000	42500
1970 - 1979	Exponential	6500	18000	26000
1980 - 1989	Exponential	8000	20000	32000
1990 - 1999	Exponential	3000	20500	24000
Network 9				
1900 - 1999	Exponential	0	27500	8700

Due to the lack of data in the monthly series, a unique variogram model was estimated for the first, second and third decades (1900–1929) from network 5, for each month (Table C.1 of the Appendix C). For the same reason, the fourth and fifth decades' data were also combined in order to obtain another single variogram model. Seven variogram models were prepared for each monthly series, in a total of 84. The estimated variogram models for network 5 were also used in network 9 (Table C.1 of the Appendix C), since the reduced number of stations in this network did not allow to obtain a reliable estimate of the variogram model.

4.3.2 Specifications of the homogenisation exercises

Several homogenisation exercises were undertaken for the precipitation networks 5 and 9 from the COST-HOME benchmark using different sets of parameters (Table 8). We investigated the impact of two strategies on the definition of the simulation grids. The homogenisation exercises used a regular grid comprising 9882 cells (81 x 122 cells) for a grid cell size of 1 km, except one exercise with annual data that used a regular grid with 425 cells (17 x 25 cells) having a cell size of 5 km. Different values of the local radius parameter (*r*) were also considered, ranging from 1 to 5 cells (Table 8). All homogenisation exercises with the gsimcli method used the following common set of parameters:

- Candidates order = descending order of the stations' data variance;
- Number of simulations (*m*) = 500;
- Detection parameter (*p*) = 0.95;
- Correction parameter =percentile value of 0.975.

Table 8 - Parameters of the homogenisation exercises with the gsimcli method.

Test #	Grid cell size	Local radius parameter (r)
Annual time series		
1	1000 m	1
2	1000 m	2
3	1000 m	3
4	1000 m	4
5	1000 m	5
6	5000 m	1
Monthly time series		
7	1000 m	0
8	1000 m	1
9	1000 m	2

The annual series were homogenised using both the automatic and semi-automatic versions of gsimcli. Considering that the later did not significantly improve the method's efficiency, the monthly series were only homogenised using the automatic gsimcli. In the adjustments stage of the annual series, whenever the candidate series had missing values in the beginning of the time series, it was considered the existence of a breakpoint at the first date with data in the automatically corrected candidate series. Therefore, the missing values that were automatically estimated were also adjusted, despite the fact that these data are not used in the computation of the performance metrics.

The original geostatistical approach (Costa and Soares, 2009a) was also used to homogenise the annual time series from networks 5 and 9. This was accomplished by setting the local radius parameter (r) to zero, and the correction parameter to the mean of the local pdf of the candidate station. No further adjustments were applied in these homogenisation exercises. The simulation grid was defined with cells of 1 km^2.

4.4 Results and discussion

The automatic gsimcli method homogenises candidate time series using the correction parameter derived from the estimated local pdf for the inhomogeneous periods. In the semi-automatic version, the automatically corrected candidate series are further adjusted using correction factors derived from comparison series. These are based on composite reference series corresponding to the series of means computed from the estimated local pdfs. The different parameters used in the homogenisation of the precipitation series are described in Section 4.3.2. The following sections detail the results of the precipitation data homogenisation.

4.4.1 Annual precipitation series

For illustration purposes, Figure 7 shows the candidate time series of station 21142001 from network 5, and the homogenised series that were obtained using the gsimcli method with the parameters specified for Test #6 (Table 8), as well as the corresponding composite reference series and comparison series. The Buishand range test identified a significant sudden shift in 1952 in this candidate time series. No other significant breakpoints were identified in the segments before and after this year.

The irregular years identified in the automatic stage of gsimcli, as well as the years corresponding to significant sudden-shifts identified by the Buishand's test are listed in Table C.2 (Appendix C). This table also presents the years defined by HOME project as breakpoints and outliers. It is important to note that these irregularities were introduced in the monthly time series of the benchmark data set. Certain inhomogeneities might only be evident at certain timescales of variability (Yan and Jones, 2008). In this study, those monthly irregularities were considered as annual breakpoints for comparison purposes, thus the detection results should be analysed with caution. Those inhomogeneities might not be detected as breakpoints in the homogenisation exercises, since the annual amounts of precipitation may smooth those monthly irregularities.

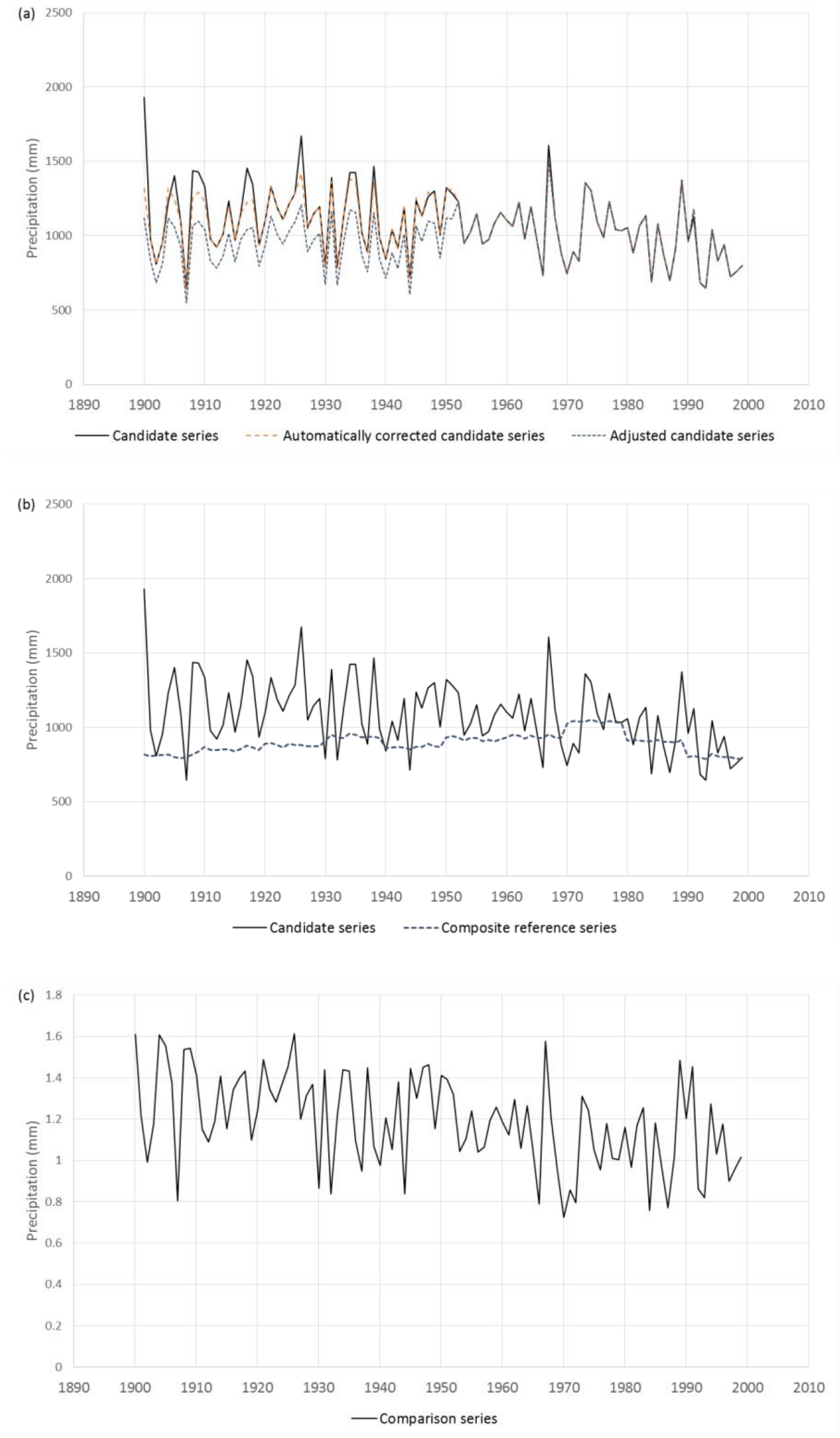

Figure 7 - Graphic (a) shows the candidate time series of station 21142001 from network 5, and the corresponding homogenised series using the automatic and semi-automatic versions of gsimcli with the parameters specified for Test # 6. Other graphics show the comcomposite reference series (b) and the comparison series (c) used in the homogenisation of this candidate series with the semi-automatic gsimcli.

The number of years with detected irregularities by the automatic gsimcli does not seem to be dependent of the local radius parameter (*r*), since it is similar in the different homogenisation exercises. It is higher than the number of breakpoints defined by the HOME project. In some cases, a sequence of more than two consecutive years with irregularities is detected, which can be assumed as the detection of a trend-type inhomogeneity in the candidate series by the automatic gsimcli (e.g., in the station 21454001 from network 5 there are breakpoints detected consecutively from 1911 to 1914, from 1940 to 1946, and from 1987 to 1993).

The breakpoint years detected by the Buishand’s test are similar for all homogenisation exercises, varying from zero to three, thus the performance metrics obtained with the automatic and semi-automatic gsimcli are also similar (Table 9).The automatically corrected candidate series from stations 21454001, 21501003 and 21584001 from network 5, and station 21109001 from network 9, were considered as homogeneous by the Buishand’s test in all homogenisation exercises. One additional breakpoint year (1984) was identified in the automatically corrected candidate series from station 21310001 from network 5, and two additional years (1906 and 1918) in station 21584001 from network 9, using the homogenisation Test #6. Only one breakpoint year (1926) was identified in station 21425001 from network 9 using the homogenisation Test #1, whereas all other homogenisation exercises identified two breakpoint years (1917 and 1926) in this station. The breakpoint year of 1937 was not identified in station 21711001 from network 9 using the homogenisation Test #2.

The performance metrics were computed for the homogenisation exercises considering the application of the automatic gsimcli method (without adjustments for sudden shifts), and the semi-automatic version (with the additional adjustments stage) (Table 9 and Table 10). The performance metrics were also computed for the homogenisation activities undertaken with the original geostatistical approach, and for the blind submissions to the HOME project that homogenised networks 5 and 9. All the homogenisation exercises undertaken with the annual precipitation data from network 9 (Table 10) made the data more inhomogeneous, i.e. had a Station improvement quotient over the inhomogeneous data above one. However, the original geostatistical approach was the only homogenisation activity undertaken that made the data from network 5 (Table 9) more inhomogeneous. The higher number of stations in network 5 might explain the better results obtained for this network than for network 9. All the values of the Station CRMSE of the gsimcli method are at least 24% smaller than those of the original geostatistical approach. Considering the Network CRMSE, the efficiency increase of the gsimcli method is greater for the automatic version (at least 44%) than for the semi-automatic one (at least 24%). Accordingly, the gsimcli method is more efficient than the original geostatistical approach. Nonetheless, the gsimcli method underperformed all the blind submissions to the HOME project, except the absolute method (h008 - PMFred abs) for network 5.

Table 9 - Performance metrics of the annual precipitation series from network 5 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Test #	Adjustment method	Station CRMSE	Station Improvement	Network CRMSE	Network Improvement
1	**Automatic gsimcli**	7.13	0.99	2.88	1.07
2	**Automatic gsimcli**	7.13	0.99	2.9	1.08
3	**Automatic gsimcli**	7.12	0.98	2.88	1.07
4	**Automatic gsimcli**	7.14	0.99	2.88	1.07
5	**Automatic gsimcli**	7.13	0.98	2.9	1.08
6	**Automatic gsimcli**	7.09	0.98	2.86	1.07
1	**Semi-automatic gsimcli**	7.03	0.97	3.9	1.45
2	**Semi-automatic gsimcli**	7.03	0.97	3.93	1.46
3	**Semi-automatic gsimcli**	7.02	0.97	3.9	1.45
4	**Semi-automatic gsimcli**	7.03	0.97	3.9	1.45
5	**Semi-automatic gsimcli**	7.03	0.97	3.93	1.46
6	**Semi-automatic gsimcli**	6.93	0.96	3.78	1.41
Original geostatistical approach		9.38	1.3	5.19	1.93
Inhomogeneous data		*7.232*	*1.0*	*2.685*	*1.0*
h002 - PRODIGE main		3.948	0.546	2.525	0.940
h006 - C3SNHT		5.556	0.768	2.588	0.964
h007 - PMTred rel		6.130	0.848	2.934	1.092
h008 - PMFred abs		8.655	1.197	2.260	0.842
h009 - MASH Marinova		3.851	0.532	2.062	0.768
h010 - Climatol		5.930	0.820	2.962	1.103
h011 - MASH main		3.288	0.455	1.699	0.632
h013 - PRODIGE trendy		3.948	0.546	2.525	0.940
h018 - AnClim main		5.744	0.794	2.552	0.950
h021 - PRODIGE monthly		3.277	0.453	2.040	0.760

Considering the performance of the automatic and semi-automatic versions of gsimcli, both provide similar results. For the Station's CRMSE and Improvement, the semi-automatic gsimcli was more efficient (in average, 2%) for network 5, and less harmful (in average, 11%) for network 9. Regarding the Network's CRMSE and Improvement, the automatic gsimcli provided better results than the semi-automatic gsimcli (in average, 35% in network 5 and 16% in network 9). These results seem to indicate that the automatic gsimcli increases the temporal consistency of the regional climate signal more than the semi-automatic version.

Table 10 - Performance metrics of the annual precipitation series from network 9 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Test #	Adjustment method	Station CRMSE	Station Improvement	Network CRMSE	Network Improvement
1	**Automatic gsimcli**	6.48	1.19	2.89	1.54
2	**Automatic gsimcli**	6.48	1.19	2.89	1.54
3	**Automatic gsimcli**	6.4	1.18	2.88	1.53
4	**Automatic gsimcli**	6.4	1.18	2.88	1.54
5	**Automatic gsimcli**	6.38	1.17	2.87	1.53
6	**Automatic gsimcli**	6.45	1.19	2.88	1.53
1	**Semi-automatic gsimcli**	5.92	1.09	3.42	1.82
2	**Semi-automatic gsimcli**	5.83	1.07	3.08	1.64
3	**Semi-automatic gsimcli**	5.62	1.03	3.44	1.83
4	**Semi-automatic gsimcli**	5.61	1.03	3.45	1.84
5	**Semi-automatic gsimcli**	5.59	1.03	3.44	1.83
6	**Semi-automatic gsimcli**	5.88	1.08	3.28	1.74
Original geostatistical approach		10.58	1.95	6.85	3.64
Inhomogeneous data		*5.433*	*1.0*	*1.880*	*1.0*
h002 - PRODIGE main		3.308	0.609	1.284	0.683
h006 - C3SNHT		3.794	0.698	1.146	0.609
h007 - PMTred rel		4.126	0.759	1.606	0.854
h008 - PMFred abs		5.380	0.990	1.653	0.879
h009 - MASH Marinova		3.484	0.641	1.188	0.632
h010 - Climatol		5.039	0.927	2.936	1.275
h011 - MASH main		3.083	0.567	0.920	0.490
h012 - SNHT DWD		4.009	0.738	1.654	0.880
h013 - PRODIGE trendy		3.308	0.609	1.284	0.683
h018 - AnClim main		4.217	0.776	2.107	1.121
h021 - PRODIGE monthly		2.981	0.549	1.185	0.630

In network 5 (Table 9), the smallest Network metrics were obtained for the homogenisation Test #6 with both the automatic (Network CRMSE = 2.86; Network Improvement = 1.07), and the semi-automatic (Network CRMSE = 3.78; Network Improvement = 1.41) versions of gsimcli. The efficiency of the semi-automatic gsimcli Test #6 was lower than the Climatol (h010) and AnClim main (h018) procedures by 17% and 21%, respectively, in terms of Station CRMSE. However, all automatic versions of gsimcli were more efficient (at least 2%) than the Climatol (h010) in terms of Network CRMSE. The efficiency of the automatic gsimcli Test #6 was lower than the C3SNHT (h006), AnClim main (h018) and PRODIGE main (h002) procedures by 11%, 12% and 13%, respectively, in terms of Network CRMSE. Considering the results of network 9 (Table 10), the automatic gsimcli homogenisation Test #5 was the less harmful (Network CRMSE

= 2.87; Network Improvement = 1.53), whereas using the semi-automatic version the "best" homogenisation Test was #2 (Network CRMSE = 3.08; Network Improvement = 1.64). These results indicate that using the local radius parameter (*r*) with values greater than 1 does not conclusively increase the gsimcli's efficiency. However, using larger grid cells generally improves the gsimcli method efficiency and decreases the processing time, since the size of the simulation grid cells has a significant impact in the computational effort. These results are consistent with a preliminary sensitivity analysis of the gsimcli's parameters that was undertaken using monthly precipitation data from the benchmark's network 16 (Ribeiro et al., 2015a), which is located in Austria and comprises 15 stations.

4.4.2 Monthly precipitation series

The monthly series were homogenised with the automatic gsimcli method as described in Section 4.3.2. Even though the performance metrics of the homogenisation exercises provided similar values, the best results were obtained in Tests #8 and #9 (Table 11), which used a local radius parameter greater than zero. In the previous homogenisation exercises, using a local radius parameter (r) equal to zero provided similar results to the Tests #8 and #9. These results might be explained by the fact that the correction parameter was the percentile of 0.975, whereas the original geostatistical approach used the mean as the correction parameter in the homogenisation of the annual series. This suggests the high importance of the correction parameter in the overall homogenisation efficiency.

The efficiency of the automatic gsimcli was higher than the C3SNHT (h006), AnClim main (h018) and Climatol (h010) procedures by 24%, 19% and 8%, respectively, in terms of Station CRMSE. However, it underperformed the PRODIGE monthly (h021) and the MASH Marinova (h009) procedures by 9% and 22%, respectively. It is noticeable that, in comparison with other procedures, the efficiency of gsimcli in the homogenisation of monthly series is higher when compared to the annual series.

Table 11 - Performance metrics of the monthly precipitation series from networks 5 and 9 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Method	**Station CRMSE**	**Station Improvement**	**Network CRMSE**	**Network Improvement**
gsimcli Test #7	10.38	1.02	4.20	1.13
gsimcli Test #8	10.34	1.02	4.17	1.12
gsimcli Test #9	10.34	1.02	4.18	1.12
Inhomogeneous data	*10.142*	*1.0*	*3.713*	*1.0*
h002 - PRODIGE main	7.665	0.762	3.454	0.932
h006 - C3SNHT	13.634	1.344	6.095	1.637
h007 - PMTred rel	9.449	0.930	3.754	1.009
h008 - PMFred abs	10.930	1.072	3.503	0.944
h009 - MASH Marinova	8.499	0.842	3.822	1.026
h010 - Climatol	11.224	1.120	4.804	1.299
h011 - MASH main	8.059	0.796	3.244	0.872
h013 - PRODIGE trendy	7.665	0.762	3.454	0.932
h018 - AnClim main	12.750	1.266	4.071	1.100
h021 - PRODIGE monthly	9.522	0.941	3.709	0.994

4.5 Concluding remarks

In the original geostatistical approach (Costa and Soares, 2009a; Costa et al., 2008a), the detection and correction stages of the homogenisation process were automatic procedures based on individual pieces of data. The proposed gsimcli algorithm includes a new parameter (local radius) that aims to provide more local information to the calculation of the local pdf in order to reproduce the climatic signal of that location more realistically. Moreover, the gsimcli method may include another stage that aims at further adjusting the candidate time series by examining the characteristics of segments of data (semi-automatic version). Both automatic and semi-automatic versions of the gsimcli method proved to be more efficient in the homogenisation of the benchmark's precipitation series than the geostatistical approach proposed by Costa and Soares (2009a) and Costa et al. (2008a).

The semi-automatic version of gsimcli uses comparison series that can be statistically tested in order to proceed with further inhomogeneities detection and adjustments. In this study, both gsimcli versions provided similar results in the homogenisation of annual precipitation series. We used the Buishand's test in the semi-automatic gsimcli, but the application of other techniques should be investigated.

Even though the geostatistical homogenisation made the data slightly more inhomogeneous in many experiments, the gsimcli approach outperformed a few procedures in the homogenisation of monthly precipitation data (Table 11), and Climatol in the homogenisation of monthly temperature series (as shown in Table S.3 of Appendix D). It is also important to point out that the benchmark's networks are relatively small, and that the gsimcli method is more appropriate for larger networks. Ribeiro et al. (2016b) tested the gsimcli method with a real data set of 66 monitoring stations from Portugal (0.0015 stations/km^2 in a simulation grid with 1 km^2 cells), whereas networks 5 and 9 have nine and five stations, respectively (0.0009 and 0.0005 stations/km^2 in the simulation grids with 1 km^2 cells, respectively).

Geostatistical techniques are suitable for variables that exhibit spatial correlation, which is modelled by the variogram. A higher number of observations that are spatially well distributed allows for a more accurate estimation of the variogram, thus improving the quality of the kriging predictions. A major limitation of this study was the reduced number of points available to estimate the variogram models. The modelling was particularly difficult for the shorter lag distances, which tend to have very few pairs of points. This is an important weakness of gsimcli, since the variogram's behaviour near the origin is the most important to characterise. Accordingly, further research with larger networks should be pursued.

Another direction for future research is the application of Direct Sequential Cosimulation (coDSS; Soares, 2001), which is an extension of the DSS algorithm that allows incorporating covariates such as elevation. Such extension of the gsimcli procedure could be suited for homogenising climatological networks from mountainous regions. However, the variography analysis would be even more challenging, because the coDSS algorithm requires a linear model of coregionalisation (i.e., modelling the spatial correlation structure through the simple and cross variograms). Another potential drawback is that the computational effort would highly increase.

The proposed approach is a valuable contribution to this research field, particularly the new methods' capability for filling missing values, and irregularities filtering. However, data corresponding to missing data or outliers were not taken into account in the computation of the performance metrics.

5 Conclusion

This research aimed at evaluating the efficiency of the geostatistical simulation approach (Costa et al. 2008a), and also envisioned the investigation of an extension of this procedure, named gsimcli, which should better estimate the climatic signal of the surrounding area of the candidate station's location. The efficiency of both the original geostatistical simulation approach and of gsimcli was evaluated using annual and monthly precipitation series of the benchmark data set from COST Action ES0601 “HOME” (Ribeiro et al., 2015a, 2015b, 2016c, 2016d).

The formulation of the gsimcli method follows closely the original geostatistical approach proposed by Costa et al. (2008a) to detect inhomogeneities in climate time series. One of the changes introduced on the original procedure aims to provide more local information to the calculation of the probability distribution function at the candidate station’s location (local pdf) in order to better estimate the climatic signal of its surrounding area. Furthermore, a different approach to adjust for sudden shifts in the inhomogeneous series has been proposed for the gsimcli method.

The geostatistical simulation approach uses the direct sequential simulation (DSS) algorithm to generate a set of equally probable and independent realisations and to estimate the local pdf. When an irregularity is detected, the corresponding value is replaced by a statistical value (correction parameter) derived from the estimated local pdf. The local radius parameter of the gsimcli procedure allows enhancing the local pdf estimation by including values simulated within a neighbourhood of the candidate station’s location. The detection and correction stages of the homogenisation process are automatic procedures based on individual observations of the climatic time series. The semi-automatic version of gsimcli includes another stage that takes advantage of a comparison series to examine the characteristics of segments of data using traditional homogenisation techniques.

Both automatic and semi-automatic versions of the gsimcli method proved to be more efficient in the homogenisation of the benchmark’s precipitation series than the original geostatistical simulation approach (Ribeiro et al., 2015a, 2016c, 2016d). Results also show that gsimcli outperformed a few well-established procedures in the homogenisation of monthly precipitation series (Section 4; Ribeiro et al., 2016d).

According to the specific objectives listed in Section 1.5, detailed conclusions are as follows.

The literature review (Section 2; Ribeiro et al., 2016a) emphasised the importance of the development of homogenisation methods to ensure the accuracy of climate records. It was also highlighted the reduced number of homogenisation methods dedicated to variables with high variability, such as precipitation, and dealing with high temporal resolution data sets (monthly/sub-monthly). Relevant contributions of the literature review are the comprehensive summary and description of the available homogenisation methods and a summary of their applications, which may help climatologists and other researchers to select adequate method(s) for their particular needs. Another important contribution is the discussion of the advantages and disadvantages of the homogenisation methods, which depicts lessons learned regarding good homogenisation practices.

Previous works (Costa et al., 2008a; Costa and Soares, 2009a) required a lot of time and interaction from its users. The gsimcli software (http://iled.github.io/gsimcli) allowed homogenising the climate data series in an intuitive and straightforward way. The computational performance has been an important factor in the design and implementation of the algorithms, both in processing time and required system memory (Caineta et al., 2015a, 2015b).

The extension of the study presented by Costa et al. (2008a) (Section 3; Ribeiro et al., 2016b) comprised the detection of irregularities in a real precipitation data set (66 monitoring stations) located in the south of Portugal. The analysed climate variable was the annual number of wet days. By comparing the detection skills of the geostatistical simulation approach with other homogenisation methods, it was possible to conclude that all methods indicate the presence of inhomogeneities around the same time periods. The geostatistical approach detected the existence of irregularities in a larger sequential interval, which can be an indicator that it is able to detect trends. Some of the analysed parameters were the number of simulations and the number of nodes included in the simulations. A higher number of simulations lead to better detection results, because the empirical local distribution function tends to be less irregular. The increase of the number of nodes included in the simulations did not bring enough benefits to justify the increasing computing time.

The original geostatistical approach and the gsimcli method were tested against artificial annual and monthly precipitation data provided by the HOME project (Section 4; Ribeiro et al., 2016d). The sensitivity analysis of the modelling parameters showed a high influence of the correction method in the efficiency of the homogenisation. The original geostatistical approach used the mean as the correction parameter in the homogenisation of the annual series. However, the best results of the performance metrics were obtained with a correction parameter equal to the 0.975 percentile (Ribeiro et al., 2015a, 2015b, 2016d). It can also be concluded that increasing the size of the grid cell accelerates the simulation process, without decreasing the quality of homogenisation significantly (Ribeiro et al., 2015a, 2016d). The local radius parameter of the gsimcli algorithm brings the local characteristics of the climate variable into the calculation of the local pdf. The advantages of such innovation are expressed in the improvement of the performance metrics, as demonstrated by Ribeiro et al. (2015a, 2016c, 2016d).

The automatic and semi-automatic (with Buishand-test) versions of gsimcli provided similar results in the homogenisation of annual precipitation data. Even though the geostatistical homogenisation made the data slightly more inhomogeneous in many experiments, the gsimcli approach outperformed a few procedures in the homogenisation of monthly precipitation data. The capability of gsimcli to fill in missing values and to filter irregularities is an advantage when compared to other methods.

5.1 Limitations

Similarly to any other geostatistical approach, the homogenisation of climate data with gsimcli assumes that the study variable is spatially autocorrelated, which is assessed by the study of its variogram. The higher the number of observations, the more accurate the variogram estimation, and so the kriging predictions computed based on it. The reduced number of stations available in the benchmark networks proved to be the major limitation of the study.

5.2 Future research

Based on the above-mentioned conclusions, some recommendations for future work are made. The performance of gsimcli in the homogenisation of large network data sets should be further assessed, since a high number of observations will improve the estimation of the variogram model.

Testing the efficiency of gsimcli with other benchmark data sets (e.g., the International Surface Temperature Initiative's benchmark), and for other climatic regions of the World, should also be pursued. The International Surface Temperature Initiative's project will be the first global benchmarking study and it will enable the assessment of homogenisation methods' performance in quite diverse climatic areas (Willett et al., 2014).

Domonkos (2013b) argues that the optimal homogenisation method should be a combination of the best segments of homogenisation methods, such as the best detection part, the best correction part, etc. Further research is also needed to find the optimal way of spatial comparison (Domonkos, 2011a). In this context, the evaluation of the semi-automatic gsimcli procedure using other techniques (alternative to the Buishand-test) is encouraged.

The cost-benefit analysis of the inclusion of covariates in the homogenisation is also suggested, in regions where statistically significant correlation is observed.

In summary, the following research questions should be investigated in future works:

- Is gsimcli more efficient than other state-of-the-art methods in the homogenisation of dense monitoring networks?

- Can the gsimcli algorithm be further improved by directly incorporating data from neighbouring stations in the estimation of the local pdf?
- How does gsimcli perform for other climate variables?
- How does gsimcli perform in other regions of the World?
- Can the gsimcli method be extended, or incorporated in other homogenisation procedures, to take advantage of the comparison series that are derived from the local pdfs of the candidate station?
- Is it worthwhile to increase the algorithm complexity by using multivariate geostatistical simulation?

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidelines on Climate Metadata and Homogenization. WMO/TD No. 1186, WCDMP No. 53, Llansó P (ed.). World Meteorological Organization, Geneva, Switzerland, 55 pp.

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology*, **6**, 661–675. doi: 10.1002/joc.3370060607.

Alexandersson H, Moberg A. 1997. Homogenization of Swedish Temperature Data. Part I: Homogeneity Test for Linear Trends. *International Journal of Climatology* **17**(1): 25–34. doi: 10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J.

Allen RJ, De Gaetano AT. 2000. A method to adjust long-term temperature extreme series for non climatic inhomogeneities. *Journal of Climate* **13**(2): 3680–3695. doi: 10.1175/1520-0442(2000)013<3680:AMTALT>2.0.CO;2.

Allen R, Pereira L, Raes D, Smith M. 1998. Statistical analysis of weather data sets, Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements, FAO – Food and Agriculture Organization of the United Nations (eds.). *FAO Irrigation and drainage paper 56*, Annex IV, Rome.

Auer I, Bohm R, Jurkovic A, Orlik A, Potzmann R, Schoner W, Ungersbock M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, Moisselin JM, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajic-Capka M, Zaninovic K, Majstorovic Z, Szalai S, Szentimrey T, Mercalli L. 2005. A new instrumental precipitation dataset for the greater alpine region for the period 1800-2002. *International Journal of Climatology* **25**(2): 139–166. doi: 10.1002/joc.1135.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X, Boulet G, Yagouti A. 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44**(2). doi: 10.1029/2006WR005615.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X. 2009. Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change-point models. *Water Resources Research* **45**: W08410. doi: 10.1029/2008WR007501.

Beaulieu C, Ouarda TBMJ, Seidou O. 2010. A Bayesian normal homogeneity test for the detection of artificial discontinuities in climatic series. *International Journal of Climatology* **30**(15): 2342–2357. doi: 10.1002/joc.2056.

Begert M, Schlegel T, Kirchhofer W. 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology* **25**(1): 65–80. doi: 10.1002/joc.1118.

Bourennane H, King D, Couturier A, Nicoullaud B, Mary B, Richard G. 2007. Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. *Ecological Modelling* **205**: 323–335. doi: 10.1016/j.ecolmodel.2007.02.034.

Brunet M, Jones P. 2011. Data rescue initiatives: bringing historical climate data into the 21st century. *Climate Research* **47**(1): 29–40. doi: 10.3354/cr00960.

Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, Moberg A, Lister D, Walther A, Lopes D, Almarza C. 2006. The Development of a New Dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850-2003). *International Journal of Climatology* **26**: 1777–1802. doi: 10.1002/joc.

Brunetti M, Maugeri M, Monti F, Nanni T. 2006. Temperature and precipitation variability in Italy in the last two centuries from homogenized instrumental time series. *International Journal of Climatology* **26:** 345–381. doi: 10.1002/joc.1251.

Brunetti M, Caloiero T, Coscarelli R, Gullà G, Nanni T, Simolo C. 2012. Precipitation variability and change in the Calabria region (Italy) from a high resolution daily dataset. *International Journal of Climatology* **32**(1): 57–73. doi: 10.1002/joc.2233.

Buishand T. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**(1): 11–27. doi: 10.1016/0022-1694(82)90066-X.

Buishand TA, De Martino G, Spreeuw JN, Brandsma T. 2013. Homogeneity of precipitation series in the Netherlands and their trends in the past century. *International Journal of Climatology*, **33**(4): 815–833. doi: 10.1002/joc.3471.

Caers J. 2000. Adding local accuracy to Direct Sequential Simulation. *Mathematical Geology* **32**(7): 815–850. doi: 10.1023/A:1007596423578.

Caineta J, Ribeiro S, Henriques R., Costa AC. 2015a. “A Package for the homogenisation of climate data using geostatistical simulation”. In: *GEOProcessing 2015: The Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services*, IARIA (ed.), Lisbon, Portugal, 22-27 February 2015, pp. 123-126.

Caineta J, Ribeiro S, Soares A, Costa AC. 2015b. “Workflow for the homogenisation of climate data using geostatistical simulation”. In: *Informatics, Geoinformatics and Remote Sensing,* *International Multidisciplinary Scientific GeoConference-SGEM*, Albena, Bulgaria, Jun 18-24, 2015, Vol. 1, 921-929. DOI: 10.5593/SGEM2015/B21/S8.118

Caussinus H, Mestre O. 1996. “New mathematical tools and methodologies for relative homogeneity testing”. In: *Proceedings of the First Seminar for Homogenization of Surface Climatological Data*, Hungarian Meteorological Service (ed.). Budapest, Hungary, 63–82.

Caussinus H, Lyazrhi F. 1997. Choosing a Linear Model with a Random Number of Change-Points and Outliers. *Annals of the Institute of Statistical Mathematics* **49**(4): 761–775. doi: 10.1023/A:1003230713770.

Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Applied Statistics* **53:** 405–425. doi: 10.1111/j.1467-9876.2004.05155.x.

Costa AC, Soares A. 2006. “Identification of inhomogeneities in precipitation time series using SUR models and the Ellipse test”. In: *Proceedings of Accuracy 2006 - 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Caetano M, Painho M (eds.), Instituto Geográfico Português, pp. 419–428.

Costa AC, Soares A. 2009a. Homogenization of climate data: Review and new perspectives using geostatistics. *Mathematical Geosciences* **41**(3): 291–305. doi: 10.1007/s11004-008-9203-3.

Costa AC, Soares A. 2009b. Trends in extreme precipitation indices derived from a daily rainfall database for the South of Portugal. *International Journal of Climatology* **9**(13): 1956–1975. doi: 10.1002/joc.1834

Costa AC, Soares A. 2012. Local spatiotemporal dynamics of a simple aridity index in a region susceptible to desertification. *Journal of Arid Environments* **87**: 8–18. doi: 10.1016/j.jaridenv.2012.05.005.

Costa AC, Negreiros J, Soares A. 2008a. Identification of inhomogeneities in precipitation time series using stochastic simulation. In: *geoENV VI – Geostatistics for Environmental Applications*, Soares A, Pereira MJ, Dimitrakopoulos R (eds.). Springer, pp. 275–282. doi: 10.1007/978-1-4020-6448-7_23.

Costa AC, Durão R, Pereira MJ, Soares A. 2008b. Using stochastic space-time models to map extreme precipitation in southern Portugal. *Natural Hazards and Earth System Sciences* **8**(4): 763–773. doi: 10.5194/nhess-8-763-2008.

Costa AC, Soares A, Henriques R, Ribeiro S, Caineta J. 2015. Final Scientific Report of the Project. GSIMCLI – Geostatistical simulation with local distributions for the homogenization and interpolation of climate data, PTDC/GEO-MET/4026/2012. NOVA Information Management School, 30 November 2015, 72 pp.

Craddock JM. 1979. Methods of comparing annual rainfall records for climatic purposes. *Weather* **34:** 332–346. doi: 10.1002/j.1477-8696.1979.tb03465.x.

de Lima MIP, Santo FE, Ramos AM, de Lima JL. 2013. Recent changes in daily precipitation and surface air temperature extremes in mainland Portugal, in the period 1941–2007. *Atmospheric Research* **127**: 195–209. doi: 10.1016/j.atmosres.2012.10.001.

Deutsch CV, Journel AG. 1998. *GSLIB: Geostatistical Software Library*. Oxford University Press.

Domonkos P. 2006. Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation. *Időjárás* **110**(1): 63–87.

Domonkos P. 2011a. Homogenising time series: beliefs, dogmas and facts. *Advances in Science and Research* **6**: 167–172. doi: 10.5194/asr-6-167-2011.

Domonkos P. 2011b. Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods. *Theoretical and Applied Climatology* **105**(3-4): 455–467. doi: 10.1007/s00704-011-0399-7.

Domonkos P. 2011c. Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). *International Journal of Geosciences* **02**(03): 293–309. doi: 10.4236/ijg.2011.23032.

Domonkos P. 2011d. "ACMANT: Why is it efficient?". In: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.), World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 33–44.

Domonkos P. 2013a. Measuring performances of homogenization methods. *Időjárás* **117**(1): 91–112.

Domonkos P. 2013b. Efficiencies of inhomogeneity-detection algorithms: comparison of different detection methods and efficiency measures. *Journal of Climatology*, Article ID 390945: 15 pages. doi: 10.1155/2013/390945.

Domonkos P. 2015. Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology* **122**(1): 303–314. doi: 10.1007/s00704-014-1298-5.

Domonkos P, Štěpánek P. 2009. Statistical characteristics of detectable inhomogeneities in observed meteorological time series. *Studia Geophysica et Geodaetica*, **53**(2): 239–260. doi: 10.1007/s11200-009-0015-9.

Domonkos P, Coll J. 2016. Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. *International Journal of Climatology*. doi: 10.1002/joc.4822, in press.

Domonkos P, Poza R, Efthymiadis D. 2011a. Newest developments of ACMANT. *Advances in Science Research* **6:** 7–11. doi: 10.5194/asr-6-7-2011.

Domonkos P, Venema V, Mestre O. 2011b. “Efficiencies of homogenisation methods: our present knowledge and its limitation”. In: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.). World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 11–24.

Domonkos P, Venema V, Auer I, Mestre O, Brunetti M. 2012. The historical pathway towards more accurate homogenisation. *Advances in Science Research* **8**(1): 45–52. doi: 10.5194/asr-8-45-2012.

Ducré-Robitaille J-F, Vincent LA, Boulet G. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**(9): 1087–1101. doi: 10.1002/joc.924.

Durão RM, Pereira MJ, Costa AC, Delgado J, del Barrio G, Soares A. 2010. Spatial-temporal dynamics of precipitation extremes in southern Portugal: a geostatistical assessment study. *International Journal of Climatology* **30**(10): 1526–1537. doi: 10.1002/joc.1999.

Easterling DR, Peterson TC. 1995. A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369–377. doi: 10.1002/joc.3370150403.

Eccel E, Cau P, Ranzi R. 2012. Data reconstruction and homogenization for reducing uncertainties in high-resolution climate analysis in Alpine regions. *Theoretical and Applied Climatology* **110**(3): 345–358. doi: 10.1007/s00704-012-0624-z.

Emery X. 2004. Properties and limitations of sequential indicator simulation. *Stochastic Environmental Research and Risk Assessment* **18**(6): 414–424. doi: 10.1007/s00477-004-0213-5.

Franco C, Soares A, Delgado J. 2006. Geostatistical modelling of heavy metal contamination in the topsoil of Guadiamar river margins (S Spain) using a stochastic simulation technique. *Geoderma*, **136**: 852–864. doi: 10.1016/j.geoderma.2006.06.012.

Gaffen DJ, Sargent MA, Habermann RE, Lanzante JR. 2000. Sensitivity of Tropospheric and Stratospheric Temperature Trends to Radiosonde Data Quality. *Journal of Climate* **13**: 1776–1796. doi: 10.1175/1520-0442(2000)013<1776:SOTAST>2.0.CO;2.

Gallagher C, Lund R, Robbins M. 2012. Changepoint detection in daily precipitation data. *Environmetrics*, **23**(5): 407–419. doi: 10.1002/env.2146.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series, Oxford University Press.

Guijarro J. 2006. “Homogenization of a dense thermo-pluviometric monthly database in the Balearic Islands using the free contributed R package “Climatol””. In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.). World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1493, pp. 25–36.

Guijarro J. 2011. “Influence of network density on homogenization performance”. In: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.). World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 11–18.

Guijarro J. 2013. Climatological series shift test comparison on running windows. *Időjárás* **117**(1): 35–45.

Hannart A, Naveau P. 2009. Bayesian multiple change points and segmentation: Application to homogenization of climatic series, *Water Resources Research* **45** (10): W10444. doi: 10.1029/2008WR007689.

HOME, 2006. Memorandum of Understanding for the implementation of a European Concerted Research Action designated as COST Action ES0601: Advances in Homogenization Methods of Climate Series: An Integrated Approach (HOME), COST Secretariat, 2006.

HOME, 2011. *Monitoring Progress Report 03/05/2007 – 01/06/2011*. HOME – Advances in Homogenisation Methods of Climate Series: an Integrated Approach (COST Action ES0601), Chair of the Management Committee of the Action, submitted to the relevant Domain Committee.

Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org (accessed November 2015).

Journel AG. 1994. “Modeling uncertainty: some conceptual thoughts”. In: *Geostatistics for the Next Century*, Dimitrakopoulos R (ed.), Kluwer Academic Pub., Dordrecht, The Netherlands, pp. 30–43. doi: 10.1007/978-94-011-0824-9_5

Kendall M. 1975. *Rank correlation methods*. Charles Griffin, London.

Klein Tank AMG, Zwiers FW, Zhang X. 2009. *Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation*. Climate data and monitoring, WCDMP No.72, WMO-TD No.1500, World Meteorological Organization, Geneva, Switzerland, pp. 1–56.

Kohler M. 1949. Double-mass analysis for testing the consistency of records and for making adjustments. *Bulletin of the American Meteorological Society* **30**: 188–189.

Kruskal W. 1952. A non-parametric test for the several sample problem. *The Annals of Mathematical Statistics* **23**: 525–540.

Kruskal W, Wallis W. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**: 583– 621.

Lakatos M, Szentimrey T, Bihari Z, Szalai S. 2008. Homogenization of daily data series for extreme climate indeces calculation". In: *Proceedings of the Sixth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.), World Meteorological Organization, Budapest, Hungary, WCDMP No. 76, WMO-TD No. 1576, pp. 100–109.

Li Q, Zhang H, Liu X, Huang J. 2004. Urban heat island effect on annual mean temperature during the last 50 years in China. *Theoretical and Applied Climatology* **79**(3-4): 165–174. doi: 10.1007/s00704-004-0065-4.

Li-Juan C, Zhong-Wei Y. 2012. Progress in research on homogenization of climate data. *Advances in Climate Change Research* **3**(2): 59–67. doi: 10.3724/SP.J.1248.2012.00059.

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

Lindau R, Venema V. 2016. The uncertainty of break positions detected by homogenization algorithms in climate records. *International Journal of Climatology* **36**: 576–589. doi: 10.1002/joc.4366.

Mann H. 1945. Non-parametric tests against trend. *Econometrica* **13**(3): 245–259. doi: 10.2307/1907187.

Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* **18**: 50–60.

Maronna R, Yohai VJ. 1978. A bivariate test for the detection of a systematic change in mean. *Journal of the American Statistical Association* **73**: 640–645. doi: 10.1080/01621459.1978.10480070.

Maugeri M, Brunetti M, Monti F, Nanni T. 2004. Sea-level pressure variability in the Po Plain (1765–2000) from homogenized daily secular records. *International Journal of Climatology* **24**(4): 437–455. doi: 10.1002/joc.991.

Menne MJ, Williams Jr. CN. 2009. Homogenization of temperature series via pairwise comparisons. *Journal of Climate* **22**(7): 1700–1717. doi: 10.1175/2008JCLI2263.1.

Menne MJ, Williams Jr. CN, Vose RS. 2009. The U. S. historical climatology network monthly temperature data, version 2. *Bulletin of American Meteorology Society* **90**: 993–1007. doi: 10.1175/2008BAMS2613.1.

Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S. 2011. SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology* **50**(11): 2343–2358. doi: 10.1175/2011JAMC2641.1.

Mestre O, Domonkos P, Picard F, Auer I, Robin S, Lebarbier E, Bohm R, Aguilar E, Guijarro J, Vertachnik G, Klancar M, Dubuisson B, Štěpánek P. 2013. HOMER: a homogenization software – methods and applications. *Időjárás* **117**(1): 47–67.

Moberg A, Alexandersson H. 1997. Homogenization of Swedish Temperature Data. Part II: Homogenized Gridded Air Temperature Compared with a Subset of Global Gridded Air Temperature Since 1861. *International Journal of Climatology* **17:** 35–54. doi: 10.1002/(SICI)1097-0088(199701)17:1<35::AID-JOC104>3.0.CO;2-F.

Mourato S, Moreira M, Corte-Real J. 2010. Interannual variability of precipitation distribution patterns in Southern Portugal. *International Journal of Climatology* **30**(12): 1784–1794. doi: 10.1002/joc.2021.

Nemec J, Gruber C, Chimani B, Auer I. 2013. Trends in extreme temperature indices in Austria based on a new homogenised dataset. *International Journal of Climatology* **33**(6): 1538–1550. doi: 10.1002/joc.3532.

Oliveira AR, Branquinho C, Pereira M, Soares A. 2013. Stochastic Simulation Model for the Spatial Characterization of Lung Cancer Mortality Risk and Study of Environmental Factors. *Mathematical Geosciences* **45**(4): 437–452. doi: 10.1007/s11004-013-9443-8.

Pandžić K, Likso T. 2010. Homogeneity of average annual air temperature time series for Croatia. *International Journal of Climatology* **30**: 1215–1225. doi: 10.1002/joc.1022.

Pereira LS, Louro V, Rosário L, Almeida A. 2006. “Desertification, territory and people, a holistic approach in the Portuguese context”. In: *Desertification in the Mediterranean Region: a Security Issue*. Kepner WG, Rubio JL, Mouat DA, Pedrazzini F (eds.), NATO Security Through Science Series, Vol. 3, pp. 269–289. doi: 10.1007/1-4020-3760-0_11.

Perreault L, Bernier J, Bobée B, Parent E. 2000. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology* **235**(3-4): 221–241. doi: 10.1016/S0022-1694(00)00270-5.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology* **18**(13): 1493–1517. doi: 10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Pettitt AN. 1979. A non-parametric approach to the change-point detection. *Journal of Applied Statistics* **28**(2): 126–135. doi: 10.2307/2346729.

Plummer N, Lin Z, Torok S. 1995. Trends in the diurnal temperature range over Australia since 1951. *Atmospheric Research* **37**(1-3): 79–86. doi: 10.1016/0169-8095(94)00070-T.

Potter, K., 1981. Illustration of a new test for detecting a shift in mean in precipitation series. *M. Weather Review* **109**: 2040–2045.

Puglisi A, Bartolini G, Orlandini S. 2010. Climatic variability in Tuscany : homogeneity methods of climatic data and analysis of impacts on grapevine and olive trees. *Italian Journal of Agrometeorology* **1/2010**: 45–52.

Reeves J, Chen J, Wang XL, Lund R, Lu QQ. 2007. A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology* **46**(6): 900–915. doi: 10.1175/JAM2493.1.

Ribeiro MC, Pinho P, Llop E, Branquinho C, Soares A, Pereira MJ. 2014. Associations between outdoor air quality and birth weight: a geostatistical sequential simulation approach in Coastal Alentejo, Portugal. *Stochastic Environmental Research and Risk Assessment* **28**(3): 527–540. doi: 10.1007/s00477-013-0770-6.

Ribeiro S, Caineta J, Costa AC, Soares A. 2015a. Establishment of detection and correction parameters for a geostatistical homogenisation approach. *Procedia Environmental Sciences* **27**: 83–88. doi: 10.1016/j.proenv.2015.07.115.

Ribeiro S., Caineta J, Costa AC, Henriques R. 2015b. “Analysing the detection and correction parameters in the homogenisation of climate data series using gsimcli”. In: *The 18th AGILE International Conference on Geographic Information Science*, F. Bacao, M. Y. Santos, M. Painho (eds), Lisbon, Portugal, 9-12 June 2015.

Ribeiro S, Caineta J, Costa AC. 2016a. Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth* **94**: 167–179. doi: 10.1016/j.pce.2015.08.007.

Ribeiro S, Caineta J, Costa AC, Henriques R, Soares A, 2016b. Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation. *Atmospheric Research* **171**: 147–158. doi: 10.1016/j.atmosres.2015.11.014.

Ribeiro S, Caineta J, Costa AC. 2016c. “Assessing the performance of the gsimcli homogenisation method with precipitation monthly data from the COST-HOME benchmark”. In: *Geostatistics Valencia 2016*, J.J. Gómez-Hernández, J. Rodrigo-Ilarri, M.E. Rodrigo-Clavero, E. Cassiraga, J.A. Vargas-Guzmán (eds.), Quantitative Geology and Geostatistics, Springer International Publishing, Vol. 19, Ch. 64, in press. doi: 10.1007/978-3-319-46819-8.

Ribeiro S, Caineta J, Costa AC, Henriques R, 2016d. gsimcli: a geostatistical procedure for the homogenisation of climatic time series. *International Journal of Climatology*, in press. doi: 10.1002/joc.4929.

Robertson RK, Mueller UA, Bloom LM. 2006. Direct sequential simulation with histogram reproduction: A comparison of algorithms. *Computers & Geosciences* **32**(3): 382–395. doi: 10.1016/j.cageo.2005.07.002.

Ruggieri E. 2013. A Bayesian approach to detecting change points in climatic records. *International Journal of Climatology* **33**(2): 520–528. doi: 10.1002/joc.3447.

Rustemeier E, Kapala A, Venema V, Simmer C. 2011. “Detection and correction of breakpoints in long-term German precipitation series”. In: *Proceedings of the Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.), World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 62–71.

Sahin S, Cigizoglu HK. 2010. Homogeneity analysis of Turkish meteorological data set. *Hydrological Processes* **24**(8): 981–992. doi: 10.1002/hyp.7534.

Santos M, Fragoso M. 2013. Precipitation variability in Northern Portugal: Data homogeneity assessment and trends in extreme precipitation indices. *Atmospheric Research* **131:** 34–45. doi: 10.1016/j.atmosres.2013.04.008.

Seidou O, Ouarda TBMJ. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43:** W07404, doi: 10.1029/2006WR005021.

Seidou O, Asselin JJ, Ouarda TBMJ. 2007. Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. *Water Resources Research* **43**: W08401, doi: 10.1029/2005WR004835.

Sevruk B, Ondrás M, Chvíla B. 2009. The WMO precipitation measurement intercomparisons. *Atmospheric Research* **92**(3): 376–380. doi: 10.1016/j.atmosres.2009.01.016.

Soares A. 2001. Direct sequential simulation and cosimulation. *Mathematical Geology* **33**(8): 911–926. doi: 10.1023/A:1012246006212.

Solow A. 1987. Testing for climate change: An application of the two-phase regression model. *Journal of Applied Meteorology and Climatology* **26**: 1401–1405. doi: 10.1175/1520-0450(1987)026<1401:TFCCAA>2.0.CO;2.

Štěpánek P. 2008a. AnClim – Software for Time Series Analysis. Dept. of Geography, Fac. of Natural Sciences, MU, Brno. http://www.climahom.eu/AnClim.html (accessed October 2013).

Štěpánek P. 2008b. ProClimDB – Software for Processing Climatological Datasets. CHMI, Regional Office Brno. http://www.climahom.eu/ProcData.html (accessed October 2013).

Štěpánek P, Řezníčková L, Brázdil R. 2006. “Homogenization of daily air pressure and temperature series for Brno (Czech Republic) in the period 1848–2005”. In: *Proceedings of the fifth seminar for homogenization and quality control in climatological databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.), World Meteorological Organization, Budapest, Hungary 2006, WMO-TD No. 1493, pp. 106-121.

Štěpánek P, Zahradníček P, Skalák P. 2009. Data quality control and homogenization of the air temperature and precipitation series in the Czech Republic in the period 1961–2007, *Advances in Science and Research* **3**: 23–26. doi: 10.5194/asr-3-23-2009.

Szentimrey T. 1999. “Multiple analysis of series for homogenization (MASH)”. In: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, WMO, WCDMP-No. 41, WMO-TD No. 1962, pp. 27–46.

Szentimrey T. 2006a. An overview on the main methodological questions of homogenization. In *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.), World Meteorological Organization, Budapest, Hungary 2006, WCDMP-No. 71, WMO-TD No. 1493, pp. 1–6.

Szentimrey T. 2006b. “Development of MASH homogenization procedure for daily data”. In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.). World Meteorological Organization, Budapest, Hungary 2006, WCDMP-No. 71, WMO-TD No. 1493, pp. 123–130.

Szentimrey T. 2007. Manual of homogenization software MASH v3.02. Hungarian Meteorological Service, 65 pp.

Szentimrey T. 2011. “Theoretical Aspects of Homogenization”. In: *Proceedings of the Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.), World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 2–10.

Talaee PH, Kouchakzadeh M, Shifteh Some’e B. 2014. Homogeneity analysis of precipitation series in Iran. *Theoretical and Applied Climatology* **118**(1-2): 297–305. doi: 10.1007/s00704-013-1074-y.

Tayanç M, Dalfes HN, Karaca M, Yenigu O. 1998. A comparative assessment of different methods for detecting inhomogeneities in Turkish temperature. *International Journal of Climatology* **18**: 561–578. doi: 10.1002/(SICI)1097-0088(199804)18:5<561::AID-JOC249>3.0.CO;2-Y.

Teegavarapu R, Chandramouli V. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology* **312**(1-4): 191–206. doi: 10.1016/j.jhydrol.2005.02.015.

Trewin B. 2010. Exposure, instrumentation, and observing practice effects on land temperature measurements. WIREs *Climate Change* **1**: 490–506. doi: 10.1002/wcc.46.

Trewin B. 2013. A daily homogenized temperature data set for Australia. *International Journal of Climatology* **33**(6): 1510–1529. doi: 10.1002/joc.3530.

Tuomenvirta H. 2001. Homogeneity adjustments of temperature and precipitation series? Finnish and Nordic data. *International Journal of Climatology* **21**(4): 495–506. doi: 10.1002/joc.616.

Venema VKC, 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 MJ, 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, Brandsma T. 2011. Description of the COST-HOME monthly benchmark dataset and the submitted homogenized contributions. HOME – Advances in Homogenisation Methods of Climate Series: an Integrated Approach (COST Action ES0601), Meteorological Institute, University of Bonn, Germany, 18 Juli 2011, 32 pp.

Venema V, Mestre O, Aguilar E, Auer I, Guijarro J, Domonkos P, Vertacnik G, Szentimrey T, Štěpánek P, Zahradníček P, Viarre J, Muller-Westermeier G, Lakatos M, Williams C, 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, Brandsma T. 2012. Benchmarking homogenization algorithms for monthly data. *Climate of the Past* **8**(1): 89–115. doi: 10.5194/cp-8-89-2012.

Venema VKC, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey P, Štěpánek P, Zahradníček P, Viarre J, Muller-Westermeier G, Lakatos M, Williams CN, Menne MJ, Lindau R, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquaotta F, Fratiannil S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P, Bradsma T, Willett K. 2013. Benchmarking homogenization algorithms for monthly data. In *AIP Conference Proceedings* **204**: pp. 1060–1065. doi: 10.1063/1.4819690.

Vincent LA. 1998. A Technique for the Identification of Inhomogeneities in Canadian Temperature Series. *Journal of Climate* **11**: 1094–1104. doi: 10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics* **12**(4): 367–395.

Wald A, Wolfowitz J. 1943. An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics* **14**(4): 378–388.

Wang XL. 2008. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *Journal of Applied Meteorology and Climatology* **47**(9): 2423–2444. doi: 10.1175/2008JAMC1741.1.

Wijngaard JB, Klein Tank AMG, Können GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**(6): 679–692. doi: 10.1002/joc.906.

Wilcoxon F. 1945. Individual comparison by ranking methods. *Biometrics Bulletin* **1**: 80–83.

Willett K, Williams C, Jolliffe IT, Lund R, Alexander LV, Brönnimann S, Vincent LA, Easterbrook S, Venema VKC, Berry D, Warren RE, Lopardo G, Auchmann R, Aguilar E, Menne MJ, Gallagher C, Hausfather Z, Thorarinsdottir T, Thorne PW. 2014. A framework for benchmarking of homogenisation algorithm performance on the global scale. *Geoscientific Instrumentation, Methods and Data Systems* **3**(2): 187-200. doi:10.5194/gi-3-187-2014.

World Meteorological Organization 2010. *Guide to Climatological Practices*. WMO No. 100, third edition, World Meteorological Organization, Geneva, Switzerland.

Yamamoto JK. 2005. Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology* **37**(1): 69–94. doi: 10.1007/s11004-005-8748-7.

Yan ZW, Jones PD. 2008. Detecting inhomogeneity in daily climate series using wavelet analysis. *Advances in Atmospheric Sciences* **25**(2): 157–163. doi: 10.1007/s00376-008-0157-7.

Yozgatligil C, Yazici C. 2016. Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology* **36**(1): 62–81. doi: 10.1002/joc.4329.

Yue S, Wang C. 2004. The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management* **18**(3): 201–218. doi: 10.1023/B:WARM.0000043140.61082.60.

Appendix A

Table A.1 - Characteristics of the homogenisation methods

Method	Type	Temporal resolution	Breakpoints detection	Characteristics
Von Neumann ratio test (Von Neumann, 1941)	Non-parametric	Annual	Single-breakpoint No date identified	Used as absolute detection method or applied to composite reference series Qualifying test on homogeneity diagnosis
Wald-Wolfowitz runs test (Wald and Wolfowitz, 1943)	Non-parametric	Annual	Single-breakpoint No date identified	Used as absolute detection method or applied to composite reference series Requires supporting tests Qualifying test on homogeneity diagnosis
Pettitt's test (Pettitt, 1979)	Non-parametric	Annual	Single-breakpoint Date of break identified	Used as absolute detection method or applied to composite reference series More sensitive to breaks in the middle of the time series Qualifying test on homogeneity diagnosis
Mann-Kendall test (Mann, 1945; Kendall, 1975)	Non-parametric	Annual/ Monthly		Tests the significance of trends
Wilcoxon-Mann-Whitney test (Mann and Whitney, 1947; Wilcoxon, 1945)	Non-parametric	Annual	Single-breakpoint No date identified	Relative detection method Based on rank order breakpoint detection Qualifying test on homogeneity diagnosis
Kruskal-Wallis (Kruskal, 1952; Kruskal and Wallis, 1952)	Non-parametric	Any	Single-breakpoint No date identified	Relative detection method Qualifying test on homogeneity diagnosis

Method	Type	Temporal resolution	Breakpoints detection	Characteristics
Craddock's test (Craddock, 1979)	Classical	Any/ Annual	Single-breakpoint Date of break identified	Relative detection method with pairwise comparison Subjective method Requires a homogeneous reference series (or long enough homogeneous sub-periods) Included in HOCLIS (Austria) and THOMAS (Switzerland) homogenisation tools Correction estimation based on mean of multiple comparisons
Buishand range test (Buishand, 1982)	Classical	Annual	Single-breakpoint Date of break identified	Used as absolute detection method or applied to composite reference series More sensitive to breaks in the middle of the time series Qualifying test on homogeneity diagnosis
Double mass analysis (Kohler, 1949)	Classical	Annual	Single-breakpoint	Relative detection method Used for exploratory analysis Subjective method
Bivariate test (Potter, 1981)	Classical	Annual	Single-breakpoint	Relative detection method Based on maximum likelihood estimations Closely resembles the double mass analysis
Two-phase regression (Easterling and Peterson, 1995)	Regression method	Any	Single-breakpoint Date of break identified	Relative detection method Hierarchic method for identifying multiple breakpoints (cutting algorithm)

Method	**Type**	**Temporal resolution**	**Breakpoints detection**	**Characteristics**
				Stable high performance in detection skill
Multiple linear regression analysis (Vincent, 1998)	Regression method	Any	Single-breakpoint	Relative detection method Detection of gradual linear changes Objective detection method
Method of cumulative residuals (Allen et al., 1998)	Regression method	Any	Single- breakpoint	Relative detection method Used for exploratory analysis Qualifying test on homogeneity diagnosis
SNHT (Alexandersson, 1986)	Homogenisation procedure	Annual/ Monthly	Single-breakpoint	One of the most widely-used relative detection methods Usually applied to composite reference series
SNHT with trend (Alexandersson and Moberg, 1997)	Homogenisation procedure	Any	Single-breakpoint	Detection of gradual linear changes Comparison based on reference series Correction is estimated directly from comparison series
MASH (Szentimrey, 1999)	Homogenisation procedure	Monthly/ Daily	Multiple-breakpoint	Objective homogenisation method Executable program with automatic (and interactive) primary operation Deals with multiple inhomogeneous references Corrected series receive further corrections, until no break is found

Method	Type	Temporal resolution	Breakpoints detection	Characteristics
PRODIGE (Caussinus and Mestre, 2004)	Homogenisation procedure	Annual/ monthly	Multiple-breakpoint	Relative detection method with pairwise comparison Penalized likelihood as detection criterion Detection search based on Dynamic Programming Correction estimation based on ANOVA
Geostatistical simulation (Costa et al., 2008)	Homogenisation procedure	Annual/ Monthly	Multiple-breakpoint	Based on Direct Sequential Simulation using reference series Corrections can be applied by a statistic value of the local probability density function simulated at the candidate's location
ACMANT (Domonkos, 2011a)	Homogenisation procedure	Monthly	Multiple-breakpoint	Fully objective and fully automatic homogenisation method Executable program with automatic primary operation Relative detection method based on reference series Penalized likelihood as detection criterion Temperature only
ACMANT2 (Domonkos, 2015)	Homogenisation procedure	Monthly/ daily	Multiple-breakpoint	Extension of ACMANT (Domonkos, 2011a) for precipitation
Climatol (Guijarro, 2006)	Homogenisation software package	Monthly	Single-breakpoint	Objective homogenisation method R package with automatic primary operation Relative detection method based on reference and pairwise comparison

Method	Type	Temporal resolution	Breakpoints detection	Characteristics
RHTest (Wang, 2008)	Homogenisation software package	Monthly/ Daily	Single-breakpoint	Objective homogenisation method R source program with interactive primary operation Relative detection method with reference series
AnClim and Proclim DB (Štěpánek, 2008a; 2008b)	Homogenisation software package	Any	Single-breakpoint	Objective homogenisation method Executable program with interactive (and automatic) primary operation Relative detection method based on reference and pairwise comparison
USHCN (Menne and Williams Jr., 2009)	Homogenisation software package	Monthly	Single-breakpoint	Objective homogenisation method Fortran source program with automatic primary operation Trend-like inhomogeneities can be detected Relative detection method with pairwise comparison
HOMER (Mestre et al., 2013)	Homogenisation software package	Monthly	Multiple-breakpoint	Allows user to add subjective decisions based on metadata or research experiences R source program with interactive primary operation Relative detection method with pairwise comparison Correction estimation based on ANOVA
BAMS (Seidou and Ouarda, 2007)	Bayesian approaches	Any/ Monthly	Single-breakpoint	Relative detection method

Method	Type	Temporal resolution	Breakpoints detection	Characteristics
BARE (Seidou et al., 2007)	Bayesian approaches	Any/ Monthly	Single-breakpoint	Relative detection method
Bayesian change-point algorithm (Ruggieri, 2013)	Bayesian approaches	Any/ Monthly	Single-breakpoint	Absolute detection method Provides a measure of uncertainty
Bayesian multiple change-points and segmentation algorithm (Hannart and Naveau, 2009)	Bayesian approaches	Any/ Monthly	Single-breakpoint	Absolute detection method
Change-point detection algorithm (Gallagher et al., 2012)	Bayesian approaches	Daily	Single-breakpoint	Absolute detection method
BNHT (Beaulieu et al., 2010)	Bayesian approaches	Any/ Monthly	Single-breakpoint	Used as absolute detection method or applied to composite reference series Allows the integration of prior knowledge on the date of change from other sources (e.g. metadata or expert belief)

Table A.2 - List of studies where the homogenisation methods were applied

Homogenisation method	Reference	Climate variable	Temporal resolution	Study area	Analysed time period
Von Neumann ratio test	Barring et al. (1999)	Sea level pressure	Monthly	Southern Sweden	1780-1997
Von Neumann ratio test	Rodriguez et al. (1999)	Precipitation	Monthly	Barcelona, Spain	1850-1991
Von Neumann ratio test	Rodriguez et al. (2001)	Surface pressure	Daily and monthly	Barcelona, Spain	1780-1989
Von Neumann ratio test	Sahin and Cigizoglu (2010)	Temperature, precipitation, relative humidity and local pressure	Monthly	Turkey	1974-2002
Von Neumann ratio test	Santos and Fragoso (2013)	Precipitation	Daily	Northern Portugal	1950-2000
Von Neumann ratio test	Talaee et al. (2014)	Precipitation	Annual and monthly	Iran	1966-2005
Von Neumann ratio test	Wijngaard et al. (2003)	Temperature and precipitation	Daily	Europe	1901-1999
Wald-Wolfowitz Runs test	Costa et al. (2008)	Precipitation	Annual	Southern Portugal	1980-2001
Wald-Wolfowitz Runs test	Tayanç et al. (1998)	Temperature	Annual	Turkey	1951-1990
Mann-Kendall test	Baule and Shulski (2014)	Wind speed	Monthly	Beaufort/Chukchi Sea (Arctic)	1979-2009
Mann-Kendall test	Begert et al. (2005)	Temperature and precipitation	Monthly	Switzerland	1864-2000
Mann-Kendall test	Bohm et al. (2001)	Temperature	Monthly	Alps	1760-1998
Mann-Kendall test	Freiwan and Kadioglu (2008)	Precipitation	Annual and monthly	Jordan	1923-2000
Mann-Kendall test	Maugeri et al. (2004)	Sea level pressure	Daily	Po Plain	1765-2000
Mann-Kendall test	Piccarreta et al. (2013)	Precipitation	Daily	Southern Italy	1951-2010
Mann-Kendall test	Santos and Fragoso (2013)	Precipitation	Daily	Northern Portugal	1950-2000
Mann-Kendall test	Serra et al. (2001)	Temperature	Daily	Spain	1917-1998
Mann-Kendall test	Toreti and Desiato (2008)	Temperature	Daily	Italy	1961-2004
Mann-Kendall test	Turkes et al. (2009)	Precipitation	Secular trends	Turkey	1930-2002
Wilcoxon-Mann-Whitney test	Costa et al. (2012)	Precipitation	Annual	Portugal	1961-2000
Kruskal-Wallis test	Tayanç et al. (1998)	Mean temperatures	Annual	Turkey	1951-1990
Kruskal-Wallis test	Turkes et al. (2009)	Precipitation	Secular trends	Turkey	1930-2002

Homogenisation method	Reference	Climate variable	Temporal resolution	Study area	Analysed time period
Pettitt's test	Ashagrie et al. (2006)	Precipitation	Daily	Western Europe	1911-2000
Pettitt's test	Costa et al. (2008)	Precipitation	Annual	Southern Portugal	1980-2001
Pettitt's test	Costa and Soares (2009)	Precipitation	Daily	Southern Portugal	1941-2001
Pettitt's test	Firat et al. (2010)	Precipitation	Annual and monthly	Turkey	1968-1998
Pettitt's test	Firat et al. (2012)	Mean temperature	Annual	Turkey	1968-1998
Pettitt's test	Konnen et al. (2003)	Pressure and temperature	Recovering instrumental records	Japan	1819-1872
Pettitt's test	Rahimzadeh and Zavareh (2014)	Temperature	Annual	Iran	1960-2010
Pettitt's test	Sahin and Cigizoglu (2010)	Temperature, precipitation, relative humidity and local pressure	Monthly	Turkey	1974-2002
Pettitt's test	Salinger and Griffiths (2001)	Temperature and precipitation	Daily	New Zealand	1930-1998
Pettitt's test	Santos and Fragoso (2013)	Precipitation	Daily	Northern Portugal	1950-2000
Pettitt's test	Servat et al. (1997)	Total precipitation and number of rainy days	Annual	Ivory Coast	1950-1980
Pettitt's test	Talaee et al. (2014)	Precipitation	Annual and monthly	Iran	1966-2005
Pettitt's test	Tomozeiu et al. (2005)	Precipitation	Seasonal	Romania	1961-1996
Pettitt's test	Wijngaard et al. (2003)	Temperature and precipitation	Daily	Europe	1901-1999
Craddock's test	Brugnara et al. (2012)	Total precipitation and wet days	daily	Central Alps	1922-2009
Craddock's test	Maugeri et al. (2004)	Sea level pressure	Daily	Po Plain	1765-2000
Craddock's test	Puglisi et al. (2010)	Temperature		Tuscany, Italy	1955-2005
Buishand range test	Feidas et al. (2007)	Precipitation	Annual and seasonal	Greece	1955-2001
Buishand range test	Sahin and Cigizoglu (2010)	Temperature, precipitation, relative humidity and local pressure	Monthly	Turkey	1974-2002
Buishand range test	Santos and Fragoso (2013)	Precipitation	Daily	Northern Portugal	1950-2000
Buishand range test	Talaee et al. (2014)	Precipitation	Annual and monthly	Iran	1966-2005

Homogenisation method	Reference	Climate variable	Temporal resolution	Study area	Analysed time period
Double mass analysis	Burt and Howden (2011)	Precipitation	Daily	Oxford, UK	1827
Double mass analysis	Tsakalias and Koutsoyiannis (1999)	Precipitation	Annual	Greece	1961-1983
Double mass analysis	Wilson et al. (2005)	Precipitation (dendroclimatic reconstruction)	Seasonal	Bavarian Forest region, Germany	1510-2005
Bivariate test	Bližňák et al. (2015)	Temperature, precipitation and pressure	Annual and monthly	Portugal, Cape Verde, Angola, Mozambique, Goa (India), and Macau (China)	1863-2006
Bivariate test	Bradzil et al. (2000)	Temperature	Annual, seasonal and monthly	Czech Republic	1961-1999
Bivariate test	Sahin and Cigizoglu (2010)	Temperature, precipitation, relative humidity and local pressure	Monthly	Turkey	1974-2002
Bivariate test	Štěpánek and Zahradníček (2008)	Temperature, precipitation, water vapour pressure and wind speed	Daily	Czech Republic	1961-2007
Bivariate test	Zahradníček et al. (2014)	Precipitation	Monthly	Croatia	1940-2010
Two-phase regression	El Kenawy et al. (2013)	Temperature	Daily	Northeastern Spain	1900-2006
Two-phase regression	Sherwood et al. (2008)	Radiosonde data	Twice-daily	World	1959-2005
Multiple linear regression	El Kenawy et al. (2013)	Temperature	Daily	Northeastern Spain	1900-2006
Multiple linear regression	Li and Dong (2009)	Temperature	Annual	Southeastern China	1960-2001
Method of cumulative residuals	Costa and Soares (2006)	Precipitation	Annual	Southern Portugal	1931-2000
SNHT	Buishand et al. (2013)	Precipitation	Daily	Netherlands	1910-2009
SNHT	Firat et al. (2012)	Temperature	Annual	Turkey	1968-1998
SNHT	Jovanovic (2000)	Precipitation	Annual	Former Yugoslavia	1951-1998
SNHT	Klingbjer and Moberg (2003)	Temperature	Monthly	Northern Sweden	1802-2002
SNHT	Saboohi et al. (2012)	Temperature	Annual and monthly	Iran	1950-2007
SNHT	Sahin and Cigizoglu (2010)	Temperature, precipitation, relative humidity and local pressure	Monthly	Turkey	1974-2002

Homogenisation method	Reference	Climate variable	Temporal resolution	Study area	Analysed time period
SNHT	Santos and Fragoso (2013)	Precipitation	Daily	Northern Portugal	1950-2000
SNHT	Tomozeiu et al. (2005)	Precipitation	Seasonal	Romania	1961-1996
SNHT	Vicente-Serrano et al. (2010)	Precipitation	Daily	Northeastern Spain	1901-2002
SNHT	Wang et al. (2014)	Temperature and precipitation	Monthly	China (Jiangxi province)	1951-1999
SNHT	Zahradníček et al. (2014)	Precipitation	Monthly	Croatia	1940-2010
SNHT with trend	Piccarreta et al. (2013)	Precipitation	Daily	Southern Italy	1951-2010
MASH	Freitas et al. (2013)	Temperature	Monthly	Northern Portugal	1941-2010
MASH	Lakatos et al. (2013)	Temperature, precipitation, wind speed and direction, sunshine, cloud cover, global radiation, relative humidity and pressure	Daily	Carpathian Region (Czech Republic, Slovakia, Poland, Hungary, Ukraine, Romania and Serbia)	1961-2010
MASH	Li and Yan (2010)	Temperature	Daily	Beijing (China)	1960-2006
MASH	Mamara et al. (2013)	Temperature	Monthly	Greece	1960-2004
MASH	Seleshi and Camberlin (2006)	Precipitation	Seasonal	Ethiopia	1965-2002
PRODIGE	Alexandrov et al. (2004)	Temperature and precipitation	Monthly	Bulgaria	1893-2001
PRODIGE	Nemec et al. (2013)	Temperature	Daily	Austria	1948-2009
Geostatistical simulation	Costa and Soares (2009)	Precipitation	Annual	Southern Portugal	1980-2001
ACMANT	Freitas et al. (2011)	Temperature	Monthly	Portugal	1864-2010
ACMANT	Mamara et al. (2014)	Temperature	Monthly	Greece	1960-2004
Climatol	Mamara et al. (2013)	Temperature	Monthly	Greece	1960-2004
RHTest	Bližňák et al. (2015)	Temperature, precipitation and pressure	Annual and monthly	Portugal, Cape Verde, Angola, Mozambique, Goa (India) and Macau (China)	1863-2006
RHTest	Tsidu (2012)	Precipitation	Monthly	Ethiopia	1978-2007
RHTest	Wan et al. (2010)	Wind speed	Monthly	Canada	1953-2006
AnClim and Proclim DB	Azorin-Molina et al. (2014)	Wind speed	Monthly	Iberian Peninsula	1961-2011

Homogenisation method	Reference	Climate variable	Temporal resolution	Study area	Analysed time period
AnClim and Proclim DB	Bližňák et al. (2015)	Temperature, precipitation and pressure	Annual and monthly	Portugal, Cape Verde, Angola, Mozambique, Goa (India) and Macau (China)	1863-2006
USHCN	Menne and Williams Jr. (2009)	Temperature	Annual and monthly	USA	1900-2006
HOMER	Bližňák et al. (2015)	Temperature, precipitation and pressure	Annual and monthly	Portugal, Cape Verde, Angola, Mozambique, Goa (India) and Macau (China)	1863-2006
HOMER	Freitas et al. (2013)	Temperature	Monthly	Northern Portugal	1941-2010
HOMER	Mamara et al. (2014)	Temperature	Monthly	Greece	1960-2004

Table A.3 - Summary of comparison studies for homogenisation methods

Climatic variable, periodicity and location	Methods	Main conclusions	References
Temperature data set: annual mean difference series (Turkey)	Different methods to detect inhomogeneities using relative homogeneity techniques: graphical analysis, the non-parametric Kruskal-Wallis homogeneity test and the Wald-Wolfowitz Runs test. Monte Carlo Simulation was carried out to determine the efficiency of the detection.	• Kruskal–Wallis homogeneity test is sensitive to (a) jump, (b) trend and (c) different U values. • Sensitivities of Wald–Wolfowitz Runs test are (a) jump and (b) trend. • Both tests are not powerful enough to be used individually in the relative homogeneity analysis.	Tayanç et al. (1998)
Annual temperature data. Three sets of data were generated: homogeneous series (no steps), series with one step, and series with a random number of steps.	Eight methods tested: SNHT without trend; SNHT with trend; MLR; TPR; Wilcoxon rank sum test; sequential testing for equality of means; Bayesian approach without reference series; and Bayesian approach with reference series.	• Two methods seem to work slightly better than the others: SNHT without trend, and the MLR technique. • SNHT without trend, MLR and Bayesian with reference series are the most reliable techniques for the identification of homogeneous series. • SNHT without trend, MLR and TPR are the best approaches for the detection of a random number of steps, since they do not under-adjust the series as much as the other methods. • SNHT without trend has the best performance for detecting the correct number of steps.	Ducré-Robitaille et al. (2003)

Climatic variable, periodicity and location	Methods	Main conclusions	References
Simulated data series derived from: Temperature (215 data series, annual means) and precipitation (112 data series – annual totals), 98 – 100 years long, Hungary	T-test (Ducré-Robitaille et al., 2003); T-test (Kyselý and Domonkos, 2006); Buishand-test (maximum of the absolute values of accumulated anomalies); Buishand-test (difference between maximum and minimum values of accumulated anomalies); SNHT for shifts only; Wilcoxon Rank Sum test; MLR; Bayesian test with serial correlation analysis, Bayesian test with penalized maximum likelihood method for calculating number of change-points; Pettitt test; M-K test; method of Mestre; method of Mestre with parameterized minimum unit-length; SNHT for shifts and trends, TPR; MASH and MASH with parameterized minimum unit-length.	• The efficiency much more depends on the characteristics of the candidate series and quality of the reference series, than on the applied homogenisation method. • Overall, Mestre method and MASH are the most efficient homogenisation methods.	Domonkos (2006)
Annual temperature series. Those series were generated with different variance and correlation attributes.	Seven methods analysed: SNHT, Potter's method (BIVT), MLR, TPR, Bayes approach (BAYE), Parametric metadata-based test (PMETA), Non-parametric Metadata-based test (NMETA).	• Aside from PMETA and NMETA, SNHT and BIVT identified the greatest number of imposed single discontinuities within 20-or-more-year series. • TPR is able to detect multiple breaks, particularly when sequential breaks are close in time or have opposite signs. • MLR was found to be resilient to non-stationary difference series. • BAYE's performance is comparable to BIVT and SNHT for the large (<1σ anomalies) single step changes. • PMETA and NMETA detected the highest percentage of imposed single breaks.	DeGaetano (2006)

Climatic variable, periodicity and location	Methods	Main conclusions	References
Annual mean temperature series	Change point detection tested with SNHT, Wilcoxon's non-parametric test and TPR.	• There is no unique best procedure by any criteria.	Reeves et al. (2007)
60 and 100-year-long precipitation data series (Southern and central regions of the province of Quebec, Canada. Several thousand of homogeneous and inhomogeneous synthetic series)	Intercomparison of eight statistical tests to detect inhomogeneities in climatic data: SNHT, Multiple Regression (MLR), TPR, Bivariate test (BIVT), Sequential Wilcoxon test, Sequential Student t-test (STUS), Jaruskova's method (JARU), and Bayesian approach (BAYE1).	• None of these methods was efficient for all types of inhomogeneities, but some of them performed substantially better than others: BIVT, JARU, and SNHT. • Techniques such as the STUS and TPR led to the worst performances. • Techniques which gave a good performance on temperature series like the MLR were not necessarily appropriate for precipitation data. • Three methods had similar performances with all sets of synthetic series (BIVT, JARU and SNHT). • Some techniques cannot be applied efficiently to all types of series: MLR performed well for the identification of a homogeneous series and was good to identify a single shift. However, in the presence of multiple shifts, the performance of this method was poor. • BAYE1 performed well for the identification of one or multiple shifts, but detected too many non-existent shifts.	Beaulieu et al. (2008, 2009)

Climatic variable, periodicity and location	Methods	Main conclusions	References
Simulated data	Efficiencies for the detection parts of 15 homogenisation methods: Bayesian test with penalized maximum likelihood method for calculating the number of change-points; Bayesian test with serial correlation analysis (BAYE2); Buishand-test; Buishand-test extension (difference between maximum and minimum values of accumulated anomalies); PRODIGE; TPR; M-K test; MASH; MLR; Pettitt-test; SNHT for shifts only (SNHT); SNHT for shifts and trends (SNT); T-test; T-test (Kyselý and Domonkos, 2006); Wilcoxon Rank Sum test.	• TPR method has the most stable high performance in detection skill. • MASH, PRODIGE, BAYE2, SNHT and MLR have also favourably high detection skill. • Non-parametric methods, as well as t-tests and SNT have poorer results. • Surprisingly, M-K showed low detection skill in each experiment.	Domonkos (2008)
Maximum air temperature, minimum air temperature, mean air temperature, total precipitation, relative humidity and local pressure of 232 stations for the period 1974–2002 (Turkey)	Estimation of missing values using two different methods: Linear Regression (LR) and Expectation Maximization (EM) Algorithm. Homogeneity tested (for annual series) by one relative test, Bivariate test, and four absolute tests: SNHT for a single break, Buishand Range test, Pettit test and the Von Neumann ratio test	• EM Algorithm results were preferred. • Absolute tests failed to detect the inhomogeneities in the precipitation series at the significance level 1%.	Sahin and Cigizoglu (2010)

Climatic variable, periodicity and location	Methods	Main conclusions	References
OHOMs 10000 test-dataset records, in 100 year-long artificially simulated time series.	Comparison between eight methods: Multiple Linear Regression, PMT, SNHT for shifts only (SNH1 - including the common cutting algorithm and SNH2- supplied with the semi-hierarchic algorithm); SNHT for Shifts and Trends; T-test, PRODIGE and MASH.	• PRODIGE method and MASH showed the highest efficiency for power of detection, false alarm rate, detection skill and skill of linear estimation. •	Domonkos (2011b) and Domonkos et al. (2011)
Temperature	Comparison between different methods: PRODIGE, MASH, ACMANT, USHCN, the Craddock-test and the HOME-software.	• Six homogenisation methods can be recommended: PRODIGE, MASH, ACMANT, USHCN, the Craddock-test and the HOME-software. • ACMANT is a highly efficient tool for homogenising temperature datasets of mid-latitudes, but is not tailored to other variables. • For homogenising huge datasets, USHCN or ACMANT are recommendable, because these methods are fully automatic. • HOME-software, PRODIGE and MASH are usable in a wide range of tasks, but certain expertise is needed for their use. • Craddock-test is subjective and is inappropriate for homogenising large datasets.	Domonkos et al. (2012) and Domonkos (2013a)
Simulated time series (10 test data sets with different characteristics)	Bayes method (Ducré-Robitaille et al., 2003), PRODIGE, TPR, MASH, MLR, SNHT, SNT, T-test, Wilcoxon Rank Sum test	• In cases of high quality relative time series, PRODIGE is the most effective method; • Appreciably good results can be also achieved by MASH, Bayes method, MLR and SNHT.	Domonkos (2013b)

Climatic variable, periodicity and location	Methods	Main conclusions	References
Generated climatological series	Comparison of shift detection by six algorithms: T-test, SNHT, TPR, WMW, Durbin-Watson test (DW) and SRMD (Squared Relative Mean Difference). Monte Carlo Simulations were applied to find the efficiency of the tests.	• Best performances belong to T-test, SNHT and SRDM, giving almost identical results and showing that they belong to the same family of tests. • WMW follows, with good results, while DW and TPR both yield similar discouraging scores.	Guijarro (2013)
Monthly mean air temperature (Portugal)	MASH and HOMER to homogenize the air temperature database.	• MASH identifies the location of the break with the year of the shift, while HOMER is able to estimate the month of the change. • The number of breaks detected with HOMER is higher than with MASH. • The amplitude of the breaks detected with HOMER is, in general, higher than the amplitude of the breaks detected with MASH.	Freitas et al. (2013)
Generated yearly and monthly data	K-W, Friedman test (Friedman, 1937), Buishand range test, Pettitt test, Von Neumann ratio test, KPSS (Kwiatkowski *et al.*, 1992), ADF test (Said and Dickey, 1984), GAHMDI (Toreti et al., 2012), Bayesian technique of change-point analysis (Barry and Hartigan, 1992, 1993), Bayesian technique of change-point analysis with references, F-test, F-test with references, SNHT, RHTest, PRODIGE, TPR	• Best performances are provided by SNHT, RHTest, F-test with references, GAHMDI and F-test. • ADF and Von Neumann ratio tests are not reliable. • RHTest was considered the best test for trends' detection. • Poorer performances are assigned to Friedman, K-W and Pettitt tests.	Yozgatligil and Yazici (2016)

Acronyms

ACMANT – Adapted Caussinus-Mestre Algorithm for Homogenizing Networks of Temperature series (Domonkos, 2011a)

BAYE – Bayesian test (Perreault et al., 1999, 2000)

BAYE1 – Bayesian approach (Rasmussen, 2001)

BAYE2 – Bayesian test with serial correlation analysis (Ducré-Robitaille et al., 2003; Sneyers, 1999)

BIVT – Bivariate Test (Potter, 1981)

DW – Durbin-Watson (method)

HOMER – Homogenisation Software in R (Mestre et al., 2013)

JARU – Jaruskova's (method) (Jaruskova, 1996)

K-W – Kruskal-Wallis (test)

MASH – Multiple Analysis of Series for Homogeneity (Szentimrey, 1999)

M-K – Mann-Kendall (test)

MLR – Multiple Linear Regression (Vincent, 1998)

NMETA – Non-parametric Metadata-based (test) (Allen and DeGaetano, 2000)

PMETA – Parametric Metadata-based (test) (Karl and Williams, 1987)

PMT – Penalised Maximal t-test (Wang et al., 2007)

SNH1 – SNHT including the common cutting algorithm

SNH2 – SNHT supplied with the semi-hierarchic algorithm

SNHT – Standard Normal Homogeneity Test (Alexandersson, 1986)

SNT – Standard Normal Homogeneity Test with trend (Alexandersson and Moberg, 1997)

SRMD – Squared Relative Mean Difference

STUS – Sequential Student t-test

TPR – Two-Phase Regression (Easterling and Peterson, 1995)

USHCN – United State Historical Climatology Network (Menne and Williams Jr., 2009)

WMW – Wilcoxon-Mann-Whitney

References

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661–675.

Alexandersson H, Moberg A. 1997. Homogenization of Swedish Temperature Data. Part I: Homogeneity Test for Linear Trends. *International Journal of Climatology* **17**(1): 25–34. doi: 10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J.

Alexandrov V, Schneider M, Koleva E, Moisselin J-M. 2004. Climate variability and change in Bulgaria during the 20th century. *Theoretical and Applied Climatology* **79**(3-4): 133–149. doi: 10.1007/s00704-004-0073-4.

Allen RJ, De Gaetano AT. 2000. A method to adjust long-term temperature extreme series for non climatic inhomogeneities. *Journal of Climate* **13**(2): 3680–3695. doi: 10.1175/1520-0442(2000)013<3680:AMTALT>2.0.CO;2.

Allen R, Pereira L, Raes D, Smith M. 1998. Statistical analysis of weather data sets, Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements, FAO – Food and Agriculture Organization of the United Nations (eds.). *FAO Irrigation and drainage paper 56*, Annex IV, Rome.

Ashagrie AG, De Laat PJM, De Wit MJM, Tu M, Uhlenbrook S. 2006. Detecting the influence of land use changes on discharges and floods in the Meuse River Basin – the predictive power of a ninety-year rainfall-runoff relation? *Hydrology and Earth System Sciences* **10**: 691–701. doi: 10.5194/hess-10-691-2006.

Azorin-Molina C, Vicente-Serrano SM, McVicar TR, Jerez S, Sanchez-Lorenzo A, López-Moreno J-I, Revuelto J, Trigo RM, Lopez-Bustins J, Espírito-Santo F. 2014. Homogenization and Assessment of Observed Near-Surface Wind Speed Trends over Spain and Portugal, 1961–2011. *Journal of Climate* **27**(10): 3692–3712. doi: 10.1175/JCLI-D-13-00652.1.

Barring L, Jonsson P, Achberger C, Ekstrom M, Alexandersson H. 1999. The Lund instrumental record of meteorological observations : reconstruction of monthly sea-level pressure 1780 – 1997. *International Journal of Climatology* **19**: 1427–1443. doi: 10.1002/(SICI)1097-0088(19991115)19:13<1427::AID-JOC429>3.0.CO;2-.

Barry D, Hartigan JA. 1992. Product partition models for change point problem. *Annals of Statistics* **20**: 260–279.

Barry D, Hartigan JA. 1993. A Bayesian analysis for change point problems. *Journal of the American Statistical Association* **88**: 309–319.

Baule WJ, Shulski MD. 2014. Climatology and trends of wind speed in the Beaufort/Chukchi Sea coastal region from 1979 to 2009. *International Journal of Climatology* **34**(8): 2819–2833. doi: 10.1002/joc.3881.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X, Boulet G, Yagouti A. 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44**: W02425. doi: 10.1029/2006WR005615.

Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X. 2009. Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change-point models. *Water Resources Research* **45**: W08410. doi: 10.1029/2008WR007501.

Beaulieu C, Ouarda TBMJ, Seidou O. 2010. A Bayesian normal homogeneity test for the detection of artificial discontinuities in climatic series. *International Journal of Climatology* **30**(15): 2342–2357. doi: 10.1002/joc.2056.

Begert M, Schlegel T, Kirchhofer W. 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology* **25**(1): 65–80. doi: 10.1002/joc.1118.

Bližňák V, Valente MA, Bethke J. **2015**. Homogenization of time series from Portugal and its former colonies for the period from the late 19th to the early 21st century. *International Journal of Climatology*, **35**(9): 2400–2418. doi: 10.1002/joc.4151.

Bohm R, Auer I, Brunetti M, Maugeri M, Nann T, Schoner W. 2001. Regional Temperature Variability in the European Alps: 1760-1998 from Homogenized Instrumental Time Series. *International Journal of Climatology* **21**: 1779–1801. doi: 10.1002/joc.689.

Brazdil R, Štěpánek P, Kveton V. 2000. "Air Temperature Fluctuations in the Czech Republic in the period 1961-1999". In: *Proceedings of the Third Seminar for Homogenization and Quality Control in Climatological Databases*, World Meteorological Organization, Budapest, Hungary, pp. 15.

Brugnara Y, Brunetti M, Maugeri M, Nanni T, Simolo C. 2012. High-resolution analysis of daily precipitation trends in the central Alps over the last century. *International Journal of Climatology* **32**(9): 1406–1422. doi: 10.1002/joc.2363.

Buishand T. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**(1): 11–27. doi: 10.1016/0022-1694(82)90066-X.

Buishand TA, De Martino G, Spreeuw JN, Brandsma T. 2013. Homogeneity of precipitation series in the Netherlands and their trends in the past century. *International Journal of Climatology* **33**(4): 815–833. doi: 10.1002/joc.3471.

Burt TP, Howden NJK. 2011. A homogenous daily rainfall record for the Radcliffe Observatory, Oxford, from the 1820s. *Water Resources Research* **47**: W09701. doi: 10.1029/2010WR010336.

Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Applied Statistics* **53:** 405–425. doi: 10.1111/j.1467-9876.2004.05155.x.

Costa AC, Soares A. 2006. "Identification of inhomogeneities in precipitation time series using SUR models and the Ellipse test". In: *Proceedings of Accuracy 2006* Caetano M, Painho M (eds.), 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Instituto Geográfico Português, pp. 419–428.

Costa AC, Soares A. 2009. Homogenization of climate data: review and new perspectives using geostatistics. *Mathematical Geosciences* **41**(3): 291–305. doi: 10.1007/s11004-008-9203-3.

Costa AC, Negreiros J, Soares A. 2008. "Identification of inhomogeneities in precipitation time series using stochastic simulation". In: *geoENV VI – Geostatistics for Environmental Applications* Soares A, Pereira MJ, Dimitrakopoulos R (eds.) , Springer, Netherlands, pp. 275–282.

Costa AC, Santos JA, Pinto JG. 2012. Climate change scenarios for precipitation extremes in Portugal. *Theoretical and Applied Climatology* **108**(1-2): 217–234. doi: 10.1007/s00704-011-0528-3.

Craddock JM. 1979. Methods of comparing annual rainfall records for climatic purposes. *Weather* **34**: 332–346. doi: 10.1002/j.1477-8696.1979.tb03465.x.

DeGaetano AT. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate* **19**(5): 838–853. doi: 10.1175/JCLI3662.1

Domonkos P. 2006. "Testing of homogenization methods purposes, tools, and problems of implementation". In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases* Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.), World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1493, pp. 131–148.

Domonkos P. 2008. "Quantifying efficiency of homogenization methods". In: *Proceedings of the Sixth Seminar for Homogenization and Quality Control in Climatological Databases* Lakatos M. Szentimrey T, Bihari Z, Szalai S (eds.) , World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1576, pp. 20–32.

Domonkos P. 2011a. Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). *International Journal of Geosciences* **02**(03): 293–309. doi: 10.4236/ijg.2011.23032.

Domonkos P. 2011b. Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods. *Theoretical and Applied Climatology* **105**(3-4): 455–467. doi: 10.1007/s00704-011-0399-7.

Domonkos P. 2013a. Measuring performances of homogenization methods. *Időjárás* **117**(1): 91–112.

Domonkos P. 2013b. Efficiencies of inhomogeneity-detection algorithms: comparison of different detection methods and efficiency measures. *Journal of Climatology* **390945**: 15 pages. doi: 10.1155/2013/390945.

Domonkos P. 2015. Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology* **122**(1): 303–314. doi: 10.1007/s00704-014-1298-5.

Domonkos P, Venema V, Mestre O. 2011. "Efficiencies of homogenization methods: our present knowledge and its limitation". In: *Proceedings of the Seventh Seminar for Homogenization and Quality Control in Climatological Databases* Lakatos M. Szentimrey T. Vincze E (eds.) World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 11–18.

Domonkos P, Venema V, Auer I, Mestre O, Brunetti M. 2012. The historical pathway towards more accurate homogenisation. *Advances in Science and Research* **8**: 45–52. doi: 10.5194/asr-8-45-2012.

Ducré-Robitaille J-F, Vincent LA, Boulet G. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**(9): 1087–1101. doi: 10.1002/joc.924.

Easterling DR, Peterson TC. 1995. A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369–377. doi: 10.1002/joc.3370150403.

El Kenawy A, López-Moreno JI, Štěpánek P, Vicente-Serrano SM. 2013. An assessment of the role of homogenization protocol in the performance of daily temperature series and trends: application to northeastern Spain. *International Journal of Climatology* **33**(1): 87–108. doi: 10.1002/joc.3410.

Feidas H, Noulopoulou C, Makrogiannis T, Bora-Senta E. 2007. Trend analysis of precipitation time series in Greece and their relationship with circulation using surface and satellite data: 1955–2001. *Theoretical and Applied Climatology* **87**(1-4): 155–177. doi: 10.1007/s00704-006-0200-5

Firat M, Dikbas F, Koc AC, Gungor M. 2010. Missing data analysis and homogeneity test for Turkish. *Sadhana* **35**(6): 707–720.

Firat M, Dikbas F, Koc AC, Gungor, M. 2012. Analysis of temperature series: estimation of missing data and homogeneity test. *Meteorological Applications* **19**(4): 397–406. doi: 10.1002/met.271.

Freitas L, Pereira MG, Amorim L, Caramelo L, Mendes M, Nunes LF. 2011. “Portuguese temperature dataset homogeneity with Home R”. In: *Proceedings of the Seventh Seminar for Homogenization and Quality Control in Climatological Databases* Lakatos M, Szentimrey T, Vincze E (eds.), World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, pp. 33–44.

Freitas L, Pereira MG, Mendes M, Nunes LF. 2013. Homogeneity of monthly air temperature in Portugal. *Időjárás* **117**(1): 69–90.

Freiwan M, Kadioglu M. 2008. Climate variability in Jordan. International Journal of Climatology **28**: 69–89. doi: 10.1002/joc.

Friedman M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* **32**(200): 675–701.

Gallagher C, Lund R, Robbins M. 2012. Changepoint detection in daily precipitation data. *Environmetrics*, **23**(5): 407–419. doi: 10.1002/env.2146.

Guijarro J. 2006. “Homogenization of a dense thermo-pluviometric monthly database in the Balearic Islands using the free contributed R package “Climatol””. In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.). World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1493, pp. 25–36.

Guijarro JA. 2013. Climatological series shift test comparison on running windows. *Időjárás* **117**(1): 35–45.

Hannart A, Naveau P. 2009. Bayesian multiple change points and segmentation: Application to homogenization of climatic series, *Water Resources Research* **45** (10): W10444. doi: 10.1029/2008WR007689.

Jaruskova D. 1996. Change-point detection in meteorological measurement, *Monthly Weather Review* **124**: 1535–1543. doi:10.1175/1520-0493(1996)124<1535:CPDIMM>2.0.CO;2.

Jovanovic O. 2000. “The homogenization of long-term precipitation time series in FR Yugoslavia”. In: *Proceedings of the Third Seminar for Homogenization and Quality Control in Climatological Databases*, World Meteorological Organization, Budapest, Hungary, pp. 18.

Karl TR, Williams CN. 1987. An approach to adjusting climatological time series for discontinuous inhomogeneties. *Journal of Applied Meteorology and Climatology* **26**: 1744–1763. doi: 10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.

Kendall M. 1975. *Rank correlation methods*. Charles Griffin, London.

Klingbjer P, Moberg A. 2003. A composite monthly temperature record from Tornedalen in northern Sweden, 1802-2002. *International Journal of Climatology* **23**(12): 1465–1494. doi: 10.1002/joc.946.

Kohler M. 1949. Double-mass analysis for testing the consistency of records and for making adjustments. *Bulletin of the American Meteorological Society* **30**: 188–189.

Konnen GP, Zaiki M, Baede APM, Mikami T, Jones PD, Tsukahara T. 2003. Pre-1872 Extension of the Japanese Instrumental Meteorological Observation Series back to 1819. *Journal of Climate* **16**: 118–131. doi: 10.1175/1520-0442(2003)016<0118:PEOTJI>2.0.CO;2.

Kruskal W. 1952. A non-parametric test for the several sample problem. *The Annals of Mathematical Statistics* **23**: 525–540.

Kruskal W, Wallis W. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**: 583– 621.

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Economics* **4**: 159–178.

Kyselý J, Domonkos P. 2006. Recent increase in persistence of atmospheric circulation over Europe: comparison with long-term variations since 1881. International Journal of Climatology **26**(4): 461–483. doi: 10.1002/joc.1265.

Lakatos M, Szentimrey T, Bihari Z, Szalai S. 2013. Creation of a homogenized climate database for the Carpathian region by applying the MASH procedure and the preliminary analysis of the data. *Időjárás* **117**(1): 143–158.

Li Q, Dong W. 2009. Detection and Adjustment of Undocumented Discontinuities in Chinese Temperature Series Using a Composite Approach. *Advances in Atmospheric Sciences* **26**(1): 143–153. doi: 10.1007/s00376-009-0143-8.

Li Z, Yan Z. 2010. Application of Multiple Analysis of Series for Homogenization to Beijing Daily Temperature Series (1960-2006). *Advances in Atmospheric Sciences* **27**(4): 777–787. doi: 10.1007/s00376-009-9052-0.

Mamara A, Argiriou AA, Anadranistakis M. 2013. Homogenization of mean monthly temperature time series of Greece. *International Journal of Climatology* **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. *International Journal of Climatology* **34**(10): 3024–3043. doi: 10.1002/joc.3888.

Mann H. 1945. Non-parametric tests against trend. *Econometrica* **13**(3): 245–259. doi: 10.2307/1907187.

Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* **18**: 50–60.

Maugeri M, Brunetti M, Monti F, Nanni T. 2004. Sea-level pressure variability in the Po Plain (1765–2000) from homogenized daily secular records. International Journal of Climatology **24**(4): 437–455. doi: 10.1002/joc.991.

Menne MJ, Williams Jr. CN. 2009. Homogenization of Temperature Series via Pairwise Comparisons. *Journal of Climate* **22**(7): 1700–1717. doi: 10.1175/2008JCLI2263.1.

Mestre O, Domonkos P, Picard F, Auer I, Robin S, Lebarbier E, Bohm R, Aguilar E, Guijarro J, Vertachnik G, Klancar M, Dubuisson B, Štěpánek P. 2013. HOMER: a homogenization software – methods and applications. *Időjárás* **117**(1): 47–67.

Nemec J, Gruber C, Chimani B, Auer I. 2013. Trends in extreme temperature indices in Austria based on a new homogenised dataset. *International Journal of Climatology* **33**(6): 1538–1550. doi: 10.1002/joc.3532.

Perreault L, Haché M, Slivitzky M, Bobée B. 1999. Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment* **13**: 201–216. doi: 10.1007/s004770050039.

Perreault L, Bernier J, Bobée B, Parent E. 2000. Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology* **235**: 242–263. doi: 10.1016/S0022-1694(00)00271-7.

Pettitt AN. 1979. A non-parametric approach to the change-point detection. *Journal of Applied Statistics* **28**(2): 126–135. doi: 10.2307/2346729.

Piccarreta M, Pasini A, Capolongo D, Lazzari M. 2013. Changes in daily precipitation extremes in the Mediterranean from 1951 to 2010: the Basilicata region, southern Italy. *International Journal of Climatology* **33**(15): 3229–3248. doi: 10.1002/joc.3670.

Potter, K., 1981. Illustration of a new test for detecting a shift in mean in precipitation series. *M. Weather Review* **109**: 2040–2045.

Puglisi A, Bartolini G, Orlandini S. 2010. Climatic variability in Tuscany : homogeneity methods of climatic data and analysis of impacts on grapevine and olive trees. *Italian Journal of Agrometeorology* **1/2010**: 45–52.

Rahimzadeh F, Zavareh MN. 2014. Effects of adjustment for non-climatic discontinuities on determination of temperature trends and variability over Iran. *International Journal of Climatology* **34**(6): 2079–2096. doi: 10.1002/joc.3823.

Rasmussen P. 2001. Bayesian estimation of change-points using the general linear model, Water Resources Research **37**: 2723–2731. doi: 10.1029/2001WR000311.

Reeves J, Chen J, Wang XL, Lund R, Lu QQ. 2007. A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology* **46**(6): 900–915. doi: 10.1175/JAM2493.1.

Rodriguez R, Llasat M-C, Wheeler D. 1999. Analysis of the Barcelona precipitation series 1850-1991. *International Journal of Climatology* **19**(7): 787–801. doi: 10.1002/(SICI)1097-0088(19990615)19:7<787::AID-JOC382>3.0.CO;2-2.

Rodriguez R, Barriendos M, Jones PD, Martin-Vide J, Pena JC. 2001. Long pressure series for Barcelona (Spain). Daily reconstruction and monthly homogenization. *International Journal of Climatology* **21**: 1693–1704. doi: 10.1002/joc.696.

Ruggieri E. 2013. A Bayesian approach to detecting change points in climatic records. *International Journal of Climatology* **33**(2): 520–528. doi: 10.1002/joc.3447.

Saboohi R, Soltani S, Khodagholi M. 2012. Trend analysis of temperature parameters in Iran. *Theoretical and Applied Climatology* **109**(3-4): 529–547. doi:10.1007/s00704-012-0590-5.

Sahin S, Cigizoglu HK. 2010. Homogeneity analysis of Turkish meteorological data set. *Hydrological Processes* **24**(8): 981–992. doi: 10.1002/hyp.7534.

Said SE, Dickey DA. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**: 599–607. doi: 10.1093/biomet/71.3.599.

Salinger MJ, Griffiths GM, 2001. Trends in New Zealand daily temperature and Rainfall Extremes. *International Journal of Climatology* **21**: 1437–1452. doi: 10.1002/joc.694.

Santos M, Fragoso M. 2013. Precipitation variability in Northern Portugal: Data homogeneity assessment and trends in extreme precipitation indices. *Atmospheric Research* **131**: 34–45. doi: 10.1016/j.atmosres.2013.04.008.

Seidou O, Ouarda TBMJ. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43:** W07404, doi: 10.1029/2006WR005021.

Seidou O, Asselin JJ, Ouarda TBMJ. 2007. Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. *Water Resources Research* **43**: W08401, doi: 10.1029/2005WR004835.

Seleshi Y, Camberlin P. 2006. Recent changes in dry spell and extreme rainfall events in Ethiopia. *Theoretical and Applied Climatology* **83**(1-4): 181–191. doi: 10.1007/s00704-005-0134-3.

Serra C, Burgueno A, Lana X. 2001. Analysis of maximum and minimum daily temperatures recorded at fabra observatory (Barcelona, NE Spain) in the period 1917 – 1998. *International Journal of Biometeorology* **21**: 617–636. doi: 10.1002/joc.633.

Servat E, Paturel JE, Lubes H, Kouame B, Ouedraogo M, Masson JM. 1997. Climatic variability in humid Africa along the Gulf of Guinea Part I : detailed analysis of the phenomenon in Cote d' Ivoire. *Journal of Hydrology* **191**: 1–15. doi: 10.1016/S0022-1694(96)03068-5.

Sherwood SC, Meyer CL, Allen RJ, Titchner HA. 2008. Robust Tropospheric Warming Revealed by Iteratively Homogenized Radiosonde Data. *Journal of Climate* **21**(20): 5336–5352. doi: 10.1175/2008JCLI2320.1.

Sneyers R. 1999. "Homogenizing time series of climatological observations". In: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*. WCDMP 41, WMO-TD 1962, WMO, Geneva, pp. 27–46.

Štěpánek P. 2008a. AnClim – Software for Time Series Analysis. Dept. of Geography, Fac. of Natural Sciences, MU, Brno. http://www.climahom.eu/AnClim.html (accessed October 2013).

Štěpánek P. 2008b. ProClimDB – Software for Processing Climatological Datasets. CHMI, Regional Office Brno. http://www.climahom.eu/ProcData.html (accessed October 2013).

Štěpánek P, Zahradníček P. 2008. "Experiences with quality control and homogenization of daily series of various meteorological elements in Czech Republic 1961-2007". In: *Proceedings of the Sixth Seminar for Homogenization and Quality Control in Climatological Databases* Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.) World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1576, pp. 20–32.

Szentimrey T. 1999. "Multiple analysis of series for homogenization (MASH)". In: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, WMO, WCDMP-No. 41, WMO-TD No. 1962, pp. 27–46.

Talaee PH, Kouchakzadeh M, Shifteh Some'e B. 2014. Homogeneity analysis of precipitation series in Iran. *Theoretical and Applied Climatology* **118**(1-2): 297–305. doi: 10.1007/s00704-013-1074-y118.

Tayanç M, Dalfes HN, Karaca M, Yenigu O. 1998. A comparative assessment of different methods for detecting inhomogeneities in Turkish temperature. *International Journal of Climatology* **18**(5): 561–578. doi: 10.1002/(SICI)1097-0088(199804)18:5<561::AID-JOC249>3.0.CO;2-Y.

Tomozeiu R, Stefan S, Busuioc A. 2005. Winter precipitation variability and large-scale circulation patterns in Romania. *Theoretical and Applied Climatology* **81**(3-4): 193–201. doi: 10.1007/s00704-004-0082-3.

Toreti A, Desiato F. 2008. Temperature trend over Italy from 1961 to 2004. *Theoretical and Applied Climatology* **91**(1-4): 51–58. doi: 10.1007/s00704-006-0289-6.

Toreti A, Kuglitsch FG, Xoplaki E, Luterbacher J. 2012. A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology* **51**: 317–326. doi: 10.1175/JAMC-D-10-05033.1.

Tsakalias G, Koutsoyiannis D. 1999. A Comprehensive System for the Exploration and Analysis of Hydrological Data. *Water Resources Management* **13**: 269–302. doi: 10.1023/A:1008197511426.

Tsidu GM. 2012. High-Resolution Monthly Rainfall Database for Ethiopia: Homogenization, Reconstruction, and Gridding. *Journal of Climate* **25**(24): 8422–8443. doi: 10.1175/JCLI-D-12-00027.1.

Turkes M, Koc T, Saris F. 2009. Spatiotemporal variability of precipitation total series over Turkey. *International Journal of Climatology* **29**: 1056–1074. doi: 10.1002/joc.1768.

Vicente-Serrano SM, Beguería S, López-Moreno JI, García-Vera MA, Štěpánek P. 2010. A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *International Journal of Climatology* **30**(8): 1146–1163. doi: 10.1002/joc.1850.

Vincent LA. 1998. A Technique for the Identification of Inhomogeneities in Canadian Temperature Series. *Journal of Climate* **11**: 1094–1104. doi: 10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics* **12**(4): 367–395.

Wald A, Wolfowitz J. 1943. An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics* **14**(4): 378–388.

Wan H, Wang XL, Swail VR. 2010. Homogenization and Trend Analysis of Canadian Near-Surface Wind Speeds. *Journal of Climate* **23**(5): 1209–1225. doi: 10.1175/2009JCLI3200.1.

Wang Q, Riemann D, Vogt S, Glaser R. 2014. Impacts of land cover changes on climate trends in Jiangxi province China. *International Journal of Biometeorology* **58**(5): 645–60. doi: 10.1007/s00484-013-0645-z.

Wang XL. 2008. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *Journal of Applied Meteorology and Climatology* **47**(9): 2423–2444. doi: 10.1175/2008JAMC1741.1.

Wang XL, Wen QH, Wu Y. 2007 Penalized maximal t test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology* **46**(6): 916–-931. doi:10.1175/JAM2504.1.

Wijngaard JB, Klein Tank AMG, Konnen GP. 2003. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**(6): 679–692. doi: 10.1002/joc.906.

Wilcoxon F. 1945. Individual comparison by ranking methods. *Biometrics Bulletin* **1**: 80–83.

Wilson RJS, Luckman BH, Esper J. 2005. A 500 year dendroclimatic reconstruction of spring-summer precipitation from the lower Bavarian Forest region, Germany. *International Journal of Climatology* **25**(5): 611–630. doi: 10.1002/joc.1150.

Yozgatligil C, Yazici C. 2016. Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology* **36**: 62–81. doi: 10.1002/joc.4329.

Zahradníček P, Rasol D, Cindrić K, Štěpánek P. 2014. Homogenization of monthly precipitation time series in Croatia. *International Journal of Climatology* **34**: 3671–3682. doi: 10.1002/joc.3934.

Appendix B

Table B.1 - List of the 66 monitoring stations used in the study depicting the role of the station series (candidate in the set of 4 stations, candidate in set of 10 stations, or reference station).

ID	Name/Location	Role
SNIRH 21K.01	Azaruja	Candidate (set of 10)
SNIRH 22E.01	Águas de Moura	Reference station
SNIRH 22H.02	Santiago do Escoural	Candidate (set of 4)
SNIRH 22L.01	Redondo	Candidate (set of 10)
SNIRH 22M.01	Santiago Maior	Reference station
SNIRH 23E.01	Comporta	Candidate (set of 10)
SNIRH 23F.01	Montevil	Reference station
SNIRH 23G.01	Barragem de Pego do Altar	Reference station
SNIRH 23I.01	Alcáçovas	Reference station
SNIRH 23K.01	São Manços	Reference station
SNIRH 23L.01	Reguengos	Reference station
SNIRH 24I.01	Viana do Alentejo	Candidate (set of 10)
SNIRH 24J.02	Alvito	Reference station
SNIRH 24J.03	Cuba	Reference station
SNIRH 24K.01	Portel	Reference station
SNIRH 24K.02	Vidigueira	Reference station
SNIRH 24N.01	Amareleja (D.G.R.N.)	Reference station
SNIRH 25G.01	Azinheira Barros	Reference station
SNIRH 25P.01	Barrancos	Reference station
SNIRH 26I.01	Santa Vitória	Reference station
SNIRH 26I.02	Barragem do Roxo	Reference station
SNIRH 26J.04	Albernoa	Reference station
SNIRH 26K.01	Salvada	Reference station
SNIRH 26L.01	Serpa	Reference station
SNIRH 26L.02	Santa Iria	Reference station
SNIRH 26M.01	Herdade de Valada	Reference station
SNIRH 27G.01	Relíquias	Reference station
SNIRH 27H.01	Panóias	Reference station
SNIRH 27H.02	Barragem do Monte da Rocha	Reference station
SNIRH 27J.01	São Marcos da Ataboeira	Reference station
SNIRH 27J.02	Corte Pequena	Reference station
SNIRH 27J.03	Vale de Camelos	Reference station
SNIRH 27K.01	Algodôr	Reference station
SNIRH 27K.02	Corte da Velha	Reference station

ID	Name/Location	Role
SNIRH 28F.01	Odemira	Candidate (set of 10)
SNIRH 28H.01	Aldeia de Palheiros	Candidate (set of 10)
SNIRH 28I.01	Almodôvar	Reference station
SNIRH 28J.01	Alcaria Longa	Reference station
SNIRH 28J.03	Santa Barbara de Padrões	Reference station
SNIRH 28K.01	São João dos Caldeireiros	Reference station
SNIRH 28K.02	Álamo	Reference station
SNIRH 28L.01	Mértola	Reference station
SNIRH 29G.01	Sabóia	Candidate (set of 10)
SNIRH 29I.02	Santa Clara-a-Nova	Reference station
SNIRH 29J.05	Guedelhas	Reference station
SNIRH 29K.01	Martim Longo	Reference station
SNIRH 29K.03	Malfrades	Reference station
SNIRH 29L.03	Monte dos Fortes	Reference station
SNIRH 30E.01	Aljezur	Candidate (sets of 4 & 10)
SNIRH 30E.02	Marmelete	Reference station
SNIRH 30E.03	Barragem da Bravura	Reference station
SNIRH 30G.01	Alferce	Candidate (set of 4)
SNIRH 30H.03	São Bartolomeu de Messines	Reference station
SNIRH 30H.04	Santa Margarida	Reference station
SNIRH 30J.01	Barranco do Velho	Reference station
SNIRH 30K.01	Mercador	Reference station
SNIRH 30K.02	Picota	Candidate (set of 10)
SNIRH 30L.04	Alcaria (Castro Marim)	Reference station
SNIRH 31G.02	Porches	Reference station
SNIRH 31H.02	Algoz	Reference station
SNIRH 31J.01	São Brás de Alportel	Reference station
SNIRH 31J.04	Estoi	Reference station
SNIRH 31K.01	Santa Catarina (Tavira)	Reference station
SNIRH 31K.02	Quelfes	Reference station
ECA 666	Beja	Candidate (sets of 4 & 10)
ECA 675	Lisboa Geofísica	Reference station

Appendix C

Table C.1 - Variogram models of the monthly precipitation series from networks 5 and 9.

	Decade	1900-1929	1930-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999
January	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	18000	20000	21000	25000	27500	27500	22500
	Partial Sill	2100	2100	950	2800	1650	970	1110
February	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	22500	27500	25000	30000	21000	22000	20000
	Partial Sill	2300	2350	1970	1700	1430	650	1180
March	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	22500	25000	20000	25000	20000	17500	17500
	Partial Sill	3600	3000	2150	1400	1900	1700	1300
April	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	27500	25000	18500	30000	31000	22500	23000
	Partial Sill	1600	2800	3450	4100	4500	1500	1500
May	Model	Exponential	Exponential	Exponential	Spherical	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	23000	25000	22000	28000	20500	30000	20000
	Partial Sill	2350	2560	2350	3400	1500	1810	1100
June	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	17000	20000	25000	21500	21500	21000	22500
	Partial Sill	2100	1570	2700	1800	3400	2100	910
July	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Spherical
	Nugget	0	0	0	0	0	0	0
	Range	18500	18500	28000	25000	20000	26000	27000
	Partial Sill	1700	1400	2800	1600	1750	3130	1350
Aug	Model	Exponential	Exponential	Exponential	Spherical	Exponential	Exponential	Exponential

	Decade	1900-1929	1930-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999
	Nugget	0	0	0	0	0	0	0
	Range	18500	25000	30000	30000	15000	24000	18000
	Partial Sill	2900	1700	1000	900	1600	400	1500
September	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	27000	16500	23500	23000	24000	30000	15000
	Partial Sill	2100	1200	1900	1300	1110	2290	2100
October	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	15000	24000	19500	18000	22500	15000	15500
	Partial Sill	1500	1030	2340	950	1900	2700	1030
November	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	19000	14000	25000	28000	27000	19000	26000
	Partial Sill	1700	2400	2500	3100	1400	1240	830
December	Model	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential	Exponential
	Nugget	0	0	0	0	0	0	0
	Range	24000	30000	23500	30000	18500	28000	30000
	Partial Sill	2100	1350	2700	1300	2000	1450	715

Table C.2 - List of the years with breakpoints and outliers defined by the HOME project (the "truth"), and of the irregular years that were detected in the homogenisation exercises. Years marked in bold are correctly detected breakpoints (with a tolerance of 2 years), and years marked in bold and underlined are correctly detected outliers.

		BENCHMARK ("truth")		TEST 1		TEST 2		TEST 3		TEST 4		TEST 5		TEST 6	
	Station Code	**Years with outliers**	**Breakpoint years**	**Automatic gsimcli**	**Semi-automatic gsimcli**	**Automatic gsimcli**	**Semi-automatic gsimcli**	**Automatic gsimcli**	**Semi-automatic gsimcli**	**Automatic gsimcli**	**Semi-automatic gsimcli**	**Automatic gsimcli**	**Semi-automatic gsimcli**	**Automatic gsimcli**	**Semi-automatic gsimcli**
Network 5	21109001	----	1908, 1918, 1931, 1941, 1955, 1971, 1978, 1982, 1994	**1956**, **1957**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**	**1956**, **1957**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**	**1956**, **1957**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**	**1956**, **1957**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**	**1956**, **1957**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**	**1956**, 1963, 1966, **1970**, 1974-**1977**	**1918**, 1952, **1970**
	21142001	----	1934, 1969, 1973	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1947-1950, **1967**, 1989	1952	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1947-1950, **1967**, 1989	1952	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1947-1951, **1967**, 1989, 1991	1952	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1947-1951, **1967**, 1989, 1991	1952	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1947-1951, **1967**, 1989, 1991	1952	1900, 1904, 1905, 1908-1910, 1914, 1917, 1918, 1926, 1931, **1934**, **1935**, 1938, 1945-1951, **1967**, 1989, 1991	1952
	21310001	1925, 1985	1927, 1942, 1952, 1958, 1989	**1923, 1924**	1972	**1923, 1924**	1972	**1923, 1924**	1972	**1923, 1924**	1972	**1923, 1924**	1972	**1923, 1924**	1972, 1984
	21386001	----	1925, 1940, 1944, 1955	Homogeneous	1947	Homogeneous	1947	Homogeneous	1947	Homogeneous	1947	Homogeneous	1947	Homogeneous	1947
	21425001	1933, 1962	1908, 1924, 1931, 1934, 1968, 1988	**1930**, 1947, **1967**	**1924**, **1933**	**1930**, 1947, **1967**, 1998	**1924**, **1933**	**1930**, 1947, **1967**, 1998	**1924**, **1933**	**1930**, 1947, **1967**, 1998	**1924**, **1933**	**1930**, 1947, **1967**, 1998	**1924**, **1933**	**1925**-1927, **1930**, 1947, **1967**, 1998	**1924**, **1933**
	21454001	1906	1917, 1939, 1947, 1951, 1960, 1972, 1976, 1987, 1989, 1994	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940-1946**, 1953, 1956, **1976**, 1979-1981, 1984, **1987-1993**	Homogeneous	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940-1946**, 1953, 1956, **1976**, 1979-1981, 1984, **1987-1993**	Homogeneous	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940**, **1941**, 1944, **1946**, 1953, 1956, **1976**, 1979-1981,	Homogeneous	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940-1946**, 1953, 1956, **1976**, 1979-1981, 1984, **1987-1993**	Homogeneous	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940-1946**, 1953, 1956, **1976**, 1979-1981, 1984, **1987-1993**	Homogeneous	1901, 1902, 1907, 1911-1914, 1919, 1923, **1940-1946**, 1953, 1956, **1976**, 1979-1981, 1984, **1987-1993**	Homogeneous

								1984, **1987-1993**							
	21501003	----	1981, 1985	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous
	21584001	----	1955, 1957, 1963, 1970, 1975, 1978	1926, **1955**, 1967, 1989, 1994	Homogeneous	1926, **1955**, 1967, 1989, 1994	Homogeneous	1926, **1955**, 1967, 1989, 1994	Homogeneous	1926, **1955**, 1967, 1989, 1994	Homogeneous	**1955**, 1967, 1989, 1994	Homogeneous	**1955**, 1967, 1989, 1994	Homogeneous
	21711001	1948, 1956, 1979	1974	1950	1967	1950	1967	1950	1967	1950	1967	1950	1967	1950	1967
Network 9	21109001	----	1921, 1944, 1967, 1970, 1983	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous	Homogeneous
	21425001	----	1918, 1950, 1953	1900-1910, 1913-**1917**, 1921, 1939, 1946, 1960, 1974, 1993	1926	1900-1910, 1913-**1917**, 1921, 1939, 1946, 1960, 1974, 1993	**1917**, 1926	1900-1910, 1913, **1916**, **1917**, 1921, 1939, 1946, 1960, 1974, 1993	**1917**, 1926	1900-1910, 1913, 1916, **1917**, 1921, 1939, 1946, 1960, 1974, 1993	**1917**, 1926	1900-1910, 1913, 1916, **1917**, 1921, 1939, 1946, 1960, 1974, 1993	**1917**, 1926	1900-1910, 1913, 1916, **1917**, 1921, 1960, 1974	**1917**, 1926
	21454001	1977, 1990	1900, 1910, 1911, 1913, 1916, 1946, 1974	**1912**, 1931, 1951, 1986	1950	**1912**, 1931, 1951, 1986	1950	**1912**, 1931, 1951, 1986	1950	**1912**, 1931, 1951, 1986	1950	**1912**, 1931, 1951, 1986	1950	**1912**, 1986	1950
	21584001	1902, 1917	1999	1908, 1912, 1922, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1962-1973, 1978, 1979, 1983-1992, **1997**	1941	1908, 1912, 1922, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1962-1973, 1978, 1979, 1983-1992, **1997**	1941	1908, 1912, 1922, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1962-1973, 1978-1980, 1983-1992, **1997**	1941	1908, 1912, 1922, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1962-1973, 1978- 1980, 1983-1992, **1997**	1941	1908, 1912, 1922, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1962-1973, 1978-1980, 1983-1992, **1997**	1941	1907, 1908, 1912, 1923, 1927-1934, 1937-1941, 1948-1950, 1953-1958, 1964-1973, 1978, 1979, 1983-1988, 1991, 1992, **1997**	1906, 1918, 1941
	21711001	----	1935, 1937, 1951, 1987	Homogeneous	**1937**, **1952**, 1979	Homogeneous	**1952**, 1979	Homogeneous	**1937**, **1952**, 1979	Homogeneous	**1937**, **1952**, 1979	Homogeneous	**1937**, **1952**, 1979	Homogeneous	**1937**, **1952**, 1979

Appendix D

Homogenisation of a temperature benchmark data set

The automatic version of the gsimcli method (without adjustments for sudden shifts) was used to homogenise monthly temperature data of the COST-HOME benchmark (HOME project; COST Action ES0601), considering different sets of parameters. The following sections describe the study area and the surrogate temperature data, as well as the different implementation strategies of the gsimcli method. Finally, the results are detailed and discussed based on performance metrics.

Monthly temperature data series

The HOME benchmark has networks with 5, 9 and 15 stations. We selected the temperature surrogate network 4, which comprises 15 stations located in the Northwest of France (**Error! Reference source not found.**), covering a rectangular area of approximately 100000 km^2 (250 km x 400 km) with a relatively uniform orography. Network 4 is expected to be easier to homogenise when compared with the other 15-stations temperature network available in the benchmark, which is located in the Pyrenees area (Spain, Andorra and France). Two stations are located in the islands of Groix (station 56069001) and Ile-Yeu (station 85113001). Network 4 comprises temperature monthly data series for a period of 100 years (1900 – 1999). Missing data periods occur in the first three decades (1900 – 1930), and in the beginning of the fifth decade (1940 – 1945), completing a total of 180 years of missing monthly data (2160 monthly records are missing). Two stations (44184001 St. Nazaire, and 49281001 St. George des Gardes) have a complete set of 100 years of monthly temperature data. The most incomplete time series, with only 75 years of monthly data, is the station 61377001 St. Cornier des Landes.

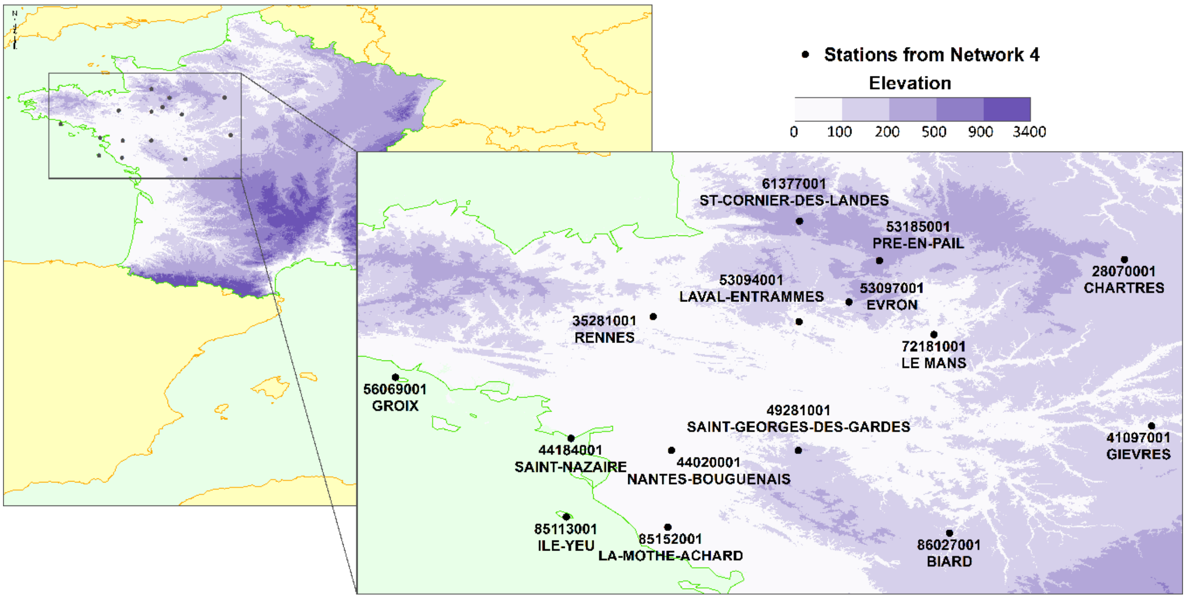

Figure S.1 - Location of stations from network 4 in the North of the Bay of Biscay (Digital Elevation Model source: Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled seamless SRTM data V4, http://srtm.csi.cgiar.org; accessed November 2015).

In network 4, the monthly temperature values vary from –0.5 ºC (observed in February 1946 in station 61377001) to 31.5 ºC (observed in July 1992, in station 41097001). Station 49281001 shows the highest range of monthly values (29.7 ºC). The stations' averages fluctuate between 12.3 ºC (station 61377001) and 16.6 ºC (station 41097001), corresponding to the stations where the minimum and the maximum values of temperature also occur. The correlation coefficients between the network stations are very high, varying from 0.893 (between stations 86027001 and 56069001) to 0.997 (between stations 53097001 and 28070001). The main spatial patterns were investigated for the annual temperature records of network 4. Three interpolation maps were elaborated for the years of 1935, 1966 and 1989 using the Inverse Distance Weighting (IDW) method. Neither an overall trend nor an anisotropic behaviour were observed in the interpolation maps, thus an overall isotropic pattern was assumed. Considering the isotropic behaviour of the variable, the high correlation coefficients between stations, and the size of the network, a single variogram model was estimated per month (**Error! Reference source not found.**). Although the correlation coefficients are high, the values of the range parameter in the monthly variogram models are surprisingly low, ranging between 77000 m (in October) and 102000 m (in December). The latter corresponds to approximately half of the minor length of the study area.

Table S.1 - Variogram models of the monthly temperature series from network 4.

Month	Model	Nugget	Partial Sill	Range
January	Exponential	0	4	90000
February	Exponential	0	4.7	95000
March	Exponential	0	4.89	83000
April	Exponential	0	5.5	88000
May	Exponential	0	4.97	88000
June	Exponential	0	6	82000
July	Exponential	0	6.8	84000
August	Exponential	0	5.7	95000
September	Exponential	0	5.413	79000
October	Exponential	0	5.25	77000
November	Exponential	0	4.5	77500
December	Exponential	0	3.7	102000

Specifications of the homogenisation exercises

Nine homogenisation exercises were undertaken for the monthly temperature series from network 4 using the automatic version of gsimcli with different sets of parameters (Table S.2). All homogenisation exercises follow a common set of parameters:

- Number of simulations (m) = 500;
- Detection parameter (p) = 0.95;
- Correction parameter = percentile value of 0.975.

Besides the size of the grid cells and the local radius parameters, which were assessed in the homogenisation exercises of precipitation, the order in which stations can be tested was also investigated. Three different strategies were evaluated: the descending order of variance (as in the homogenisation exercises of precipitation), the ascending order of variance, and the network deviation (the decreasing value of the difference between the station average and the network average). Three different grids were used: one grid with 5000 m cells (86 x 56 cells covering an area of 120400 km^2), and two grids with 10000 m cells (43 x 28 cells covering an area of 120400 km^2, and 50 x 40 cells covering an area of 200000 km^2). The values of the local radius parameter (r) vary between 0 and 2. An extended 10000 m grid was used in Test #6, where r is equal to 2 cells, because it is necessary to ensure that the minimum number of cells surrounding all the stations is at least 2.

Table S.2 - Parameters of the homogenisation exercises of monthly temperature data from network 4.

Test #	Grid cell size	Candidates order	Local radius parameter (r)
1	5000 m	Descending variance	0
2	5000 m	Descending variance	1
3	5000 m	Descending variance	2
4	10000 m	Descending variance	0
5	10000 m	Descending variance	1
6	10000 m (extended grid)	Descending variance	2
7	5000 m	Network deviation	1
8	5000 m	Ascending variance	1
9	5000 m	Network deviation	0

Results and discussion

All the homogenisation exercises undertaken with the monthly temperature data from network 4 provide identical performance metrics (Table S.3), thus it is not possible to determine which was the best modelling strategy. Changing the order of the candidate stations produced some differences regarding the adjusted values, but those differences did not significantly affect the performance metrics.

The results also show that the gsimcli homogenisation made the data slightly more inhomogeneous. Nonetheless, considering the Station CRMSE, the gsimcli method outperformed the absolute method (h008 - PMFred abs) and the Climatol (h010) by 17% and 11%, respectively. In terms of the Network CRMSE, the gsimcli homogenisation exercises show an efficiency improvement of 61% in comparison with the Climatol (h010).

Table S.3 - Performance metrics of the monthly temperature series from network 4 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Method	Station CRMSE	Station Improvement	Network CRMSE	Network Improvement
gsimcli Test #1	0.721	1.069	0.224	1.090
gsimcli Test #2	0.720	1.068	0.225	1.093
gsimcli Test #3	0.720	1.068	0.224	1.092
gsimcli Test #4	0.720	1.068	0.226	1.096
gsimcli Test #5	0.719	1.066	0.224	1.090
gsimcli Test #6	0.719	1.067	0.224	1.087
gsimcli Test #7	0.720	1.068	0.225	1.093
gsimcli Test #8	0.722	1.070	0.224	1.090
gsimcli Test #9	0.720	1.068	0.224	1.092
Inhomogeneous data	*0.674*	*1.0*	*0.206*	*1.0*
h002 - PRODIGE main	0.274	0.406	0.110	0.537
h003 - USHCN 52x	0.324	0.481	0.120	0.582
h004 - USHCN main	0.323	0.479	0.130	0.634
h005 - USHCN cx8	0.325	0.482	0.134	0.650
h006 - C3SNHT	0.569	0.844	0.196	0.951
h007 - PMTred rel	0.476	0.706	0.143	0.697
h008 - PMFred abs	0.868	1.288	0.180	0.878
h010 - Climatol	0.810	1.201	0.575	2.795
h011 - MASH main	0.285	0.423	0.109	0.531
h012 - SNHT DWD	0.498	0.739	0.191	0.928
h013 - PRODIGE trendy	0.268	0.398	0.110	0.534
h015 - ACMANT	0.300	0.444	0.127	0.618
h016 - iCraddock Vertacnik	0.284	0.422	0.108	0.526
h018 - AnClim main	0.472	0.701	0.195	0.949
h020 - PRODIGE Acquaotta	0.353	0.524	0.161	0.783
h021 - PRODIGE monthly	0.253	0.375	0.111	0.539
h022 - MASH Basic	0.302	0.448	0.128	0.622
h023 - MASH Light	0.300	0.445	0.130	0.633
h024 - MASH Strict	0.311	0.461	0.134	0.652
h025 - MASH No meta	0.317	0.471	0.138	0.673