

# Automatic Homogenization of Time Series: How to Use Metadata?

Peter Domonkos 

Independent Researcher, 43500 Tortosa, Spain; dpeterfree@gmail.com

**Abstract:** Long time series of observed climate data are often affected by changes in the technical conditions of the observations, which cause non-climatic biases, so-called inhomogeneities. Such inhomogeneities can be removed, at least partly, by the spatial comparison and statistical analysis of the data, and by the use of documented information about the historical changes in technical conditions, so-called metadata. Large datasets need the use of automatic or semiautomatic homogenization methods, but the effective use of non-quantitative metadata information within automatic procedures is not straightforward. The traditional approach suggests that a piece of metadata can be considered in statistical homogenizations only when the statistical analysis indicates a higher than threshold probability of inhomogeneity occurrence at or around the date of the metadata information. In this study, a new approach is presented, which suggests that the final inhomogeneity corrections should be done by the ANOVA correction model, and all the metadata dates likely indicating inhomogeneities according to the content of the metadata should be included in that correction step. A large synthetic temperature benchmark dataset has been created and used to test the performance of the ACMANT homogenization method both with traditional metadata use and with the suggested new method. The results show that while the traditional metadata use provides only 1–4% error reduction in comparison with the residual errors obtained by the homogenization without metadata, this ratio reaches 8–15% in the new, permissive use of metadata. The usefulness of metadata depends on the test dataset properties and homogenization method, these aspects are examined and discussed.

**Keywords:** time series; homogenization; metadata; benchmarking; ACMANT; ANOVA correction method; climate change detection



**Citation:** Domonkos, P. Automatic Homogenization of Time Series: How to Use Metadata? *Atmosphere* **2022**, *13*, 1379. <https://doi.org/10.3390/atmos13091379>

Academic Editor: Tianbao Zhao

Received: 14 August 2022

Accepted: 26 August 2022

Published: 28 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

We have long records of instrumental climate observations from a large part of the earth and for several climate variables. These records represent an enormous asset in the evaluation of past, present, and future climate changes. However, technical changes in the observation practices affect the temporal comparability of observed data. For instance, station relocations, changes in the instrumentation, observing staff, time schedule of observations, etc. may cause non-climatic biases, so-called inhomogeneities, in the time series of observed data. In fact, long series without the occurrences of such technical changes are rare [1], and the mean frequency of inhomogeneities is generally estimated to be 5–7 per 100 years [2–4]. The problem of inhomogeneities in climate records is well known since more than a century ago [5], but finding the best tools to remove inhomogeneities is still a challenging task [5,6]. The two main sources for any possible solution are the statistical analysis of time series and the use of documented information, so-called metadata, of the technical changes of observations.

The lists of documented technical changes would provide the optimal solution for inhomogeneity removal if such lists were complete, and the quantitative impact of each technical change would be known. For instance, if metadata show that a station relocation occurred in 1950 from station “A” to station “B”, we know that sections of data before 1950 are inhomogeneous when compared to data after 1950. We know the size of the change from metadata only when parallel observations in station “A” and station “B” were

performed for several years, and their results are saved among the metadata. In practice, metadata lists are incomplete, since a part of the technical changes are unintentional, and most metadata are unquantified [5,7,8].

The principal idea of statistical homogenization is that station-specific inhomogeneities can be made visible by the comparison of time series from nearby stations since the temporal evolution of climate is similar within a given climatic region. Such statistical procedures are named relative homogenization. Relative homogenization can be performed with or without the joint use of metadata. In the last few decades, several automatic methods have been constructed for the homogenization of large climatic datasets [5]. Metadata still offers additional information for effective homogenization, but the fruitful combination of a statistical method with unquantified pieces of information is not a simple task. This paper assesses the potential benefit of metadata use in automatic homogenization procedures on the example of the ACMANT homogenization method [9,10], which is tested on a large, synthetically developed monthly temperature benchmark dataset. Before the presentation of our own examinations, here we give a brief review of the usual metadata use in homogenization.

There is a wide consensus among experts regarding the generally high importance of metadata [5,8,11]. However, not all pieces of metadata have the same importance, and the most important ones are those, which point to synchronous technical changes in many or all stations of an observing network. In such cases, the basic idea of relative homogenization fails, but when metadata provide sufficient information, the inhomogeneities can be removed by a separate operation [12] performed before the general procedure of relative homogenization. The majority of inhomogeneities are station specific, and they are also referred to as station effects. Most frequently, a change in the station effect results in a sudden shift, a so-called break in the section mean values of the time series. Hereafter, metadata means station-specific metadata, except when the context specifies otherwise. In spite of the theoretical consensus about the importance of metadata, their practical treatment is varied in individual studies, and not only for the varying availability of metadata. Starting with studies in which little attention is dedicated to metadata, Pérez-Zanón et al. [13] omitted the metadata dates from break dates in homogenizing with the HOMER method [14]. In the homogenization of an integrated water vapor dataset, Nguyen et al. [15] reported that only 30–35% of the statistically detected breaks were confirmed by metadata, and they explained this by the ability of the statistical method to find relatively small breaks. By contrast, in many other studies, low ratios of metadata confirmation are often explained by the lack of metadata availability or they are interpreted as the overestimation of break frequency by statistical methods. In a few studies, statistically detected breaks without metadata confirmation are left out of consideration [16,17], partly because the limited possibilities of time series comparisons debilitated the reliability of statistical detection results in these studies. Finally, finding inconsistencies in break detection results, O’Neil et al. [18] compared the statistical break detection without metadata use to the use of imaginary maps citing an old economist [19]: “A man who uses an imaginary map, thinking that it is a true one, is likely to be worse off than someone with no map at all”.

We know a few studies in which the results of a relative homogenization method with metadata use were compared with those of the same homogenization method without metadata use. In all these studies [20–22], the compared homogenization results showed minor differences, so they could not confirm the usefulness of metadata in relative homogenization.

The value of metadata depends on the manner of their use. In automatic homogenization procedures, metadata indicating a likely break occurrence are often used for the precision of statistically detected break dates, when the metadata date falls within the confidence interval of the statistically detected break [5,23,24]. Further, lighter significance thresholds can be applied in statistical break detection when the breaks can be confirmed by metadata. For instance, coincidental statistical break detection results for seasonal and annual series of the same time series are expected when the breaks are not supported by

metadata, while breaks are accepted without such coincidences in the reverse case [25–27]. In slightly different ways, but the same logic of metadata use is applied in the newly developed “automated HOMER” of Joelsson et al. [28], and also in ACMANTv5 [9].

All the examples described in the previous paragraph represent restrictive metadata use where restriction means that the use of any piece of metadata is conditioned to some indications of statistical significance. By contrast, in permissive metadata use every piece of metadata is considered as a break position disregarding statistical break detection results. In this study, we will examine both restrictive and permissive ways of metadata use.

## 2. Materials and Methods

### 2.1. Benchmark Database

To test the usefulness of metadata, a large monthly temperature benchmark dataset has been developed. It consists of a seed dataset and further 40 datasets; in each of the latter sets, one parameter of the seed dataset is altered. Each dataset has homogeneous and inhomogeneous sections, and each of them contains 500 networks. Within a network, all time series cover the same period, and no data gap occurs. The datasets also include metadata, showing the dates of the inserted breaks. However, some pieces of metadata dates are false, i.e., they point to non-existing breaks, similarly to the occurrences of such instances in real-world metadata.

#### 2.1.1. Seed Dataset

In setting the parameters of the seed dataset, the aim was to provide a dataset, that, (a) includes several kinds of inhomogeneity problems, like multiple breaks, occurrences of gradual inhomogeneities, notable seasonal cycles of inhomogeneity magnitudes and significant network mean biases; (b) is characterized by moderately high signal-to-noise ratio; and (c) is realistic, i.e., the networks and their time series are similar to those of real-world data. To fulfill these expectations, a slightly modified version of the “U2” dataset of the MULTITEST project [29] was selected. The modifications included the reduction of the mean magnitude and maximum duration of short-term platform-shaped inhomogeneities, a slight increase in break frequency, and the enlargement of the dataset up to 500 networks. Each network consists of 10 time series, which are 60 years long. Here a shortened description of the dataset generation is provided.

The origin of the homogeneous set is from the synthetic daily temperature dataset performed for four U.S. regions [30,31]. The original series is 42 years long. The 210 homogeneous time series of the southeastern region, version 2 were taken (Wyoming data were used for U2), and 100 years long monthly series was created from them keeping the original spatial connections unchanged. See details of this step in [32]. Note that although the time series of the seed dataset is only 60 years long, 100 years long series was created for some other test datasets.

In generating the inhomogeneous set, inhomogeneities and outliers were randomly added to the time series. Three kinds of inhomogeneities are included: breaks, linearly changing biases, and short-term platform-shaped inhomogeneities. The mean frequencies for 100 years are 5 breaks, 1 linear change, 5 platform inhomogeneities, and 2 outliers, and the frequencies randomly vary between time series. The size distribution of inhomogeneities is Gaussian with 0 mean, and the standard deviation of inhomogeneity magnitudes for breaks and linear changes (platform inhomogeneities) is  $0.8\text{ }^{\circ}\text{C}$  ( $0.6\text{ }^{\circ}\text{C}$ ). The length of linear changes varies between 5 and 99 years. The length of platform inhomogeneities varies between 1 month and 60 months and the frequency quadratically increases with decreasing length. The sequence of inhomogeneities is “limited random walk” [32]. This concept means that inhomogeneity sizes are generally independent and they are simply added to the previous bias, however, threshold accumulated biases are set, which are not allowed to be exceeded in the dataset generation. The thresholds for accumulated biases differ according to the sign of the bias, and this resulted in notable network mean trend biases in the inhomogeneous dataset. Coincidental breaks in more than one time series of a

given network may accidentally occur, but synchronous or semi-synchronous breaks were not produced intentionally in the dataset creation. The seasonal cycle of inhomogeneity sizes follows a semi-sinusoid pattern in 75% of the inhomogeneities, while they are flat in the other cases. Find more details in [29].

The metadata list contains all the dates of the inserted breaks, while they do not contain dates related to gradually changing biases or short-term, platform-shaped inhomogeneities. 20% of the dates of the metadata list are false, i.e., they point to non-existing breaks. Note that in the homogenization of the seed dataset an arbitrary selected 25% of metadata were excluded, simulating metadata incompleteness.

### 2.1.2. Secondary Datasets

Secondary datasets were created from the seed dataset by changing a parameter of the dataset generation. Five parameters were varied, they are, (a) the number of time series in network, (b) time series length, (c) the standard deviation of inhomogeneity size, (d) mean spatial correlation between time series, (e) the ratio of false break dates in metadata. In addition, seeming metadata incompleteness were varied by manipulating a parameter of the homogenization procedure. As a general rule, only 1 parameter was altered in comparison with the seed dataset generation and homogenization, but there exists a deviation from this rule: each dataset with networks of 10 time series has a twin dataset with the only difference that the networks contain only 5 time series there. It is because the importance of metadata is higher for small networks than for larger networks.

The generation of secondary datasets was technically simple, only the variation of spatial correlations needs explication. The spatial correlations are calculated for the increment series of deseasonalized monthly values [32], and in this study, they are calculated for the homogeneous section of the data. The mean correlation in the source U.S. database is 0.883. Having randomly selected the time series, the mean spatial correlation in our seed dataset is the same. For raising (lowering) the mean correlation, minimum (maximum) correlation thresholds were set in the selection of time series to a given network. To lower the mean correlation more effectively, also Gaussian red noise processes of zero mean and 0.15 autocorrelation were added, but the latter had relatively smaller role than the use of correlation thresholds. When the mean correlation lowered most, down to 0.67, the standard deviation of the added noise was 0.3 °C.

The complete benchmark database consists of 41 datasets, 20,500 networks, and 170,000 time series.

## 2.2. Homogenization with ACMANT

ACMANT (Adapted Caussinus—Mestre Algorithm for the homogenization of Networks of climatic Time series) includes theoretically sophisticated and practically tested solutions for all important phases of the homogenization of climate time series. In the method comparison tests performed on the 1900 synthetic monthly temperature networks of the MULTITEST project, ACMANT produced the most accurate results [29]. ACMANTv4 is described in [10], while the changes in ACMANTv5 relative to the earlier versions are presented in [9]. Here we summarize a few important features of the method.

ACMANT homogenizes section means only. It applies a maximum likelihood multiple break detection method [33], which has univariate and bivariate modes within ACMANT [10]. The so-called combined time series comparison [9] is included, which unifies the advantages of pairwise comparisons and composite reference series use. It applies ensemble homogenization [10] to reduce random homogenization errors. The correction terms for inhomogeneity bias removal are calculated jointly for an entire network, by the equation system of the ANOVA correction model [33,34], which gives better results than any other known correction method in most practical cases [5]. ACMANT can be applied either to daily or monthly datasets, and it is characterized by high missing data tolerance. The newest version (ACMANTv5) already has both automatic and interactive versions.

When the climatic conditions suggest that inhomogeneity biases frequently have sinusoid or semi-sinusoid annual cycle, a bivariate homogenization can be applied, in which the breaks of annual means and those of summer—winter difference are jointly detected, and separate ANOVA correction term calculations are performed for these two variables [10]. In this study, both the univariate and bivariate ACMANT homogenizations are applied to each test dataset. The homogenizations are performed in fully automatic mode.

### 2.3. Metadata Use

Regarding metadata use, three kinds of homogenization were performed: (i) restrictive metadata use, (ii) permissive metadata use, and (iii) exclusion of metadata. Overall, each test dataset was homogenized in six modes by the variation of univariate and bivariate homogenizations and the variation of metadata use.

#### 2.3.1. Restrictive Metadata Use

In restrictive metadata use, the automatic mode of ACMANTv5.0 homogenization was applied [9,35]. Metadata are used in two steps of homogenization, and in both cases, the indication of some statistical results is needed to include pieces of metadata.

(a) Metadata use in the pairwise comparisons step of the homogenization: In summarizing the coincidental pieces of detection results for any date of an actually examined time series, referred to as candidate series, a confirming metadata date is considered in the same way as a confirming indication from the comparison between the candidate series and one of its neighbor series. A break existence is accepted when the total weight of the confirming pieces of information exceeds the threshold 2.1. Note that such pieces of detection results can be 0 or 1 in the original development of the automatic evaluation of pairwise comparison results [36], while they can also be fractions between 0 and 1 in ACMANT [9]. Pieces of metadata are always considered with weight 1.

(b) In the monthly precision step of the ACMANT procedure, metadata dates have a certain degree of preference. In monthly precision, the most likely break date is searched in a 28-month wide symmetric window around the date detected on an annual scale. Step functions of 1 step are fitted with varying step positions. In univariate homogenization, the sections before and after the step are flat, while the best fitting sinusoid annual cycles are included for them in bivariate homogenization [10], so that modified step functions are applied in bivariate homogenization. Generally, the step position producing the lowest sum of square errors (SSE) is selected as break date. When a metadata date occurs in the examined window, and the SSE exceeds the minimum SSE with no more than 2.0 (1.5) standard deviation of the examined data in univariate (bivariate) homogenization, the metadata date is selected as break date. When more metadata dates occur in an examined window, the one with the lowest SSE will be preferred.

#### 2.3.2. Permissive Metadata Use

In permissive metadata use, all kinds of metadata use described in Section 2.3.1 are kept, and additionally, all metadata dates are included in the final application of the ANOVA correction model regardless if any piece of statistical result indicates their significance or not. In the ANOVA model, every time series of observed climate data is considered to be the sum of a regionally common climate signal and a site-specific station effect, and the temporal evolution of station effects is described by step functions [10]. The input data of the model consists of the observed climatic data and the dates of the breaks related to known inhomogeneities or estimated inhomogeneities detected by statistical methods, hence the inclusion of metadata dates in the ANOVA correction model is straightforward.

### 2.4. Efficiency Measures

We use efficiency measures reporting directly about the success in reconstructing the climate trends and climate variability. Let  $V$  and  $Z$  stand for the vectors of homogeneous series and homogenized series with metadata use, respectively. The differences in the tem-

poral evolution in  $V$  and  $Z$  represent the residual errors after homogenization. In this study, these residual errors are characterized by 4 error measures: centered root mean squared error of monthly values (CRMSE<sub>m</sub>), centered root mean squared error of annual values (CRMSE<sub>y</sub>), mean absolute error of linear trends fitted to individual time series (Trbias), and mean absolute error of linear trends fitted to network mean time series (Trnetb).

#### 2.4.1. Centered Root Mean Squared Error

Centered root mean squared error differs from common root mean squared error in a way that the former excludes the difference between the means of the compared time series. Its use is reasoned by the fact that time series homogenization aims to reconstruct the temporal evolutions of climate data, but it does not and could not reconstruct station-specific climatic normal values. The concept of CRMSE was introduced to time series homogenization by [3], and its use is widespread since then. Equation (1) shows the calculation of CRMSE for time series of  $n$  data points. Equation (1) is usable both for monthly and annual time series.

$$\text{CRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( z_i - v_i - \frac{1}{n} \sum_{j=1}^n (z_j - v_j) \right)^2} \quad (1)$$

#### 2.4.2. Trend Bias

Linear regression with the minimization of mean quadratic difference is fitted to time series  $V$  and  $Z$ , and the trend slopes are denoted with  $\alpha_v$  and  $\alpha_z$ , respectively. Trbias and Trnetb can be calculated by Equation (2).

$$\text{Trend bias} = |\alpha_z - \alpha_v| \quad (2)$$

#### 2.4.3. Efficiency of Metadata Use

To examine the metadata effects on the homogenization results, the results with metadata use ( $z$ ) are compared to the results without metadata use ( $u$ ). Then the efficiency ( $f$ ) of metadata use is calculated as the percentage reduction of any error term ( $E$ ) by Equation (3).

$$f = 100 \cdot \frac{E(u) - E(z)}{E(u)} \quad (3)$$

The sign of  $f$  indicates if the metadata effect is positive or negative on the homogenization results, and in the theoretical case of perfect homogenization with metadata,  $f = 100\%$ . Equation (3) is used for datasets, but is not for individual time series, since  $E(u)$  can be very small or even zero for some series. In the examination of efficiencies for individual time series ( $f_1$ ) the efficiency is normalized with the dataset mean value of  $E(u)$ , Equation (4).

$$f_1 = 100 \cdot \frac{E(u) - E(z)}{\frac{1}{J} \sum_{k=1}^J E(U_k)} \quad (4)$$

In Equation (4),  $J$  stands for the number of time series in a given dataset. Differing from  $f$ ,  $f_1$  may have values above 100%.

### 3. Results

The results are presented in two main parts, firstly results averaged over the whole database are shown, then, in Section 3.2, we will examine the effects of changing dataset parameters, and compare the results of univariate and bivariate homogenizations.

### 3.1. Average Results for the Whole Database

Given that both univariate and bivariate homogenizations were performed and the simulated metadata completeness was varied, overall 102 dataset homogenizations were performed on the 41 test datasets.

Table 1 shows the mean errors for all homogenization experiments, before and after homogenization.

**Table 1.** Average errors for all homogenization experiments. CRMSE are in °C, trend biases are in °C/100 years unit.

|        | Raw Data | Metadata Use in Homogenization |            |            |
|--------|----------|--------------------------------|------------|------------|
|        |          | No Use                         | Restricted | Permissive |
| CRMSEm | 0.934    | 0.463                          | 0.458      | 0.398      |
| CRMSEy | 0.412    | 0.127                          | 0.124      | 0.115      |
| Trbias | 1.460    | 0.362                          | 0.352      | 0.324      |
| Trnetb | 0.894    | 0.326                          | 0.315      | 0.286      |

The permissive metadata use results in an 8.8–14.1% error reduction in comparison with the residual errors after homogenization without metadata use, while the reduction ratio is only 1.1–3.3% with the restrictive metadata use of ACMANT. Overall, Table 1 shows relatively small error reductions in both senses of absolute error reduction and the ratio of error removal. These results may be disappointing, but we must consider three factors for a balanced evaluation: (i) The raw data inhomogeneity magnitudes are moderated, which is not always the case in datasets of some climate variables, and even less in some datasets of less developed countries; (ii) The generally favorable signal-to-noise ratio facilitated an effective inhomogeneity removal of the ACMANT method even without metadata; (iii) The reduction of rarer large errors is more important than the reduction of average errors.

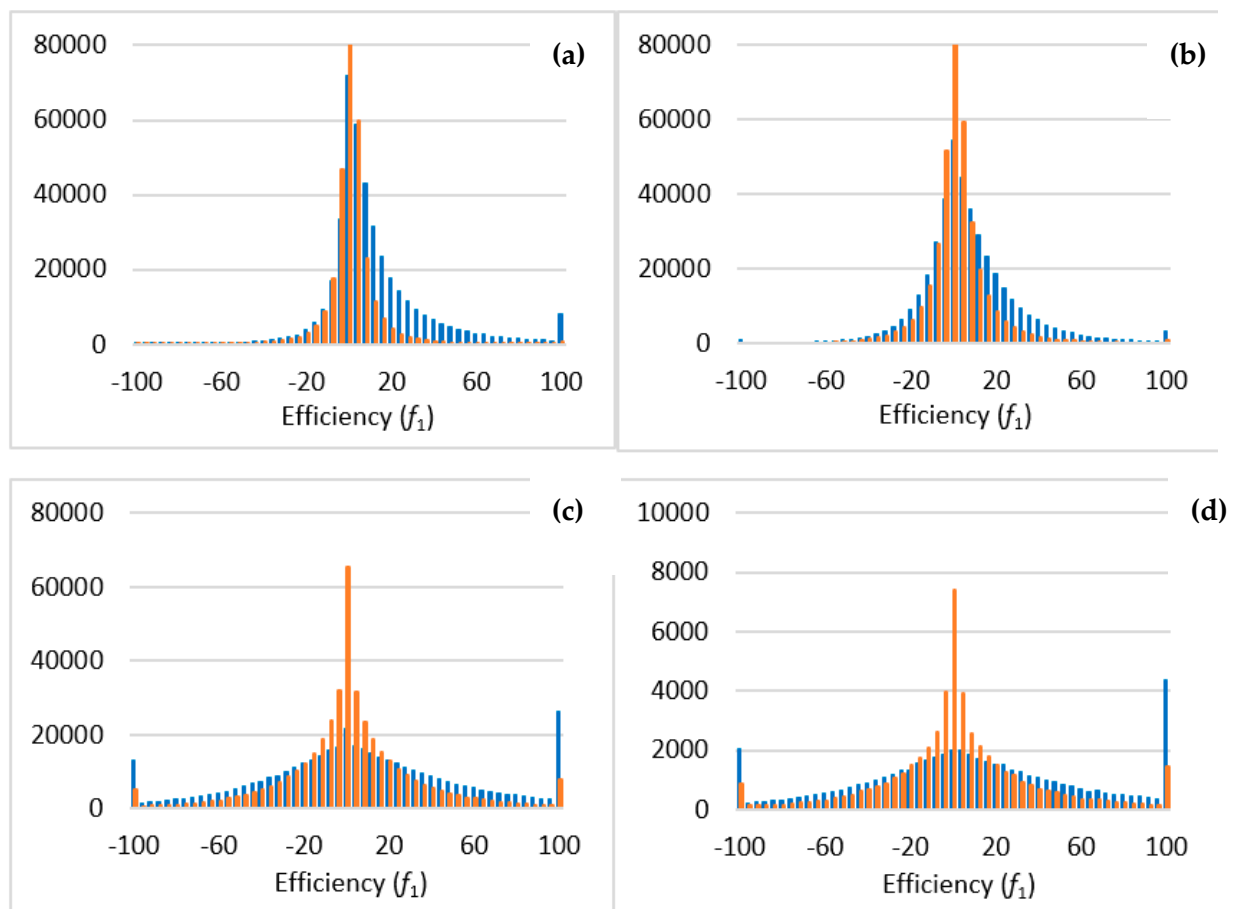
Now we examine the frequency distribution of error reductions expressed by efficiency  $f_1$ , the results are shown in Figure 1.

We can see that the spread of the error reduction efficiencies is higher for trend errors than for CRMSE, and the spread is higher in permissive metadata use than in restrictive metadata use. In trend errors both large improvement and large worsening occur with notable frequencies, particularly with permissive metadata use. With permissive metadata use, the frequency of large improvement is approximately double the frequency of large worsening.

Table 2 shows the impact of metadata use on the reduction of the largest inhomogeneity biases.

**Table 2.** Impact of metadata use on the upper tail of the probability distribution of residual errors after homogenization.

|        | Error Percentile 95            |            |            | Error Percentile 99            |            |            |
|--------|--------------------------------|------------|------------|--------------------------------|------------|------------|
|        | Metadata use in Homogenization |            |            | Metadata Use in Homogenization |            |            |
|        | No Use                         | Restricted | Permissive | No Use                         | Restricted | Permissive |
| CRMSEm | 0.962                          | 0.952      | 0.823      | 1.272                          | 1.263      | 1.142      |
| CRMSEy | 0.231                          | 0.225      | 0.208      | 0.302                          | 0.294      | 0.275      |
| Trbias | 0.982                          | 0.950      | 0.866      | 1.421                          | 1.377      | 1.280      |
| Trnetb | 0.905                          | 0.872      | 0.785      | 1.324                          | 1.282      | 1.188      |



**Figure 1.** Absolute values larger than 100 are unified in one column in each tail of the distribution function. Blue: permissive metadata use, brown: restrictive metadata use. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).

Interestingly, in Table 2 we find similar ratios of error reductions for high percentile values of the error distribution to those, which were shown for the mean error reductions in Table 1. The results of Table 2 confirm that permissive metadata use results in an 8–15% error reduction, while the restricted metadata use of ACMANT provides only a 1–4% error reduction in comparison with the results obtained without metadata use.

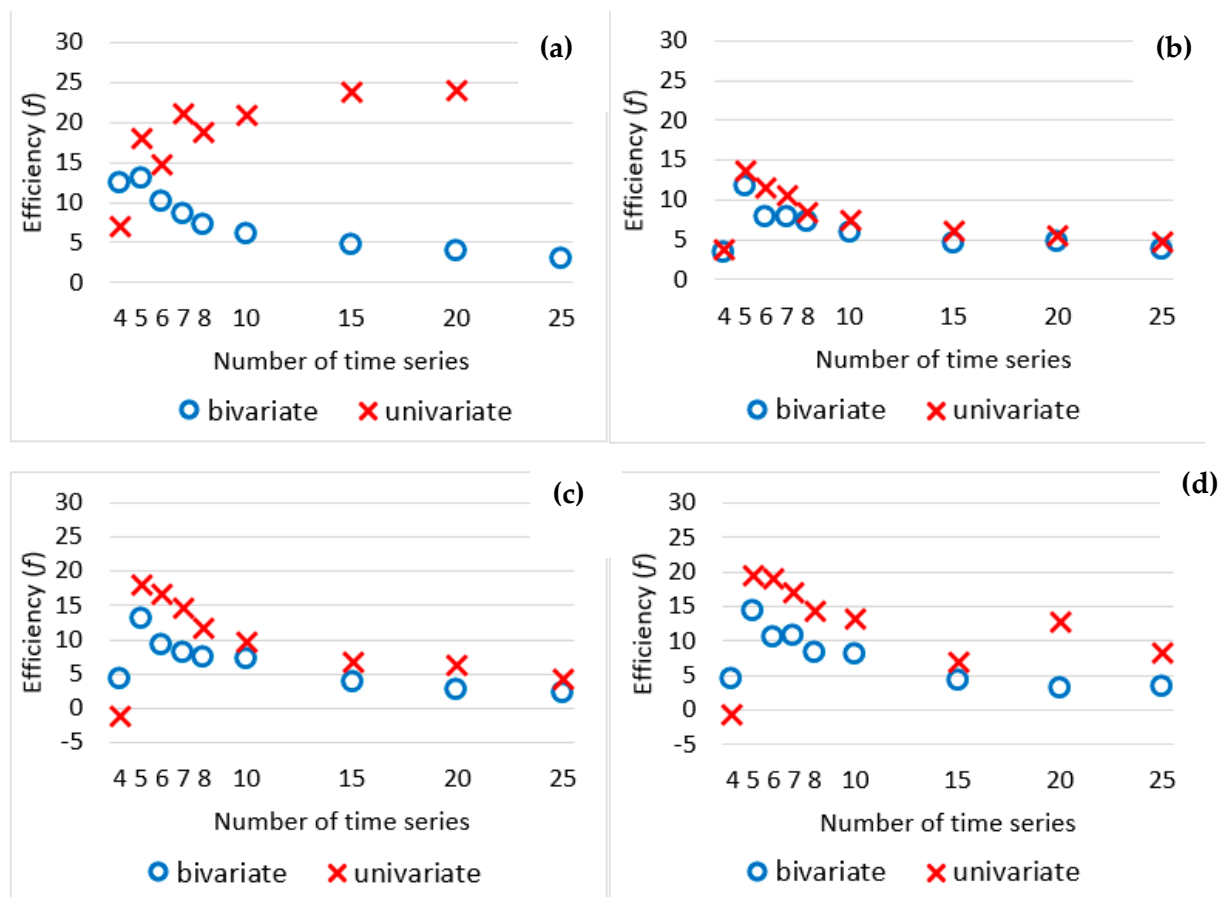
### 3.2. Effects of Changing Dataset Parameters

In this section, we study the results of univariate homogenization and bivariate homogenization separately. First, we examine the impact of the number of time series in networks (Section 3.2.1), since its importance merits distinguished attention. Then, in Section 3.2.2, further parameter change effects are presented.

#### 3.2.1. Impact of the Number of Time Series in Networks

Figure 2 shows the efficiency of metadata use in the reduction of inhomogeneity biases in the function of the number of time series in networks.





**Figure 2.** Efficiency of metadata use ( $f$ ) in the function of the number of time series in networks, and according to bivariate and univariate ACMANT homogenization. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).

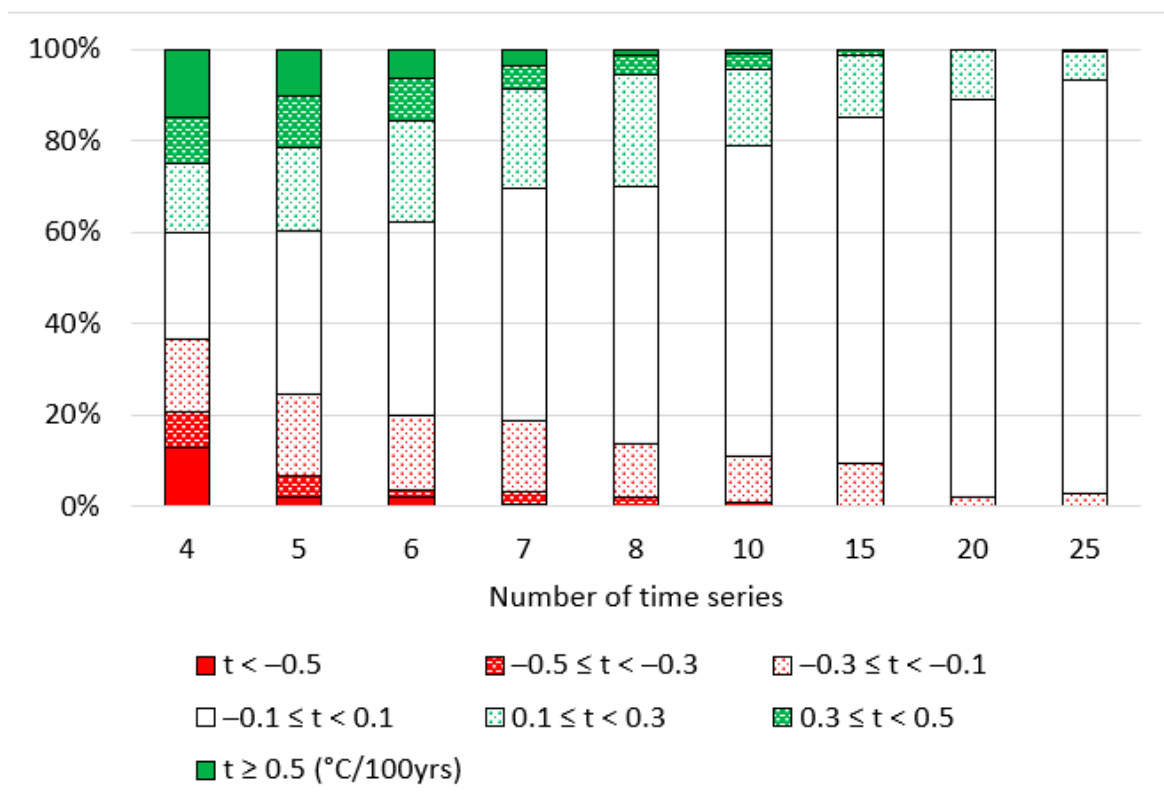
The results show that the importance of metadata generally increases with the decrease of network size, and it is the highest for networks of 5–6 time series. However, two features seem to contradict the general rule, and they are the low efficiency for networks of 4 time series, and the high efficiency in CRMSEm reduction in the univariate homogenization of large networks (Figure 2a). A further finding is that the efficiency of metadata use is higher with univariate homogenization than with bivariate homogenization for all error types, except for networks of 4 time series.

ACMANT limits the number of breaks with identical or very nearby dates to less than 50% of the number of time series. As a consequence, no synchronous or semi-synchronous breaks may occur in networks of 4 time series. ACMANT removes automatically the relatively less important breaks if this rule would be violated otherwise. The program gives preference to breaks supported by metadata, and the automatic reduction of break numbers may cause growing errors.

In ACMANT univariate homogenization, break detection is performed only for annual means. The exclusion of break detection for seasonal means is reasoned by the lower signal-to-noise ratio of seasonal series than that of annual series, which might cause the detection of spurious seasonal cycles. However, a weakness is that large breaks may remain undetected when they do not affect the annual means significantly [37]. Clearly, metadata can help a lot in reducing this kind of error in any size of network. The slightly higher efficiency of metadata use with univariate homogenization in reducing other kinds of errors than CRMSEm is unrelated to the problem of seasonal break detection. The bivariate homogenization of ACMANT is generally more powerful than the univariate homoge-

nization when many inhomogeneity biases have sinusoid or semi-sinusoid annual cycles. The results of Figure 2 point to the fact that metadata use may partly compensate for weaknesses of the applied statistical procedure.

We examine how the number of time series in networks influences the occurrences of large improvement or large worsening of inhomogeneity biases related to metadata use. More specifically, we examine network mean trend bias changes with the help of Figure 3, since Figure 1 showed that this error type is often strongly influenced by metadata use.



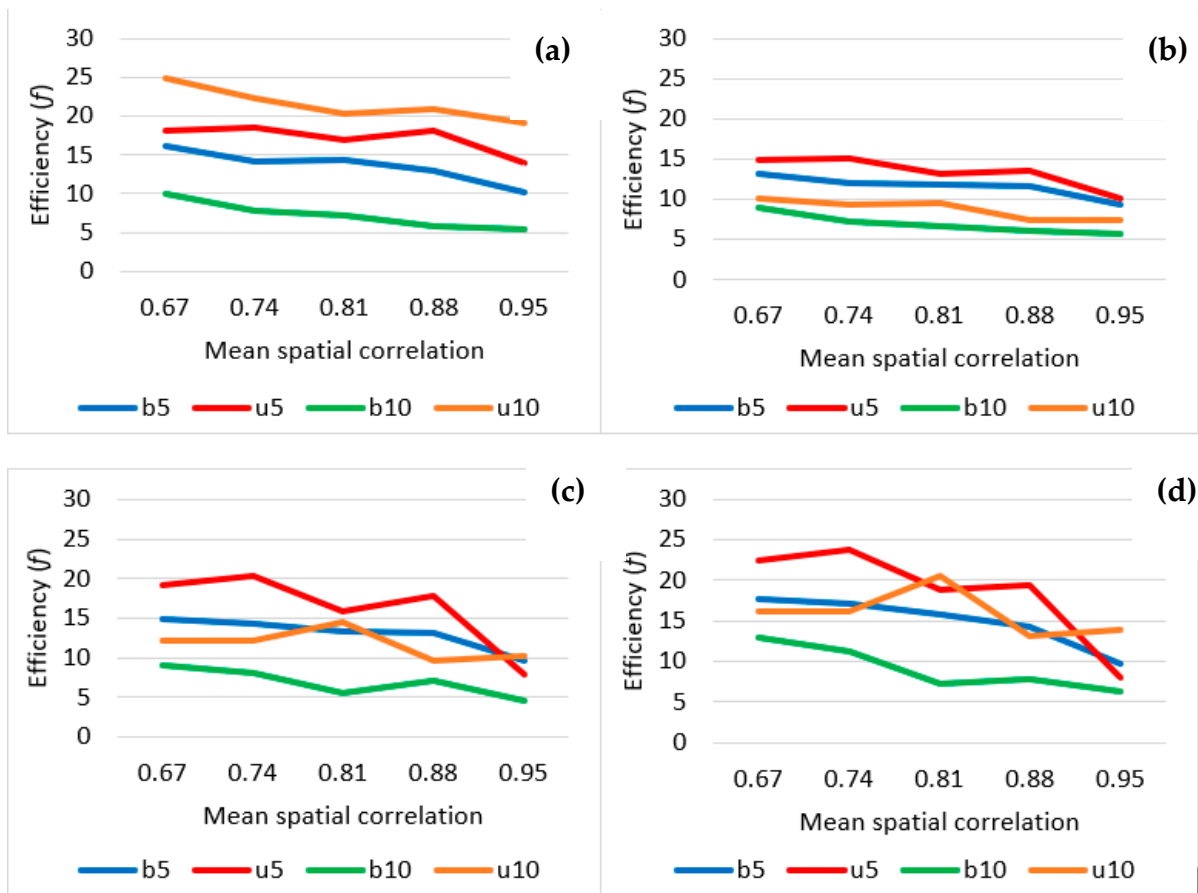
**Figure 3.** Frequency distribution of network mean trend error reduction (t) using °C/100 years unit. Green (red) color indicates improvement (worsening).

Figure 3 shows that the importance of metadata declines with the growing number of time series in networks in the sense that notable change in the network mean trend bias hardly occurs for large networks (number of time series  $\geq 20$ ). Figure 3 also confirms that the frequency of notable improvement is clearly higher than the frequency of notable worsening, except for networks of 4 time series.

### 3.2.2. Impacts of Other Parameter Changes

The efficiency of metadata use in the function of mean spatial correlation is presented in Figure 4.

The results show lowering metadata use efficiency with increasing spatial correlation. The likely explanation is that statistical homogenization is most successful when the spatial correlations are high, hence there remains little room to further improve the homogenization results by metadata use. We can see again that the metadata use efficiency is generally higher for small networks than for large networks, and it is generally higher with univariate homogenization than with bivariate homogenization. The unsmoothed curves of “u10” in Figure 4c,d indicate that in spite of the rather large sample sizes (500 networks per dataset), sampling errors still manifest, since the sizes of error reduction are generally small in absolute terms.



**Figure 4.** Efficiency of metadata use ( $f$ ) in the function of the mean spatial correlation in networks, in the examples of 4 settings of homogenization: b5—bivariate homogenization of networks of 5 time series; u5—univariate homogenization of networks of 5 time series; b10—bivariate homogenization of networks of 10 time series; and u10—univariate homogenization of networks of 10 time series. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).

The homogenization of longer time series is more challenging than that of the shorter ones, therefore we may expect that the efficiency of metadata use is higher for longer time series than for shorter ones. Figure 5 shows that these expectations are realistic.

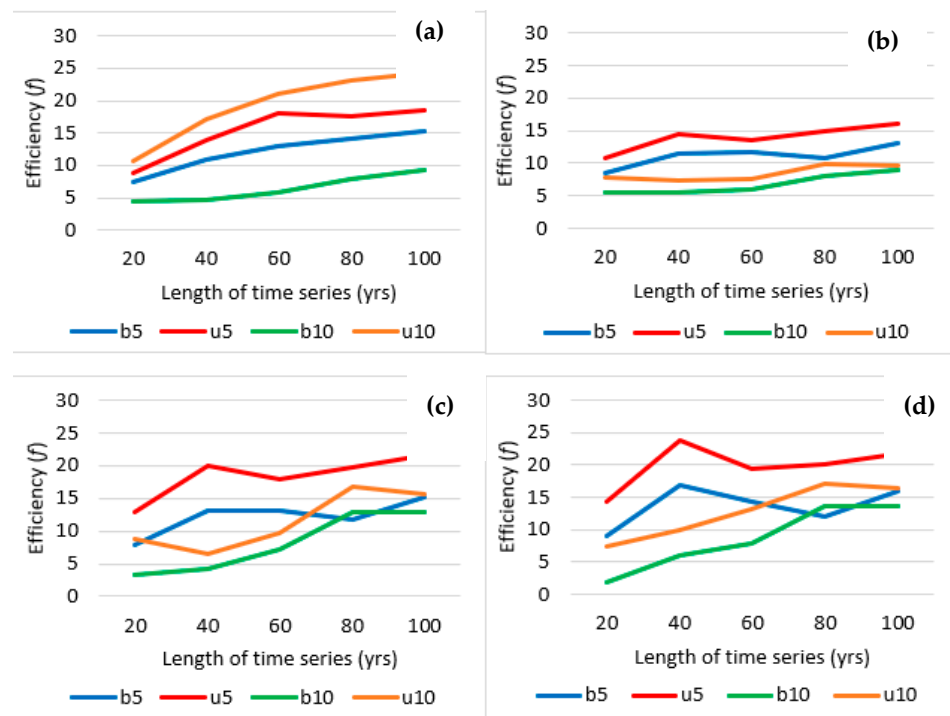
Continuing with changing the parameter that determines the inhomogeneity magnitude distribution, we find that the impact of changing this parameter is more complex (Figure 6) than the impacts of previously examined parameter changes.

The results of Figure 6 suggest that inhomogeneity magnitude has little effect on the efficiency of metadata use, except when these magnitudes are generally small. When generally small inhomogeneity magnitudes (standard deviation  $0.4\text{ }^{\circ}\text{C}$ ) cause a low signal-to-noise ratio, metadata use helps more in reducing monthly CRMSE (Figure 6a), but it cannot help a lot in reducing other kinds of inhomogeneity biases.

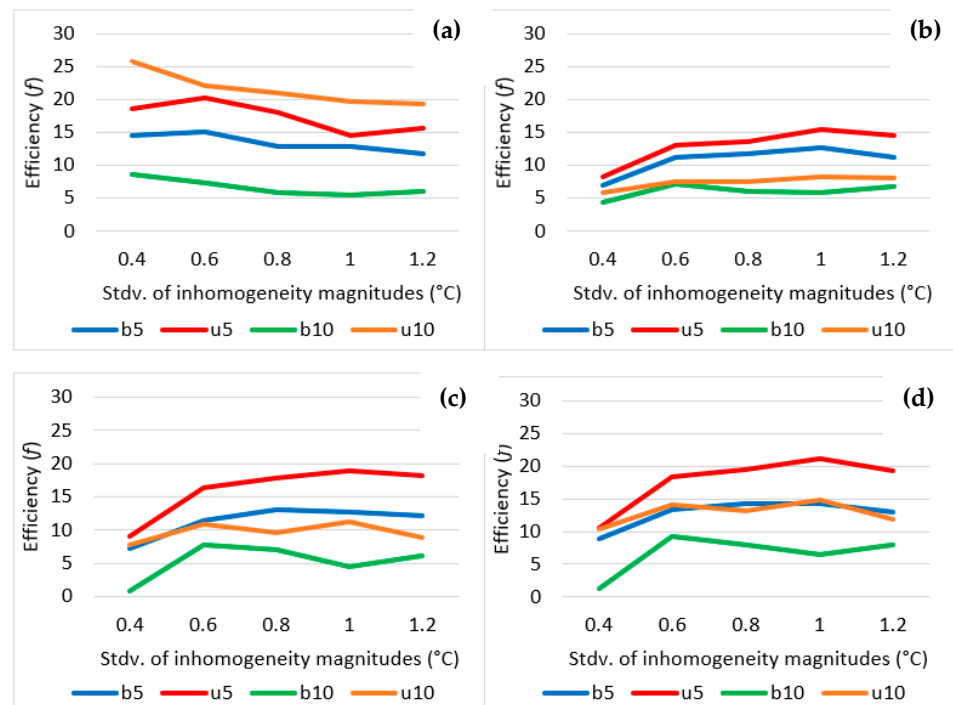
We examine the impact of metadata completeness with the help of Figure 7.

In accordance with our expectation, the efficiency of metadata use increases with the completeness of metadata. The good news is that the mean efficiency is always positive, even when only low ratios of metadata are available.

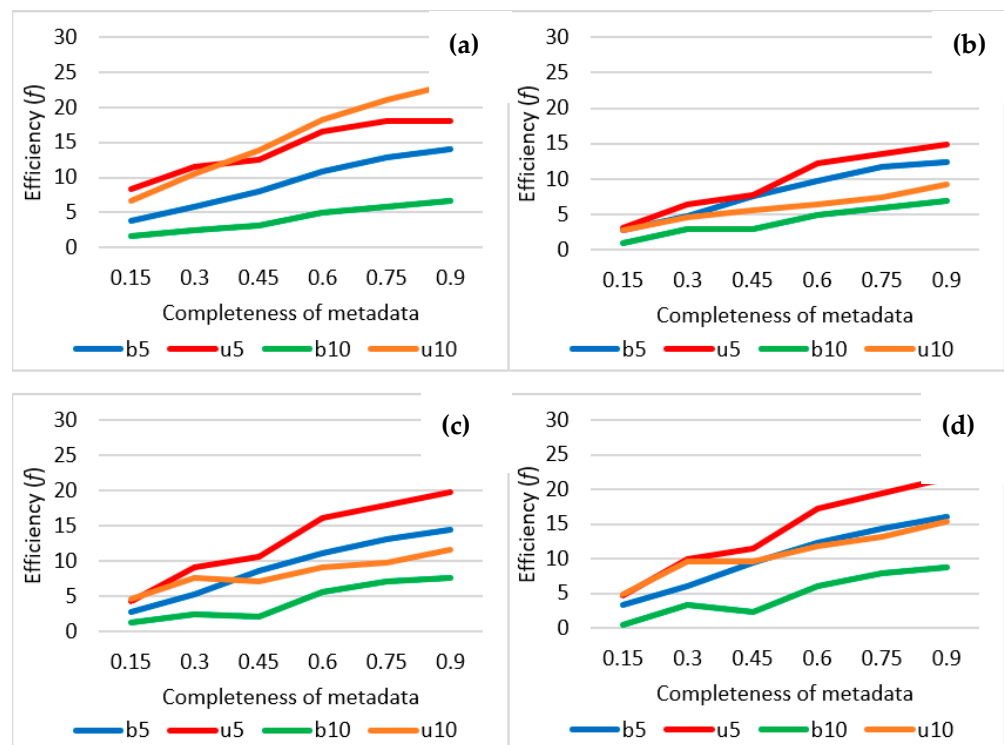
Finally, we examine the effect of the ratio of false metadata on the efficiency of metadata use (Figure 8).



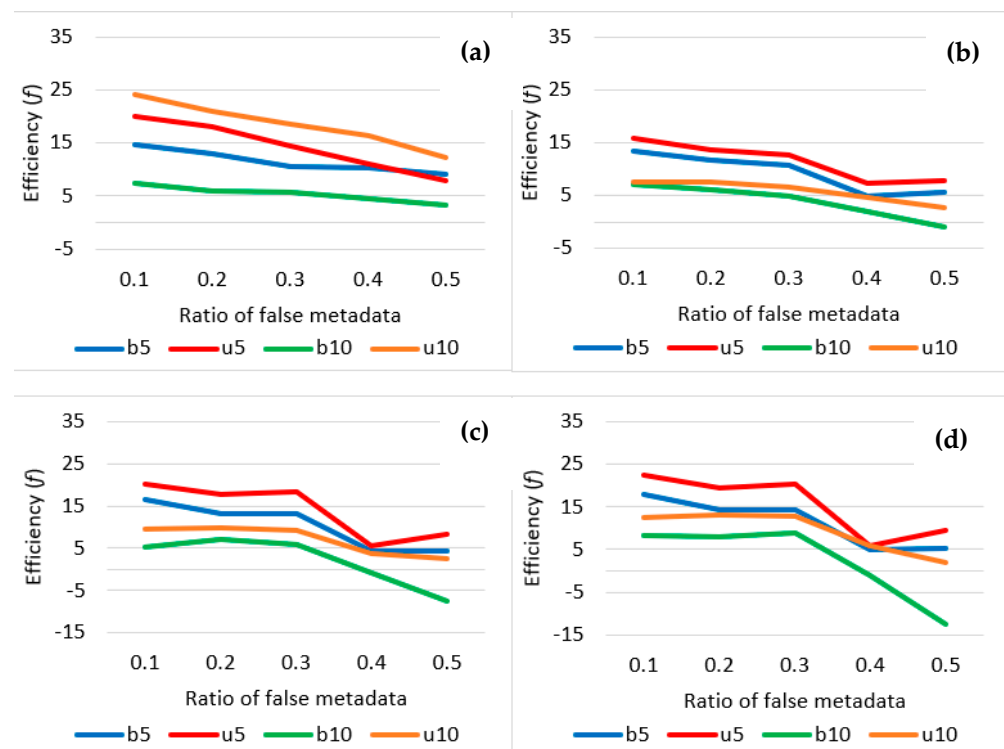
**Figure 5.** The same as Figure 4, but here the efficiencies are shown in the function of the length of time series. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).



**Figure 6.** The same as Figure 4, but here the efficiencies are shown in the function of the standard deviation of inhomogeneity magnitudes. Note that only the parameter modulating the magnitudes of breaks and gradually changing inhomogeneities is varied, while the standard deviation of short platform inhomogeneities is fixed (0.6 °C). Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).



**Figure 7.** The same as Figure 4, but here the efficiencies are shown in the function of the metadata completeness. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).

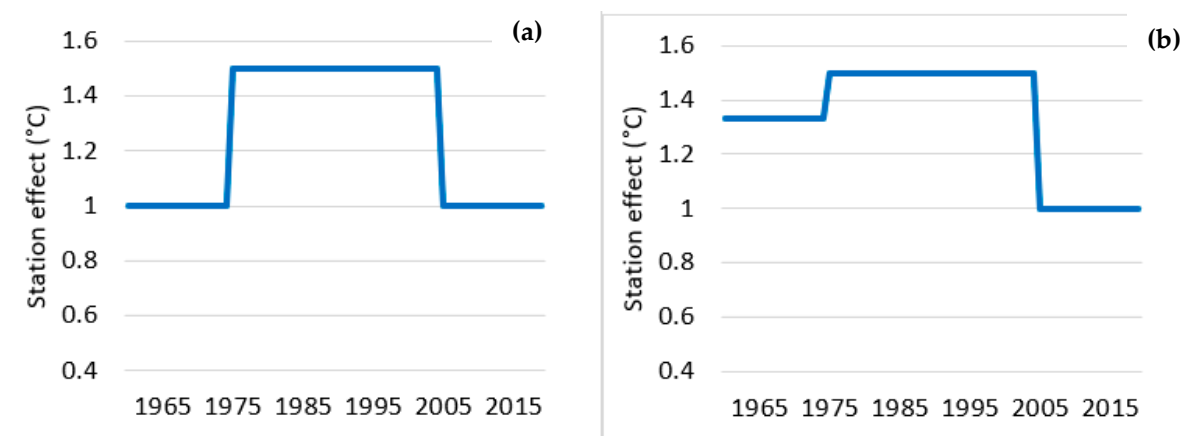


**Figure 8.** The same as Figure 4, but here the efficiencies are shown in the function of the ratio of false metadata. Upper left panel (a) CRMSEm, upper right panel (b) CRMSEy, bottom left panel (c) Trend bias for individual time series (Trbias), bottom right panel (d) Network mean trend bias (Trnetb).

As expected, the efficiency of metadata use decreases with the increase of the false metadata ratio. Our results suggest that the mean efficiency remains positive up to approximately 40% false metadata ratio, while it may be negative for higher ratios.

#### 4. Discussion

The presented examinations show that the inclusion of metadata use in automatic homogenization improves the accuracy of homogenization results. However, we have also found some unfavorable features. The average improvement of the homogenization accuracy for metadata use is not very large and may be smaller than expected. In addition, in the results of individual network homogenizations, sometimes notable worsening of the accuracy was found, although with lower frequency than the cases of notable improvement. The presented results are linked to the conditions of the test experiments i.e., (i) base statistical software (ACMANT), (ii) algorithm of metadata use, and (iii) test dataset properties. In evaluating if the settings of the presented experiments could produce some unfavorable results or not, we point to the general stochastic behavior of homogenization accuracy with a simple example (Figure 9).



**Figure 9.** Two breaks of a platform shaped inhomogeneity in a synthetic example of the variation of station effect. Left panel (a) before homogenization; right panel (b) after the adjustments for the first break when the second break has remained undiscovered.

Let us suppose that during the homogenization of a series, whose platform-shaped inhomogeneity is shown in Figure 9a, only the first break (the break in 1975) has been detected and adjusted. In the shown synthetic example, the unrevealed break in 2005 affected the calculation of the adjustment term for the break in 1975, but any additional errors for noise or neighbor series inhomogeneities (not shown) were zero. However, in the actual context, not the accuracy of the adjustment for the first break is the important point, but the fact that adjusting only one break of the existing two resulted in an increment of the trend bias for the period 1961–2020: The trend bias is zero in Figure 9a for symmetry reasons, while in Figure 9b the linear trend slope for 1961–2020 is  $-0.62$  °C/100 years. As most climate time series include multiple inhomogeneities of varied magnitude [3,38,39], the stochastic behavior of accuracy changes related to the adjustments of individual inhomogeneities is an inherent characteristic of homogenization. We can conclude that the accuracy improvements provided by metadata use are limited by the following factors: the efficiency of statistical procedure, the incompleteness of metadata, false metadata occurrences, and stochastic effects.

A general experience is that an undetected inhomogeneity tends to do more harm than the mistaken detection of a non-existing break. Nevertheless, false breaks have negative impacts on homogenization accuracy. Such negative impacts are the smallest when the breaks indicated by metadata but not confirmed by statistical tests are considered only in the final correction step of a homogenization procedure. This permissive metadata use

allows the benefit of all metadata but minimizes the error propagation between time series for the uncertainty of breaks indicated only by metadata.

The usefulness of metadata is influenced by their reliability and relevance. A metadata date is rarely erroneous, but some pieces of metadata can be irrelevant. While several types of technical changes almost always cause inhomogeneities, the relevance is less certain in some other cases, or even it can be doubtful, e.g., when metadata indicates “maintenance works”, “station inspection”, etc.

For time series of spatially dense and highly correlating observations, high homogenization accuracy can be achieved even without metadata use, and the presented tests confirm that metadata can have a minor role in such homogenization tasks. Although the mean efficiency of metadata use remains positive for any network density according to the presented tests, the possible exclusion of metadata use may sometimes be reasoned by the working load of metadata selection and metadata digitation. However, the following factors must be considered in connection with the possible exclusion of metadata use: (i) metadata of synchronous or semi-synchronous technical changes must be treated in a distinct way, and they never can be omitted from homogenization, (ii) The unevenness of spatial correlations or data gaps may cause the low spatial density of observed data for some stations or for some periods in an otherwise spatially dense dataset.

Networks of 4 time series can be homogenized by the automatic version of ACMANT when no metadata is available. When metadata is available, the use of a manual or interactive homogenization method is recommended, which can be the interactive version of ACMANTv5. Larger networks can be homogenized in automatic mode either with metadata or without metadata, although the inclusion of permissive metadata use in the ACMANT software is a task for the future at this time.

## 5. Conclusions

This study examined the use of station-specific metadata in the automatic homogenization with ACMANT. The impact of metadata use on the accuracy of homogenization results was evaluated by the use of a large-size synthetic monthly temperature database. The main conclusions are as follows:

- In the final correction step, all metadata dates should be included as break dates in the setting of the ANOVA correction model. This metadata use is called permissive metadata use. In the earlier phases of the homogenization procedure, metadata can be used only when their significance is confirmed by statistical tests.
- Permissive metadata use provides small to moderate accuracy improvement on average. Both the mean errors and the largest errors can generally be reduced by 8–15% by the inclusion of permissive metadata use, although the ratio depends on dataset properties.
- The usefulness of station-specific metadata (a) decreases with station density and spatial correlations, (b) increases with time series length, and (c) increases with the completeness and relevance of metadata.

**Funding:** This research was funded by the Catalan Meteorological Service, contract CM117.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The benchmark database is available at: <https://zenodo.org/record/6990845> (accessed on 14 June 2022).

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Moberg, A. Alexandersson, H. Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *Int. J. Climatol.* **1997**, *17*, 35–54. [CrossRef]
2. Auer, I.; Böhm, R.; Jurkovic, A.; Lipa, W.; Orlik, A.; Potzmann, R.; Schöner, W.; Ungersböck, M.; Matulla, C.; Briffa, K.; et al. HISTALP—historical instrumental climatological surface time series of the greater Alpine region. *Int. J. Climatol.* **2007**, *27*, 17–46. [CrossRef]
3. Venema, V.; Mestre, O.; Aguilar, E.; Auer, I.; Guijarro, J.A.; Domonkos, P.; Vertacnik, G.; Szentimrey, T.; Štěpánek, P.; Zahradnick, P.; et al. Benchmarking monthly homogenization algorithms. *Clim. Past* **2012**, *8*, 89–115. [CrossRef]
4. Lindau, R.; Venema, V.K.C. The joint influence of break and noise variance on the break detection capability in time series homogenization. *Adv. Stat. Clim. Meteorol. Oceanogr.* **2018**, *4*, 1–18. [CrossRef]
5. Venema, V.; Trewin, B.; Wang, X.L.; Szentimrey, T.; Lakatos, M.; Aguilar, E.; Auer, I.; Guijarro, J.; Menne, M.; Oria, C.; et al. Guidelines on homogenization. World Meteorological Organization. *Geneva* **2020**, 1245.
6. Domonkos, P. Application of homogenization methods for climate records. *Atmosphere* **2022**, *13*, 481. [CrossRef]
7. Buishand, T.A. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.* **1982**, *58*, 11–27. [CrossRef]
8. Aguilar, E.; Auer, I.; Brunet, M.; Peterson, T.C.; Wieringa, J. *Guidelines on Climate Metadata and Homogenization*; WCDMP-53, WMO-TD 1186; World Meteorological Organization: Geneva, Switzerland, 2003.
9. Domonkos, P. Combination of using pairwise comparisons and composite reference series: A new approach in the homogenization of climatic time series with ACMANT. *Atmosphere* **2021**, *12*, 1134. [CrossRef]
10. Domonkos, P. ACMANTv4: Scientific content and operation of the software. 71p. 2022. Available online: <https://github.com/dpeterfree/ACMANT> (accessed on 14 June 2022).
11. Auer, I.; Böhm, R.; Jurkovic, A.; Orlik, A.; Potzmann, R.; Schöner, W.; Ungersböck, M.; Brunetti, M.; Nanni, T.; Maugeri, M.; et al. A new instrumental precipitation dataset for the Greater Alpine Region for the period 1800–2002. *Int. J. Climatol.* **2005**, *25*, 139–166. [CrossRef]
12. Vincent, L.A.; Wang, X.L.; Milewska, E.J.; Wan, H.; Yang, F.; Swail, V. A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J. Geophys. Res.* **2012**, *117*, D18110. [CrossRef]
13. Pérez-Zanón, N.; Sigró, J.; Ashcroft, L. Temperature and precipitation regional climate series over the central Pyrenees during 1910–2013. *Int. J. Climatol.* **2017**, *37*, 1922–1937. [CrossRef]
14. Mestre, O.; Domonkos, P.; Picard, F.; Auer, I.; Robin, S.; Lebarbier, E.; Böhm, R.; Aguilar, E.; Guijarro, J.; Vertacnik, G.; et al. HOMER: Homogenization software in R—methods and applications. *Időjárás QJ Hung. Meteorol. Serv.* **2013**, *117*, 47–67.
15. Nguyen, K.N.; Quarello, A.; Bock, O.; Lebarbier, E. Sensitivity of change-point detection and trend estimates to GNSS IWV time series properties. *Atmosphere* **2021**, *12*, 1102. [CrossRef]
16. Zhang, L.; Ren, G.-Y.; Ren, Y.-Y.; Zhang, A.-Y.; Chu, Z.-Y.; Zhou, Y.-Q. Effect of data homogenization on estimate of temperature trend: A case of Huairou station in Beijing Municipality. *Theor. Appl. Climatol.* **2014**, *115*, 365–373. [CrossRef]
17. Jovanovic, B.; Smalley, R.; Timbal, B.; Siems, S. Homogenized monthly upper-air temperature data set for Australia. *Int. J. Climatol.* **2017**, *37*, 3209–3222. [CrossRef]
18. O’Neill, P.; Connolly, R.; Connolly, M.; Soon, W.; Chimani, B.; Crok, M.; de Vos, R.; Harde, H.; Kajaba, P.; Nojarov, P.; et al. Evaluation of the homogenization adjustments applied to European temperature records in the Global Historical Climatology Network Dataset. *Atmosphere* **2022**, *13*, 285. [CrossRef]
19. Schumacher, E.F. *Small is Beautiful: Economics as If People Mattered*; Reprint, Ed.; (Original 1973); Harper Perennial: New York, NY, USA, 2010; ISBN 978-0-06-199776-1.
20. Haimberger, L.; Tavolato, C.; Sperka, S. Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *J. Clim.* **2012**, *25*, 8108–8131. [CrossRef]
21. Gubler, S.; Hunziker, S.; Begert, M.; Croci-Maspoli, M.; Konzelmann, T.; Brönnimann, S.; Schwierz, C.; Oria, C.; Rosas, G. The influence of station density on climate data homogenization. *Int. J. Climatol.* **2017**, *37*, 4670–4683. [CrossRef]
22. Domonkos, P.; Coll, J.; Guijarro, J.A.; Curley, M.; Rustemeier, E.; Aguilar, E.; Walsh, S.; Sweeney, J. Precipitation trends in the island of Ireland using a dense, homogenized, observational dataset. *Int. J. Climatol.* **2020**, *40*, 6458–6472. [CrossRef]
23. Szentimrey, T. Multiple Analysis of Series for Homogenization (MASH). In *Second Seminar for Homogenization of Surface Climatological Data WMO WCDMP-41*; Szalai, S., Szentimrey, T., Szinell, C., Eds.; World Meteorological Organization: Geneva, Switzerland, 1999; pp. 27–46.
24. Williams, C.N.; Menne, M.J.; Thorne, P. Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res. Atmos.* **2012**, *117*, D05116. [CrossRef]
25. Xu, W.; Li, Q.; Wang, X.L.; Yang, S.; Cao, L.; Feng, Y. Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res. Atmos.* **2013**, *118*, 9708–9720. [CrossRef]
26. Xu, W.; Sun, C.; Zuo, J.; Ma, Z.; Li, W.; Yang, S. Homogenization of monthly ground surface temperature in China during 1961–2016 and performances of GLDAS reanalysis products. *J. Clim.* **2019**, *32*, 1121–1135. [CrossRef]
27. Minola, L.; Azorin-Molina, C.; Chen, D. Homogenization and Assessment of Observed Near-Surface Wind Speed Trends across Sweden, 1956–2013. *J. Clim.* **2016**, *29*, 7397–7415. [CrossRef]
28. Joëlsson, L.M.T.; Sturm, C.; Södling, J.; Engström, E.; Kjellström, E. Automation and evaluation of the interactive homogenization tool HOMER. *Int. J. Climatol.* **2021**, *42*, 2861–2880. [CrossRef]



29. Domonkos, P.; Guijarro, J.A.; Venema, V.; Brunet, M.; Sigró, J. Efficiency of time series homogenization: Method comparison with 12 monthly temperature test datasets. *J. Clim.* **2021**, *34*, 2877–2891. [[CrossRef](#)]
30. Killick, R.E. Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data. Ph.D. Thesis, University of Exeter, Exeter, UK, 2016. Available online: <https://ore.exeter.ac.uk/repository/handle/10871/23095> (accessed on 14 May 2022).
31. Killick, R.E.; Jolliffe, I.T.; Willett, K.M. Benchmarking the performance of homogenisation algorithms on synthetic daily temperature data. *Int. J. Climatol.* **2021**. [[CrossRef](#)]
32. Domonkos, P.; Coll, J. Time series homogenisation of large observational datasets: The impact of the number of partner series on the efficiency. *Clim. Res.* **2017**, *74*, 31–42. [[CrossRef](#)]
33. Caussinus, H.; Mestre, O. Detection and correction of artificial shifts in climate series. *JR Stat. Soc. Ser. C Appl. Stat.* **2004**, *53*, 405–425. [[CrossRef](#)]
34. Lindau, R.; Venema, V. On the reduction of trend errors by the ANOVA joint correction scheme used in homogenization of climate station records. *Int. J. Climatol.* **2018**, *38*, 5255–5271. [[CrossRef](#)]
35. Webpage. 2022. Available online: <https://www.acmant.eu> (accessed on 14 June 2022).
36. Menne, M.J.; Williams, C.N., Jr. Homogenization of temperature series via pairwise comparisons. *J. Clim.* **2009**, *22*, 1700–1717. [[CrossRef](#)]
37. Domonkos, P.; Prohom, M.; Llabrés-Brustenga, A. Removal of bias introduced by considering calendar or rainfall day as 24-hour period in daily minimum temperature series. Results from ACMANT approach. *Int. J. Climatol.* **2021**, *41/S1*, E1926–E1943. [[CrossRef](#)]
38. Menne, M.J.; Williams, C.N., Jr. Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Clim.* **2005**, *18*, 4271–4286. [[CrossRef](#)]
39. Menne, M.J.; Williams, C.N.; Gleason, B.E.; Rennie, J.J.; Lawrimore, J.H. The Global Historical Climatology Network monthly temperature dataset, version 4. *J. Clim.* **2018**, *31*, 9835–9854. [[CrossRef](#)]