On the reduction of trend errors by the ANOVA joint correction scheme used in homogenization of climate station records

Ralf Lindau | Victor Venema

Meteorological Institute, University of Bonn, Bonn, Germany

Correspondence
Ralf Lindau, Meteorological Institute, University of Bonn, Auf dem Hügel 20, D53121 Bonn, Germany.
Email: rlindau@uni-bonn.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: LI 2830/1-1, VE 366/8

Inhomogeneities in climate data are the main source of uncertainty for secular warming estimates. To reduce the influence of inhomogeneities in station data statistical homogenization compares a candidate station to its neighbours to detect and correct artificial changes in the candidate. Many studies have quantified the performance of statistical break detection tests used in this comparison. Also, full homogenization methods have been studied numerically, but correction methods by themselves have not been studied much. We analyse the ANOVA (analysis of variance) joint correction method, which is expected to be the most accurate published method. We find that, if all breaks are known, this method produces unbiased trend estimates and that in this case the uncertainty in the trend estimates is not determined by the variance of the inhomogeneities, but by the variance of the weather and measurement noise. For low signal-to-noise ratios and high numbers of breaks, the correction may also worsen the data by increasing the original random unbiased trend error. Any uncertainty in the break dates leads to a systematic undercorrection of the trend errors and in this more realistic case the variance of the inhomogeneities is also important.

KEYWORDS
adjustment, climate stations, correction scheme, homogenization, temperature trends

1 | INTRODUCTION

The main obstacle to accurate long-term trend estimates is the presence of inhomogeneities that are hidden in the data (Parker, 1994; Aguilar et al., 2003; Auer et al., 2005; Begert et al., 2005; Brohan et al., 2006; Brunetti et al., 2006; Menne et al., 2010). Important reasons for inhomogeneities are relocations of the stations and changes in their surroundings, as well as changes in the screens and instruments (Peterson et al., 1998).

Statistical homogenization algorithms aim at detecting and correcting these spurious effects by comparing a candidate station to its neighbouring reference stations. All nearby stations are assumed to observe the same regional climate signal, while the perturbations due to inhomogeneities are assumed to be different for every station (Conrad and Pollak, 1950).

Previous studies have mostly focused on the accuracy of break detection (Easterling and Peterson, 1995; Ducré-Robitaille et al., 2003; DeGaetano, 2006; Beaulieu et al., 2008; Kuglitsch et al., 2012). More recent numerical validation studies looked at the performance of both the break detection and complete homogenization algorithms (Domonkos, 2008; Domonkos, 2011; Venema et al., 2012; Williams et al., 2012; Killick, 2016; Chimani et al., 2018). The HOME benchmarking study for European climate station networks found that the performance for break detection and trend accuracy were only modestly correlated (Venema...
et al., 2012). This implies that also the other components of homogenization methods, including break correction, are important.

The performance of correction methods has hardly been studied, but Domonkos et al. (2013) found that the station trends and root-mean-square error (RMSE) of all contributions to the HOME benchmark that did not use the modern analysis of variance (ANOVA) joint correction model yet, were improved by applying this method. (One contribution was made worse due to the application of the ANOVA method, likely because the information on the break positions contained errors). The HOME benchmark did not have an explicit trend bias due to inhomogeneities, the small stochastic network-wide trend biases were hard to remove and the changes due to applying the ANOVA method more mixed (Domonkos et al., 2013).

In the HOME benchmark the error in the network temperature trends themselves were about halved by the best homogenization methods, while the error of the station trends were reduced to a quarter (Venema et al., 2012). The validation study of the Pairwise Homogenization Algorithm for U.S. climate station network did add explicit network-wide temperature trend errors due to inhomogeneities (Williams et al., 2012). For the two most realistic scenarios the remaining network trend bias is a few percent to 20%. For the hardest (unrealistic) scenario with many small biased breaks about half of the trend bias remained. Both validation studies were for high-quality well-correlated networks. The median station trend error in the Austrian validation study for relative humidity station data could be reduced by a factor of 2 due to homogenization with ACMANT and HOMER (Chimani et al., 2018). The validation data set did include an explicit trend error, but the relative humidity observations have very low correlations, the median correlation was only 0.7.

The original ANOVA method decomposes the observations of a set of climate stations into (a) a regional climate signal for all stations, (b) a break signal (step function) per station defined by the previously detected breakpoints and (c) noise (Caussinus and Mestre, 2004). The method minimizes the noise to estimate the common climate signal and the break signals. Domonkos (2017) improved this method by relaxing the assumption that all stations have the same climate signal by estimating the regional climate signal of the candidate as a squared-correlation weighted average. This gave small, but consistent improvements in his validation data sets of about 1%; for sparse networks and regions where the regional climate changes considerably in space, such as the Arctic, this new method may give larger improvements.

This study sets out to study how accurately the ANOVA joint correction model can remove trend errors in station data. Because the description of the ANOVA model in Caussinus and Mestre (2004) was short, we will detail the method in Appendix A. In section 2 on the methodology we explain how homogenization works and argue that the main task of homogenization is to get the network-wide trends right. It furthermore details how the study data was simulated and how we assess the performance of the correction algorithm.

We will analyse four scenarios: with and without a bias in the breaks that produces a systematic trend error and with and without errors in the positions (dates) of the breaks. The results for these scenarios are presented in sections 3.1–3.4. The simplest scenario of no bias and no position errors lends itself to a detailed mathematical analysis in section 3.1 where we compare our theoretical expectation of input and output errors with our empirical findings. For a signal to noise ratio of one and six breaks in each time series the input and output errors for the network-wide trend are equally large, while the number of stations is less important (section 3.1.2). For the more complex scenarios, which include also biased breaks, we show that an unbiased correction is possible if the break positions are exactly known, but that any trend bias is only partly corrected if the break positions are uncertain. The paper closes with a summary and discussions.

2 | METHODOLOGY

Statistical homogenization algorithms consist of three parts, where the first is dedicated to detecting the break positions. For this purpose, differences with neighbouring station are considered. In this way the natural variability is reduced, which would otherwise dominate the signal. Uncertainties occurring during the detection part are discussed in Lindau and Venema (2016). For pairwise methods the second step is called attribution. Here the decision is made, which of the involved stations is responsible for a detected break. Other algorithms avoid this step by using a composite reference of the surrounding network; the problem here is to produce a composite that is free of breaks. The third and last step is the correction. This step is the topic of this study. We concentrate on the ANOVA correction method, which is described in detail in Appendix A.

This paper is focused on the ability to remove linear trend errors from the data. Trends can be computed separately for each station or as regional average over the entire network. Large-scale trends are climatologically more important than station trends. Moreover, there is a much easier way to estimate the station-to-station variability of trend errors directly from the data, without running a full homogenization algorithm. The trend TRD, at a station i consists of two additive components, the spurious trend due to inhomogeneities Bi and the true climate trend C, which is the same for all stations of the network:

$$\text{TRD}_i = C + B_i$$

(1)

The average of all trend differences between a candidate station i0 and its n neighbour stations i gives a direct estimate
of the spurious break induced trend at the candidate station relative to the network mean,

\[ \frac{1}{n} \sum_{i=1}^{n} (TRD_{i0} - TRD_{i}) = \frac{1}{n} \sum_{i=1}^{n} (C + B_{i0} - C - B_{i}) = B_{i0} - \frac{1}{n} \sum_{i=1}^{n} B_{i}, \tag{2} \]

Thus, because the left hand term of Equation 2 can be computed, the trend error of each station \( B_{i0} \) is easily to infer, if the mean spurious trend of the entire network is known (i.e., the second term of the right-hand side of Equation 2). Therefore, we focus on the latter, that is, on regional averages of the inhomogeneity effects over several neighbouring stations.

The default configuration to study the performance of the ANOVA correction scheme is to simulate data for 1,000 networks consisting of \( n = 10 \) stations each. The length of each station time series is equal to \( m = 100 \) and consists of three superimposed signals.

1. The climate signal, which is identical for all stations of a network.
2. Noise added to the climate signal, which mimics the differences between the stations within a network, for example, due to measurement errors and the weather.
3. Inhomogeneities inserted at random timings and with random strength.

All three signals are randomly chosen from a normal distribution with zero mean. The first one, that is, the climate signal, can be seen as a sequence of 100 yearly means or alternatively as 100 January (or July) means, which is the typical format of temperature data to be homogenized. Such annual or monthly averages with 12-month time lag have only a low mutual dependence and can be well modelled by Gaussian white noise. The same is true for the other two signals, that is, the noise part itself and the breaks (Menne and Williams, 2005). All three signals having zero mean is justified because the correction scheme is not able to make any statement about the overall temporal mean of each station (see Appendix A).

The three standard deviations are varied, but their default values are set to \( \sigma_c = 3 \) for the climate and to \( 1 \) for the noise variance \( \sigma_n^2 \) and break variance \( \sigma_b^2 \). All three are altered later in this study to analyse their impact on the performance of the correction scheme. This is also true for the number of breaks \( k \), which we set to a default value of 5, according to the estimate of Venema et al. (2012) for Europe. In the simulation, the break positions are determined by random numbers drawn from a uniform distribution. The first break has \( m - 1 \) possible locations, the second \( m - 2 \), etc. In this way we are able to create time series with exactly five breaks each.

After creation, the temporal average of each time series is set to zero. One reason is mentioned above: the correction cannot cope with mean differences between the stations. Additionally, we are aiming at trends and these are independent from the temporal mean at each station.

### 2.1 Skill measures

As specified above, the simulated observations \( O(i,j) \), given for each station \( i \) and year \( j \), consist of three superimposed signals: the climate \( C \), the weather \( W \) and the inhomogeneities \( B \) (Equation 3). We will show in the following that the climate signal, although it is in many cases the most interesting one, is cancelled out while running the correction scheme. The inhomogeneities are the signal that has to be detected, isolated and suppressed. The weather acts as noise hampering the estimation of the break signal. Therefore, weather and noise are used in the following as synonyms:

\[ O(i,j) = C(j) + W(i,j) + B(i,j). \tag{3} \]

We are searching for the truth \( T \), that is, the observations without the inhomogeneity signal,

\[ T(i,j) = C(j) + W(i,j). \tag{4} \]

The correction scheme provides us with the detected inhomogeneity signal \( D \) and the correction is actually performed by subtracting \( D \) from \( O \). Thus, the homogenized data \( H \) is,

\[ H(i,j) = O(i,j) - D(i,j) = C(j) + W(i,j) + B(i,j) - \left( \frac{O(i,j) - D(i,j)}{R(i,j)} \right). \tag{5} \]

The most intuitive skill measure is probably the detected inhomogeneity \( D \),

\[ D(i,j) = O(i,j) - H(i,j). \tag{6a} \]

In Equation 5 the remaining inhomogeneity signal \( R \) occurs, which is in general not zero, because the correction scheme will not be perfect. Inserting Equation 4 into 3 and 5 we get for the deviations from the truth,

\[ B(i,j) = O(i,j) - T(i,j), \tag{6b} \]

\[ R(i,j) = H(i,j) - T(i,j). \tag{6c} \]

Equations 6a–6b give the three series whose statistics (e.g., mean and variance) are used to estimate the skill of the correction procedure. \( B \) is giving the deviation from the truth before the homogenization, whereas \( R \) is the deviation after the homogenization. We denote \( B \) as inserted or input error, \( D \) as detected error and \( R \) as remaining or output error. Two skill measures are used:

Measure 1: Comparison of inserted error \( B \) and detected error \( D \).

Measure 2: Comparison of inserted error \( B \) and remaining error \( R \).
The comparison is performed in the following by scatterplots, where the key parameters are the means, the variances and the correlation between $B$ and $D$ and between $B$ and $R$, respectively.

### 2.2 Analysed three steps

The main aim of this study is to analyse the ability of the ANOVA correction scheme to improve the estimates for the network-mean trends. This total process can be decomposed into three parts.

- **Step 1:** The correction itself, that is, the determination of the inhomogeneities.
- **Step 2:** The resulting correction of the trend for each station separately.
- **Step 3:** The effect of averaging the 10 station trends of each network.

Although our focus is assessing the skill of the entire three-step process as a whole, that is, which effect do inserted inhomogeneities have on the network mean trend, it helps our understanding to also show intermediate products after step 1 and 2.

### 2.3 Analysed scenarios

In the main scenario we (a) insert breaks with zero mean and (b) use the correct break positions within the ANOVA correction scheme. These two characteristics are then separately changed to biased breaks (with non-zero mean) and perturbed (intentionally worsened) break positions, so that four scenarios result:

- **Scenario 1:** Zero mean (unbiased) breaks together with correct break positions.
- **Scenario 2:** Non-zero mean (biased) breaks together with correct break positions.
- **Scenario 3:** Zero mean (unbiased) breaks together with perturbed break positions.
- **Scenario 4:** Non-zero mean (biased) breaks together with perturbed break positions.

Zero-mean breaks introduce false trends for each station and hence also a (reduced, but remaining) error for the entire network. However, on average there will be no mean trend error, but only an increased random scatter of the individual network trends. This will be different when biased (non-zero) breaks are inserted in (scenarios 2 and 4). In these cases an overall effect of the breaks on the trend is expected.

In our study the correct break positions are known, because we use artificial data. Because we focus on the last step of the homogenization procedure, the correction itself, it is justified to use these true breakpoint positions to isolate the performance of the correction scheme alone. However, in reality the break positions will not be perfectly determined. The additional impact of these position errors are analysed in scenarios 3 and 4, by adding random perturbations to the true break positions.

Most experiment results are given by scatterplots (Figures 1–8, 14–16), where the inserted error (input) is given on the abscissa and the detected or remaining error (output) is given on the ordinate. In each figure the six key parameters of any scatterplot (mean and variance of both axes plus correlation coefficient and number of observations) are given by $\bar{x}, \sigma_x^2, \bar{y}, \sigma_y^2, r$ and $N$, respectively. These six parameters are also given as a summary in Table 1.
3 | RESULTS

3.1 | Scenario 1: Unbiased breaks, correct break positions

In this scenario we assume both no overall trend bias and perfect break detection. The first study is very simplistic: We set the noise to zero and check the result after step 1. Figure 1 shows the inserted inhomogeneities for each year, station and network on the abscissa against the output of the correction scheme on the ordinate. Actually, 1,000,000 crosses should appear, but for technical reasons we displayed only a subset. However, the statistics is based on the full data set. The test yields a perfect correlation between the input inhomogeneities and the detected ones. Thus, the correction scheme works perfectly, if no noise between the stations is present and if the break positions are known. The perfect result is not completely trivial, as only the break positions, but not their sizes are used as input of the algorithm.

We repeat this study with different values for $\sigma_c$ and found no differences in the result. This is expected. In the correction scheme, only differences of overlapping time
periods between two stations of the same network are considered (Appendix A). Consequently, the climate signal, which is assumed to be the same at every station of a network, is completely cancelled out. Thus, the climate signal, while always generated in our simulations, is not relevant and not further considered.

The second study is conducted with non-vanishing noise variance $\sigma_n^2 = 1$. In Figures 2–4 the results for all three processing steps (as defined in section 2.2) are presented. The three panels show on the $x$-axis the inserted quantity and on the $y$-axis the detected one (i.e., measure 1). In Figure 2 the situation after applying the correction scheme is shown. Both means are zero, the detected variance is with 0.786 slightly higher than that for the inserted one (0.719). The correlation remains rather large with 0.955. In Figure 3 the resulting linear station trends (for 10 station times 1,000 networks) are given. These trends are calculated by linear least squares regression. Temperature trends normally have the dimension “Kelvin per time.” However, here we multiplied it with the total length of the time series so that the trend expresses the total temperature change during the considered time period (explainable by a linear trend). Compared to Figure 2, the variances of input and output both increase by a factor of about 3, while the correlation is only slightly decreased. In the third step the 10 station trends of each network are averaged. By the averaging process the variances are both reduced, but that of the detected trend less than that of the inserted one (0.133, but in the same order of magnitude). On both axes, we consider trends of inhomogeneities and no real climate trends. Therefore, we are dealing with the uncertainty introduced by the inhomogeneities on the trends. The interpretation of the found numbers is as following. If both the

$$y_j = x_j + \epsilon_j.$$  \hspace{1cm} (7)

If Equation 7 is true, the slope $a$ of the regression line is equal to 1. A short rational is given by the following equation chain, where we used Equation 7 at the forth equal sign:

$$a = r \frac{\sigma_y}{\sigma_x} = \frac{\text{cov}}{\sigma_x^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} (x_i + \epsilon_i) y_i}{\sum_{i=1}^{n} x_i^2} \approx \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} = 1,$$  \hspace{1cm} (8)

where two variables $x$ and $y$ comprising $n$ values with zero mean are considered. The terms $r$ and $\text{cov}$ denote their correlation and covariance, respectively. The standard deviations are referred to as $\sigma_x$ and $\sigma_y$. The expectation of the product of two random zero-mean variables is zero, which justifies the nearly equal sign between the fifth and sixth term.

For the above discussed reason, it is convenient to display not $y$ (the detected error $D_i$), but the difference $x - y$ (the remaining error $R$) on the ordinate (Figures 5–7). The correlation is negligible in all cases, which confirms that Equation 7 is valid here. The variable on the $x$-axis (the inserted error $B$) is unchanged compared to Figures 2–4 so that its statistics remains of course the same. The key parameters are the variances of inserted and remaining quantities $B$ and $R$.

Figure 7 shows an inserted network-mean trend variance of 0.265, the remaining one after correction is smaller with 0.133, but in the same order of magnitude. On both axes, we consider trends of inhomogeneities and no real climate trends. Therefore, we are dealing with the uncertainty introduced by the inhomogeneities on the trends. The interpretation of the found numbers is as following. If both the

![FIGURE 7](image1.png)

**FIGURE 7** As Figure 4, but for the remaining network-mean trends instead of the detected ones. This is the result for scenario 1 assuming unbiased break sizes and correct positions

![FIGURE 8](image2.png)

**FIGURE 8** As Figure 7, but the standard deviation of the inserted breaks is increased from 1 to 2, while the standard deviation of the noise remains at 1
standard deviation of noise and that of breaks are equal to 1 K, the existing inhomogeneities make it a priori rather difficult to determine the network-mean trend accurately. The x-axis variance can be referred to as the input trend error variance, and its square root is as large as \( f = 0.515 \) K. Thus, the secular trend of this small network is in same order of magnitude as the climate trend itself. The y-axis variance can be interpreted as output trend error variance. The corresponding uncertainty of the network-mean trend after correction is with \( g = 0.365 \) K comparable in size.

In the first study (Figure 1), we have shown already that the climate signal is irrelevant for the correction and that it works perfectly under the absence of noise. From this, it is obvious that the output error \( R \) depends only on the noise variance. The input error is nothing else than the trend inserted by inhomogeneities. Therefore, it is clear that the input error depends only on the break variance.

However, to show it formally, we vary the noise and break variance in the next test. First the standard deviation of the inserted breaks is increased to \( \sigma_b = 2 \) (with \( \sigma_n = 1 \)). As consequence, the input error increases by a factor of 2, while the output error remains unchanged (Figure 8). If vice versa the noise is set to \( \sigma_n = 2 \) (with \( \sigma_b = 1 \)), the output error is doubled, while no change in the input error is observed (Table 1). Further studies with various combinations of \( \sigma_b \) and \( \sigma_n \) confirm that (a) the input error depends only on the break variance and the output error only on the noise variance and that (b) the relationships are both linear.

Consequently, we have two separated processes as illustrated by the flowchart given in Figure 9. On the one hand, there is the conversion chain from the initial break variance (state in0) down to the error variance of the inserted network trends (state in3). On the other hand there is the transformation from the initial noise variance (state out0) down to the error variance of the remaining network trends (state out3). The first conversion chain considers input parameters beginning with the break variance \( \sigma_b^2 \) and ending with the trend variability before the correction \( \sigma_{n_1}^2 \). The second chain deals with output parameters. It starts from the noise variance \( \sigma_n^2 \), which finally defines the remaining trend variability \( \sigma_{out}^2 \) after the correction. Three factors for each chain, \( f_1^2 \) to \( f_3^2 \) and \( g_1^2 \) to \( g_3^2 \), denote the conversions of the variances from step to step. The total conversion factors are given by the product of the three partial factors:

\[
\begin{align*}
    f^2 &= f_1^2 \times f_2^2 \times f_3^2, \\
    g^2 &= g_1^2 \times g_2^2 \times g_3^2.
\end{align*}
\]

Let us first consider the numerical values of the factors \( f^2 \) and \( g^2 \) as whole and coming to the detailed analysis of the partial factors later in section 3.1.1. In our experiments we set the initial variances \( \sigma_b^2 \) and \( \sigma_n^2 \) to 1 so that the factors \( f^2 \) and \( g^2 \) are directly visible as the two variances in Figure 7, which are 0.265 for the input and 0.133 for the output. Their square roots give the conversion factor for the standard deviation, which describe the errors.

Thus, in this case the input error \( \sigma_{in} \) and the output error \( \sigma_{out} \) defined as the initial and the remaining standard deviation of the network-mean trend are given by

\[
\begin{align*}
    \sigma_{in} &= f \sigma_b \quad \text{with} \quad f = \sqrt{0.265} = 0.515, \\
    \sigma_{out} &= g \sigma_n \quad \text{with} \quad g = \sqrt{0.133} = 0.365.
\end{align*}
\]

The fraction of input and output error is then
\[
\sigma_{\text{in}} = f \frac{\sigma_b}{g} = \frac{f}{g} \ \text{SNR},
\]
with SNR denoting the signal-to-noise ratio of the analysed data. From Equation 12 we see that, for the considered settings, the error is enlarged by the correction, if \( \text{SNR} < \frac{f}{g} = 0.71 \).

For unbiased breaks (mean zero jump size) and perfect detection the correction scheme is reducing the error variance of the network-mean trend from 0.265 to 0.133. This improvement is obtained for networks of \( n = 10 \) stations with \( k = 5 \) breaks each and a \( \text{SNR} = 1 \). However, any alteration of these three parameters has the potential to change the obtained result. The effect of changes in \( n \) and \( k \) are discussed in section 3.1.2. However, for pure changes in the SNR the situation is clear. For \( n = 10 \) and \( k = 5 \), the correction scheme enlarges the error of the network trend if the SNR is smaller than 0.71.

### 3.1.1 Theoretical considerations for the factors \( f \) and \( g \)

In the following we will estimate the factors \( f \) and \( g \) theoretically to confirm their numerical values determined so far only empirically; see Equations 10 and 11. First, we estimate the input factor \( f \) by considering the partial factors \( f_1 \) to \( f_3 \). It is based on \( \sigma_b \), the standard deviation of the break signal.

Figure 2 shows that the variance is reduced from state \( \text{in}0 \) to \( \text{in}1 \) by a factor of \( f_1^2 = 0.719 \). The numerical value of \( f_1^2 \) is discussed in the following. Each of the simulated time series contain five breaks, that is, a step function with six sub-periods of arbitrary lengths. Their step heights are chosen from a standard normal distribution. As mentioned above, we decided to set the mean of each of these simulated time series to zero. Hereby, a fraction of \( 1/m_n \) of the input variance is lost, where \( m_n \) is defined as the number of independents. In our case, a rough estimate is \( m_n = 6 \), as this is the number of sub-periods. However, the sub-periods have different lengths and this reduces the effective number of independents. A time series with one long and five very short sub-periods will behave more as if it contains only one rather than six independents.

Thus, from the obtained variance of inserted inhomogeneities (0.719), we can conclude that the effective number of independents must be approximately 3.5, because

\[
f_1^2 = 1 - \frac{1}{m_n} \approx 0.719 \Rightarrow m_n \approx 3.5,
\]

In Appendix B we show that a good approximation for \( m_n \) is

\[
m_n = \frac{k}{2} + 1,
\]
where \( k \) is the number of breaks. As we applied \( k = 5 \) in the simulation, Equations 13 and 14 are in good agreement.

In step 2 (from state \( \text{in}1 \) to \( \text{in}2 \)), the transition from the variance of the inserted inhomogeneities (Figure 5) to the error variance of the inserted trend (Figure 6) is made. The trend is equal to the slope of the regression line and its error variance \( \sigma_{\text{slp}}^2 \) is:

\[
\sigma_{\text{slp}}^2 = \frac{\sigma_y^2(1-r^2)}{\sigma_{\text{in}}^2(m_{\text{in}}-2)} = \frac{\sigma_y^2}{\frac{m_{\text{in}}}{12} m_{\text{in}}^{-2}},
\]
where \( \sigma_y^2(1-r^2) \) denotes the variance unexplained by the trend, \( \sigma_{\text{in}}^2 \) the variance of the entire considered time period (which is equal to \( m^2/12 \)), and \( m_{\text{in}} \) the number of independent observations. The approximation made in Equation 15, is essentially justified because the trend typically explains only a small part of the total variance; and neglecting the small part it explains anyhow is further compensated by replacing the term \( m_{\text{in}}-2 \) by \( m_{\text{in}} \). In this study we treat trends as temperature changes over the entire period. Accordingly, errors of the trend are given by the product \( m \sigma_{\text{slp}} \) so that the factor \( f_2^2 \) is given by

\[
f_2^2 = \frac{(m \sigma_{\text{slp}})^2}{\sigma_{\text{in}}^2} \approx \frac{12}{m_{\text{in}}},
\]

With \( m_{\text{in}} = 3.5 \) it follows a theoretical estimate of \( f_2^2 = 3.4 \). This value is nearly perfectly obtained in the simulation (the quotient of input variances in Figures 5 and 6 is 2.354/0.719 = 3.3).

In step 3 (from state \( \text{in}2 \) to \( \text{in}3 \)), the 10 station trends of each network are averaged. As they can be regarded as independent, we expect a variance reduction by a factor of 10. The transition from Figure 6 (with an input variance of 2.354) to Figure 7 (with an input variance of 0.265) confirms this approximately.

In three steps, we have retraced the development of the input error variance \( \sigma_{\text{in}}^2 \) beginning with the initial break variance \( \sigma_b^2 \). As these are connected by the factor \( f^2 \) (Equation 9a), we have implicitly deduced a more general estimate for \( f^2 \) which depends on two parameters, the number of breaks \( k \) and the number of stations \( n \) within a network

\[
f^2 = f^2(k, n) = f_1^2(k) f_2^2(k) f_3^2(n) = \frac{m_{\text{in}} - 1}{m_{\text{in}}} \frac{12}{m_{\text{in}}} \frac{1}{n}
\]

\[
= \frac{1}{2} + \frac{12}{2} + \frac{1}{n} \frac{24 k}{n (k + 2)^2}.
\]

In the following the factor \( g \) is discussed. This is more complicated as here the matrix solving process performed within the correction scheme has to be considered. We start with the noise variance \( \sigma_n^2 \), which triggers errors in the determination of the inhomogeneities in the correction scheme. In Appendix A, we show that an estimate for the error variance of the inhomogeneities for known break positions is \( k \sigma_n^2 (m - 1) \). In our case with \( k = 5 \), \( \sigma_n^2 = 1 \) and \( m \).
= 100, this is equal to about 0.05. The simulation in Figure 5 shows an only slightly larger variance of 0.069 for the residuals, which roughly confirms our theoretical estimation. The transition from this initial variance (state out1) to the station trends variance (state out2) performs rather similar as found for the respective input parameters. The comparison of Figures 5 and 6 shows that the obtained trend variance is again larger by a factor of about 3.35 (from 0.069 to 0.231). However, in the next step, that from station (state out2) to network trends (state out3), the output error variances behave rather different. They are only reduced from 0.231 to 0.133. Although we averaged again over 10 stations, the variance is only reduced by a factor of less than 2. The reason is that the used correction scheme is a regression method and such methods have the general property of producing dependent solutions. The 10 station trends in each network are highly correlated. Consequently, we are not able to reduce the uncertainty significantly by averaging.

An exact equation for \( g^2 \) (analogue to Equation 17) is difficult to give here. However, \( g^2 \) consists analogously of three sub-factors. The first one, \( g_1^2 \), determines the variance of inhomogeneities, and is in the order of 10 times smaller than \( f_1^2 \), because \((m_{in} - 1)/m_{in}\) is in the order of 1 and \( k/(m - 1) \) is in the order of 1/10. The transition to station trends \((f_2^2 \text{ and } g_2^2, \text{ respectively})\) is similar for input and output quantities. For the third factor the circumstances are reversed: The variance reduction for the output \( g_3^2 \) is in the order of 1, whereas that of the input \( f_3^2 \) is 1/n, thus 10 times smaller. Consequently, the entire effect of all three sub-factors together is comparable. And this is the reason why \( f \) and \( g \) have the same order of magnitude for \( m = 100, k = 5 \text{ and } n = 10 \).

3.1.2 The variation of the factors \( f \) and \( g \) with break and station numbers

However, both factors \( f \) and \( g \) are expected to depend on the number of breaks \( k \) and the number of stations \( n \). We performed a fourth study and varied both, the number of breaks \( k \) between 2 and 9, and the number of stations \( n \) between 5 and 10. Figure 10 shows the result for the factor \( f \), which gives the translation factor from the standard deviation of the break heights to the resulting error of the network trend before correction. For the studied \( k-n \) combinations, \( f \) varies between 0.41 and 0.74. For \( f(5,10) \) we obtained 0.50. The slight difference to Equation 10 (giving 0.51) is within the noise. Isolines are drawn for the obtained values, which indicate the growth of \( f \) with decreasing break and station number, respectively. The theoretical estimate from Equation 17 is given by the fat isolines. It shows that theory and numerical results are in good agreement.

Figure 11 gives the corresponding result for the factor \( g \). It grows with increasing number of breaks, because the underlying uncertainties of the inhomogeneities do so. The dependence on the number of stations is, at least for small break numbers, less important. Figure 12 shows the \( k-n \) dependence of the ratio \( g/f \). Here, the dependence on break number dominates by far; the dependence on station number, which is found for the two factors separately, is largely cancelled out. The ratio \( g/f \) grows rather linearly with \( k \) and can be approximated by

$$ g = \frac{k}{f} = \frac{k}{6}. \quad (18) $$

Inserting Equation 18 into 12, we find

![Figure 9](image-url) Flowchart for the input (left) and output (right) variables as they are given in Figure 5 on the x and y-axis, respectively. The transitions of the input states in0 to in3 are given by the factors \( f_1^2 \) to \( f_3^2 \), that of the output states by the factors \( g_1^2 \) to \( g_3^2 \).

![Figure 10](image-url) The factor \( f \) as a function of break and station number. Three isolines are drawn for \( f = 0.5, 0.6 \) and 0.7. The theoretical values from Equation are given as fat isolines for the same three values.
If the SNR is smaller than $k/6$ the output error is larger than the input error, which means that the correction scheme makes the trend uncertainty larger.

In Figure 7 we compared the trend uncertainties of the uncorrected and corrected data. The correction procedure replaces the initial error of a certain network by a different one, which is uncorrelated ($r = 0.02$) and smaller, but comparable in size. Equation 19 defines the threshold when it will become larger. However, already below this threshold the corrections produce dependent solutions within each network. But independence is an important feature in data analysis and homogenized data may not be useful whenever variability shall be investigated.

---

**FIGURE 11** The factor $g$ as a function of break and station number. Isolines are drawn for $g = 0.2$ to 1.4 in steps of 0.1.

\[
\frac{\sigma_{\text{in}}}{\sigma_{\text{out}}} = \frac{6}{k} \text{SNR}.
\] (19)

**FIGURE 12** The factor $g/f$ as a function of break and station number. Isolines are drawn for $g/f = 0.4$ to 2.0 in steps of 0.1. The line $g/f = 1$ is drawn fat.

**FIGURE 13** The mean inserted inhomogeneity for a bias according to Equation is given by the fat line. Because of edge effects the introduced trend is decreased from 1 to 0.897 K/cty. If additionally random breaks with $\sigma_b = 1$ are included, the actually inserted trend is scattered and differs slightly with 0.873 K/cty.

**FIGURE 14** As Figure 7, but with an additionally inserted trend bias of 0.873 K/cty as depicted in Figure 13. This is the result for scenario 2 assuming biased break sizes and correct positions.
3.2 Biased breaks, correct break positions

So far, we did not take into account that the inhomogeneities may have a mean network-wide non-zero trend. Such an effect may occur in reality, for example, for a transition to Stevenson screens or due to the introduction of new instruments. To simulate this situation we add a fourth signal to the artificial data. The homogeneous sub-periods of the break signal are shifted by $\Delta B$ upwards or downwards depending on the middlemost year $j_m$ of the sub-period. Early inhomogeneities are shifted downwards, late ones upwards. In this way, a global mean trend bias is inserted into the data,

$$\Delta B = 0.01 \left( j_m - 50.5 \right)$$

where the middlemost year $j_m$ may vary between 1 and 100. By applying Equation 20, one may expect a resulting trend bias of 1 K/cty. However, edge effects lead to a reduced linear trend bias so that only 0.897 K/cty is resulting (Figure 13, fat line). Nevertheless, the inserted inhomogeneities still consist of two components: the discussed bias plus a random component. The latter adds noise to the data, which leads to random variations of the mean inhomogeneities and therefore to an uncertainty of the actually inserted trend bias. Figure 13 (crosses) shows the resulting mean situation for all inhomogeneities. The actually inserted mean trend is slightly different with 0.873 K/cty.

The important question is of course, whether the correction scheme is able to identify and remove this mean trend bias. Figure 14 shows that the answer is yes. The artificial systematic trend of 0.873 K/cty is almost entirely removed. The mean remaining trend error is as small as 0.001 K/cty.

A comparison of Figure 14 with Figure 7, where no trend bias exists, shows that the variation of individual network trend errors remains the same on both axes. The data cloud in Figure 14 is solely shifted to the right on the abscissa. Thus, the uncertainty to determine an individual network trend is still high, but the important issue is the successful removal of the trend bias.

3.3 Unbiased breaks, scattered break positions

Up to now, we used the known break positions of the simulated data assuming hereby perfect break detection. In the following, we use slightly scattered break positions as a simple way to study the effect of unavoidable errors. For this purpose, the true break positions are shifted by random time spans based on Gaussian noise of a certain standard deviation $\sigma_s$. To conserve the original number of inserted breaks (five), we mirror any shift that would land outside the allowed time period (1–100). Figure 15 shows the result for $\sigma_s = 1$. Please note that the chosen strength of scatter is still moderate: about 40% of the positions are in this case not altered at all, and about 50% are shifted by plus or minus one temporal item. The main effect of the now included break position errors is that input errors $B$ on the abscissa and output errors on the ordinate are slightly correlated (please compare Figures 7 and 15). Assuming error-free break positions $D$ and $R$ were uncorrelated, while their correlation is now equal to $r = 0.143$. Additionally, the output error is slightly increased from 0.133 to 0.162. In a further experiment, we doubled the break position errors to a standard deviation of 2 (Table 1). In this case the correlation is increased from $r = 0.143$ to $r = 0.218$. The correlations found between inserted and remaining errors show that the errors are only partly corrected.
3.4 | Biased breaks, scattered break positions

Finally, we considered both together biased break sizes and scattered break positions. Figure 16 shows again the scatter-plot for inserted and remaining network trend errors for $\sigma_s = 1$. However, now the inserted breaks are additionally biased. Thus, concerning the effect of the bias Figure 14 is the appropriate reference, while it is Figure 15 for the effects of the position errors. Concerning the latter, input and output errors are again correlated with $r = 0.147$, which is rather similar to Figure 15, where we found $r = 0.143$. However, more important is the ability of the correction scheme to remove overall trend biases. While it was possible to correct the trend bias nearly entirely, when the break positions are exactly known (Figure 14), the mean trend bias of 0.873 K is now removed only partly (Figure 16), 0.093 K (about 11%) remains. Increasing $\sigma_s$ to 2 (Table 1) provides an almost doubled remaining bias (0.180 K = 21%).

4 | SUMMARY AND DISCUSSION

The performance of the ANOVA correction method is studied with simulated data. A reasonable skill measure is the improvement compared to uncorrected data. Therefore, also the input errors, which are caused by the inhomogeneities, have to be quantified. We divide this break signal into a random part, characterized by the break variance (exclusively considered in scenarios 1 and 3), and a deterministic part, the trend bias, which is a constant non-zero trend in all time series (additionally considered in scenarios 2 and 4). Obviously, the latter is really harmful as it would falsify the mean trend, whereas the first merely increases the uncertainty.

For the output side, we showed that, apart from the number of breaks, two parameters are essential. First the noise (weather) variance, being the difference between the true station signal and the regional climate signal (Equation 4); and second the break position errors that may occur in the preceding detection part of the homogenization algorithm. Note, that the climate signal itself is completely cancelled out by the correction method and has no influence on the result.

In the basic scenario we assumed perfect detection and no trend bias. In this simplistic case only the break variance (on the input side) and the noise variance (on the output side) plays a role, when the number of breaks is held constant. We showed that the input error depends only on the break variance, while the output error depends only on the noise variance. Input and output errors are independent from each other, both with zero mean. Thus, the breaks introduce an unbiased random trend error to each network and the homogenization replaces this input error by a different and independent, but also random and unbiased error after the homogenization process. The strength of this output error depends on the noise. For the basic scenario we showed that under realistic circumstances, that is, six breaks in each station time series and a signal-to-noise ratio of 1, input and output trend uncertainties are equally large. If the noise becomes larger than the break variance (SNR < 1) and/or more than six breaks are hidden in the time series, the homogenization may also worsen the data by increasing the original random unbiased trend error.

Moreover, after correction the trend errors for the different stations of a network are mutually dependent. Statistical analyses based on these homogenized station data, which mistakenly assume independence, may conclude that false trends are significant, because the trends vary so little from station to station.

In this study we assumed that the breaks are noisy deviations from a baseline. However, many homogenization studies have assumed that the break signal is a random walk. Reality seems to lie in the middle for European networks (Venema et al., 2012). For the same break size distribution a random walk signal has a larger total variance. It could thus be that the above numbers are somewhat different for a random walk.

However, the strength of the ANOVA method became apparent, when we added a trend bias to the input data. An important finding is that a trend bias is almost perfectly corrected if the break positions are known. As mentioned above, the correction of possible trend biases is very important.

The focus of the study is the performance of the correction method in case of perfect knowledge about the breaks, but in practice the derived breaks will have flaws as well. In future it will be worthwhile to study the interactions between detection and correction in detail. To get a first glimpse of how the ANOVA method handles errors in the breaks we considered the impact of position errors in the input of the correction scheme. In a more realistic scenario the position error would be a function of break size (Lindau and Venema, 2016) and there would be missing and spurious breaks.

As expected for a regression method, the ANOVA method is only able to correct a part of the trend bias. For moderate perturbations of the break positions ($\sigma_s = 2$ time steps), 21% of the introduced trend bias remained in the corrected data.

In the real world, the break detection will not be perfect so that position errors are unavoidable. Consequently, we expect that any trend correction performed so far tends to be underestimated. This fits to the results of the validation study for the US network by Williams et al. (2012), where the algorithm was able to reduce trend errors, but some trend bias remained. In the most difficult case with 10 breaks per century half of the trend error remained. Formulated positively, also for this very difficult case homogenization was able to reduce the large-scale trend bias.

The correct climate trend for each network is a by-product of the ANOVA decomposition. For users who are exclusively interested in this specific parameter, it would be appropriate to make the estimate of the climate trend directly available to the community. The detour over the determination of the inhomogeneities, the correction of the station
data, the calculation of the station trends from this corrected data, and the final averaging over all stations of network could be avoided.

ACKNOWLEDGEMENTS
This study was supported by the Deutsche Forschungsgemeinschaft, by the grants LI 2830/1-1 and VE 366/8.

ORCID
Ralf Lindau https://orcid.org/0000-0003-4963-4402
Victor Venema https://orcid.org/0000-0001-6902-4869

REFERENCES
Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa, J. (2003) Guidelines on Climate Metadata and Homogenization. Geneva: World Meteorological Organization, WMO-TD No. 1186, WCDMP No. 53, 55 pp.
Auer, I., Böhm, R., Jurkovic, A., Orlík, A., Potzmann, R., Schöner, W., Ungerstöck, M., Brunetti, M., Nanni, T., Maugeri, M., Briffa, K., Jones, P., Etyhymiadis, D., Mestre, O., Moisselin, J.-M., Begert, M., Brazdil, R., Bochnicek, O., Cegnar, T., Gagic-Capka, M., Zannovic, K., Majstorovicz, Z., Szalai, S., Szentrémey, T. and Mercalli, L. (2005) A new instrumental precipitation dataset for the Greater Alpine region for the period 1800–2002. International Journal of Climatology, 25, 139–166.
Beaulieu, C., Seidou, O., Ouarda, T.B.M.J., Zhang, X., Boulet, G. and Yagouti, A. (2008) Intercomparison of homogenization techniques for precipitation data. Water Resources Research, 44, W02425. https://doi.org/10.1029/2006WR005615.
Begert, M., Schlegel, T. and Kirchhofer, W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. International Journal of Climatology, 25, 65–80.
Brohan, P., Kennedy, J.J., Harris, I., Tett, S.F.B. and Jones, P.D. (2006) Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. Journal of Geophysical Research, 111, D12106. https://doi.org/10.1029/2005JD006548.
Brunetti, M., Maugeri, M., Monti, F. and 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.
Causinsson, H. and Mestre, O. (2004) Detection and correction of artificial shifts in climate series. Journal of the Royal Statistical Society: Series C, 53, 405–425. https://doi.org/10.1111/j.1467-9876.2004.05155.x.
Chimani, B., Venema, V., Lexer, A., Andre, K., Auer, I. and Nemec, J. (2018) Inter-comparison of methods to homogenize daily relative humidity. International Journal of Climatology (2018), 38, 3106–3122. https://doi.org/10.1002/joc.5488.
Conrad, V. and Pollák, C. (1950) Methods in Climatology. Cambridge, MA: Harvard University Press, 459 pp.
DeGaetano, A.T. (2006) Attributes of several methods for detecting discontinuities in mean temperature series. Journal of Climate, 19, 838–853.
Domonkos, P. (2008) Testing of homogenization methods: purposes, tools and problems of implementation. In: Lakatos, M., Szentrémey, T., Bihari, Z. and Szalai, S. (Eds.) Proceedings of the 5th Seminar and Quality Control in Climatological Databases. Geneva: World Meteorological Organization, WCDMP No. 71, WMO/TD-No. 1493, pp. 126–145.
Domonkos, P. (2011) Efficiency evaluation for detecting inhomogeneities by objective homogenization methods. Theoretical and Applied Climatology, 105, 455–467. https://doi.org/10.1007/s00704-011-0399-7.
Domonkos, P. (2017) Time series homogenization with optimal segmentation and ANOVA correction: past, present and future. In: Proceedings of the Ninth Seminar for Homogenization and Quality Control in Climatological Databases and Fourth Conference on Spatial Interpolation Techniques in Climatology and Meteorology. Geneva: World Meteorological Organization, WCDMP-No. 85.
Domonkos, P., Venema, V. and Mestre, O. (2013) Efficiencies of homogenization methods: our present knowledge and its limitation. In: Proceedings of the Seventh Seminar for Homogenization and Quality Control in Climatological Databases. Geneva: World Meteorological Organization, WCDMP-No. 78, pp. 11–24.
Ducré-Robailleau, J.-F., Vincent, L.A. and Boulet, G. (2003) Comparison of techniques for detection of discontinuities in temperature series. International Journal of Climatology, 23, 1087–1101.
Easterling, D.R. and Peterson, T.C. (1995) A new method for detecting undocumented discontinuities in climatological time series. International Journal of Climatology, 15, 369–377.
Killick, R. (2016) Benchmarking the performance of homogenization algorithms on daily temperature data. PhD Thesis, Open Research Exeter, University of Exeter.
Kuglitsch, F.G., Auchmann, R., Bleisch, R., Brönnimann, S., Martius, O. and Stewart, M. (2012) Break detection of annual Swiss temperature series. Journal of Geophysical Research, 117, D13105. https://doi.org/10.1029/2012JD017729.
Lindau, R. and Venema, V.K.C. (2013) On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records. Idiljasor, 117, 1–34.
Lindau, R. and Venema, V.K.C. (2016) The uncertainty of break positions detected by homogenization algorithms in climate records. International Journal of Climatology, 36(2), 576–589. https://doi.org/10.1002/joc.4366.
Menne, M.J. and Williams, C.N. (2005) Detection of undocumented Change-points using multiple test statistics and composite reference series. Journal of Climate, 18, 4271–4286. https://doi.org/10.1175/JCLI3524.1.
Menne, M.J., Williams, C.N., Jr. and Palecki, M.A. (2010) On the reliability of the US surface temperature record. Journal of Geophysical Research: Atmospheres, 115, D11108. https://doi.org/10.1029/2009JD013094.
Parker, D.E. (1994) Effects of changing exposure of thermometers at land stations. International Journal of Climatology, 14, 1–31.
Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentrémey, T., Salingeri, J., Forland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. and Parker, D. (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18, 1493–1517.
Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, A.J., Domonkos, P., Vertacnik, G., Szentrémey, T., Stepneak, P., Zahradnìcek, P., Viare, J., Muller-Westerner, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kalokythas, K., Marinova, T., Andresen, L., Acquavotta, F., Fratianne, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P. and Brandsma, T. (2012) Benchmarking homogenization algorithms for monthly data. Climate of the Past, 8, 89–115.
Williams, C.N., Menne, M.J. and Thorne, P.W. (2012) Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. Journal of Geophysical Research, 117, D05116. https://doi.org/10.1029/2011JD016761.

How to cite this article: 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. https://doi.org/10.1002/joc.5728
APPENDIX: CORRECTING INHOMOGENEITIES IN CLIMATE SERIES WITH AN ANOVA-TYPE METHOD

Consider an \( n \times m \) matrix of observations \( O(i,j) \) consisting of \( i = 1 \) to \( n \) time series with length \( j = 1 \) to \( m \). The observations are fit to a linear model given by the two variables \( C(j) \) and \( B(i,j) \), where \( C(j) \) denotes the time dependent climate signal, which is considered to be identical at each station. The second variable \( B(i,j) \) is the inhomogeneity for each station and year, which is considered to be constant over each homogenous sub-period (HSP) \( h = 1 \) to \( H \) where \( H \) denotes the number of HSPs within the entire network. The position of the break points are known so that we can address a specific inhomogeneity, \( B(i,j) \), alternatively as one-dimensional vector \( B(h) \), with known \( j_1(h) \) and \( j_2(h) \) denoting the first and last year of the inhomogeneity, and \( i_0(h) \) denoting the station it belongs to.

The observations are equal to a sum of three terms: The climate signal, the station inhomogeneity, and a random noise variable \( \epsilon(i,j) \):

\[
O(i,j) = C(j) + B(i,j) + \epsilon(i,j). \tag{A1}
\]

The noise, that is, the square difference between model and observation, is minimized,

\[
\sum_{i=1}^{n} \sum_{j=1}^{m} (C(j) + B(i,j) - O(i,j))^2 = \text{min}. \tag{A2}
\]

\( C(j) \) and \( B(i,j) \) are unknowns to be determined. We perform the derivation of Equation A2 with respect to all of these unknowns and obtain \( m + H \) equations for an equal number of unknowns. The derivations with respect to a fixed, but arbitrary \( C(j) \) yield

\[
2 \sum_{i=1}^{n} (C(j) + B(i,j) - O(i,j)) = 0, \tag{A3}
\]

which can be transformed to:

\[
C(j) = \frac{1}{n} \sum_{i=1}^{n} (O(i,j) - B(i,j)). \tag{A4}
\]

In this way we have a number of \( m \) equations, one for each \( C(j) \). Analogously, we derive Equation A2 with respect to the inhomogeneities \( B(h) \). For a given, but arbitrary \( B(h) \), it follows:

\[
2 \sum_{j=1}^{j_2(h)} (C(j) + B(h) - O(i_0(h),j)) = 0, \tag{A5}
\]

where \( j_1(h) \) and \( j_2(h) \) denote the first and the last year of the considered HSP with length \( l(h) \) and \( i_0(h) \) the station it belongs to. Analogously to Equation A4, we can transform and solve for \( B(h) \),

\[
B(h) = \frac{1}{l(h)} \sum_{j=j_1(h)}^{j_2(h)} (O(i_0(h),j) - C(j)). \tag{A6}
\]

We obtain \( H \) equations, one for each of the inhomogeneities of the whole station network. According to Equation A6, a specific \( B(h) \) is equal to the mean difference between the observations of the considered station \( i_0 \) and the climate signal \( C \) averaged over the duration of the HSP. The total number of \( m + H \) equations (Equations A4 and A6) can be reduced in two ways. Either we insert Equation A6 into A4 and obtain \( m \) equations for the climate signal \( C(j) \); or we insert vice versa Equation A4 into A6 and obtain \( H \) equations for the inhomogeneities \( B(h) \). The latter alternative yields

\[
B(h) = \frac{1}{l(h)} \sum_{j=j_1(h)}^{j_2(h)} (O(i_0(h),j) - \frac{1}{n} \sum_{i=1}^{n} (O(i,j) - B(i,j))). \tag{A7}
\]

Finally, we separate the inhomogeneities \( B \) from the observations \( O \),

\[
B(h) = \frac{1}{n l(h)} \sum_{j=j_1(h)}^{j_2(h)} \sum_{i=1}^{n} B(i,j) = \frac{1}{l(h)} \sum_{j=j_1(h)}^{j_2(h)} O(i_0(h),j) - \frac{1}{n l(h)} \sum_{j=j_1(h)}^{j_2(h)} \sum_{i=1}^{n} O(i,j). \tag{A8}
\]

In Equation A8 we consider a specific inhomogeneity \( B(h) \) occurring at station \( i_0 \) during the period \( j_1 \) to \( j_2 \). The double summation on the left-hand side gives the average over the inhomogeneities of all stations in that period. The difference between the candidate and the average of all stations is relevant. An analogous difference stands on the right-hand side of Equation A8. This expression is known and can be calculated from the observations. The interpretation of Equation A8 is as following: Inhomogeneities are detectable by anomalies in the observations compared to the average of neighbour stations during the same period. As always differences of identical periods are considered, the climate effect is cancelled out. However, the inhomogeneities are still mutually dependent.

We multiply Equation A8 by \( n \) and \( l(h) \) to avoid fractional numbers in the following calculations:

\[
n l(h) B(h) = \sum_{j=j_1(h)}^{j_2(h)} \sum_{i=1}^{n} (B(i,j)) = n \sum_{j=j_1(h)}^{j_2(h)} O(i_0(h),j) - \sum_{j=j_1(h)}^{j_2(h)} \sum_{i=1}^{n} O(i,j) = O(h). \tag{A9}
\]

The known right-hand side expression is combined and called \( O(h) \). The second left-hand term in Equation A9 denotes the sum of the anomalies within the period under inspection. Only anomalies occurring at the same time (but at other stations) as the candidate anomaly are taken into...
account. Thus besides the sizes of $B(i,j)$, it is the overlapping length of each anomaly with the candidate anomaly which counts. We define $L(h,h)$ as the number of overlapping years between the candidate inhomogeneity $B(h)$ and those from neighbouring stations $B(hh)$. Using the overlapping information $L(h,hh)$, we can rewrite Equation A9:

$$n l(h)B(h) - \sum_{hh=1}^H L(h,hh)B(hh) = O'(h). \quad (A10)$$

Considering all $H$ equations together, we can write them in matrix form,

$$\vec{M} \cdot \vec{B} = \vec{O}' . \quad (A11)$$

The diagonal of the matrix $M$ is given by

$$M(h,h) = (n-1) l(h), \quad (A12)$$

which can be deduced by setting $hh = h$ in Equation A10 and using the identity $L(h,h) = l(h)$. The non-diagonal elements are equal to the negative of the number of overlapping years,

$$M(h,hh) = -L(h,hh) \text{ for } h \neq hh. \quad (A13)$$

Figure A1 shows an example with five stations containing two breaks each so that 15 HSP exist. The corresponding inhomogeneities are numbered from $B1$ to $B15$. In the upper panel, $B8$ is chosen as candidate. In brackets the number of overlapping years with $B8$ is given: $L(8,i)$. $M$ is a sparse matrix, that is, many entries are zero, because many HSPs do not overlap. As the overlap of two HSPs is mutually equal, $M$ is symmetric. A further characteristic is that the sum of all elements in one row (or column) is equal to zero, because the sum of all non-diagonals is equal with opposite sign to the diagonal element ($n-1$ stations times $l(h)$, the length of the considered HSP). Therefore, the solution is not unique: Adding an arbitrary constant to a specific solution vector $B(h)$ is also a solution. Thus, the mean over all inhomogeneities is not determinable, but that is irrelevant for the trend.

Information about the position of the breaks is used to compute the matrix $M$. Assuming a pure break signal without noise, errors in the break positions may lead to an error variance of the inhomogeneities $B$ in the order of the break variance itself, if true and predicted break positions are completely uncorrelated. However, even if the detection is completely correct, there is a minimum error variance caused by the noise $\varepsilon$. In this situation the error variance of the estimated break signal is equal to the spuriously explained variance of a pure noise time series by random breaks. Lindau and Venema (2013) showed that this variance is equal to $k/(m-1) \sigma_n^2$, where $k$ denotes the number of breaks and $m$ the length of the time series. Thus, an estimate for the minimum error variance of the inhomogeneities, occurring even if the break detection is perfect, is

\[
\Delta \bar{B} \bar{B} = \frac{k}{m-1} \sigma_n^2. \quad (A14)
\]

**APPENDIX B: NUMBER OF INDEPENDENTS FOR THE STEP FUNCTION OF AN INHOMOGENEITY SIGNAL**

Consider a time series that is generated by choosing $m$ random numbers from a standard normal distribution with zero mean. A fraction of $(m-1)/m$ of the input variance can be found as internal variance in the obtained time series, whereas the rest ($1/m$) arise as external variation of the temporal means of each times series. This is the reason why we have to divide by $m-1$, when we want to estimate the variance of a sample. It is a commonly known feature that the factor of variance reduction attained by averaging is equal to the number of independents in the sample and in the above example all $m$ values are independent.

Consider now a time series of inhomogeneities (again with random and normal distributed deviations from zero). Such a time series contains obviously less independent values. It consists of only a few constant sub-periods

| Station | B1(0) | B2(32) | B3(0) | B4(0) | B5(0) | B6(0) | B7(0) | B8(19) | B9(0) |
|--------|-------|--------|-------|-------|-------|-------|-------|--------|-------|
| 0      | 282   | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |
| 1      | 0     | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |
| 2      | 0     | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |
| 3      | 0     | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |
| 4      | 0     | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |
| 5      | 0     | 0      | 0     | 0     | 0     | 0     | 0     | 0      | 0     |

**FIGURE A1** Sketch to illustrate the setup of the matrix (lower panel) that has to be solved for the ANOVA correction. The example considers five stations with altogether 15 HSPs. The circumstances for the eighth inhomogeneity (dark shaded) are given exemplarily in the upper panel. Crucial are the overlapping periods (light shaded), the length of which are additionally given in brackets for each HSP. These numbers occur with a negative sign in line 8 of the matrix $M$, which is shown in the lower panel. In this example, the diagonal element of line 8 is equal to 128 according to Equation for $n = 5$ and $l(h) = 32$.

\[
\Delta \bar{B} \bar{B} = \frac{k}{m-1} \sigma_n^2. \quad (A14)
\]
interrupted by jumps at the \( k \) breakpoints. The temporal mean of such a time series is given by the weighted mean,

\[
\bar{x} = \frac{1}{m} \sum_{i=1}^{k+1} l_i x_i, \quad (B1)
\]

where \( l_i \) denotes the length, \( x_i \) the value of each homogeneous sub-period and \( m \) the series length. The external variance is

\[
\text{Var}(\bar{x}) = \mathbb{E} \left[ (\bar{x})^2 \right] = \frac{1}{m^2} \left( \sum_{i=1}^{k+1} l_i \right)^2 \left[ x_i^2 \right] = \frac{1}{m^2} \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} [l_i l_j x_i x_j],
\]

(B2)

where the square brackets denote the expected value: \( \mathbb{E} [\cdot] \) is zero, which allows for the first equal sign in Equation B2.

As \( x \) is a random variable all mixed terms of \( i \) and \( j \) in Equation B2 become zero; further we can separate averages over \( x_i^2 \) and \( l_i^2 \) from each other as \( x \) and \( l \) are independent; and because the variance of \( x \) is 1, we can finally write

\[
\text{Var}(\bar{x}) = \frac{1}{m^2} \sum_{i=1}^{k+1} [l_i^2 x_i^2] = \frac{1}{m^2} \sum_{i=1}^{k+1} [l_i^2] \left[ x_i^2 \right] = \frac{1}{m^2} \sum_{i=1}^{k+1} [l_i^2].
\]

(B3)

If the \( k + 1 \) sub-periods have all the same length \( l \), the external variance amounts to

\[
\text{Var}(\bar{x}) = \frac{\mu^2}{m} \sum_{i=1}^{k+1} l = \frac{\mu^2}{m^2} (k+1).
\]

(B4)

For equal lengths \( l = ml/(k+1) \) and it follows:

\[
\text{Var}(\bar{x}) = \frac{1}{k+1}.
\]

(B5)

If the lengths are constant, averaging over a time series of \( k + 1 \) homogeneous sub-periods reduces the variance by the same factor, as \( k + 1 \) is not only the nominal, but also the effective number of independents. Equation B5 states that the error variance of the mean is equal to the variance (which is set to 1) divided by the number of observations \((k+1)\). If we do not set the sample average explicitly to zero the expectation of the mean is zero, but the actual means constituted by a limited number of values will vary around zero.

However, in the real world the lengths of inhomogeneities are variable. If a 100-year time series consists of a 96-year inhomogeneity and four further 1-year long ones, it is obvious that the effective number of independents must be smaller than 5. The variance of such time series means will be closer to 1 than to 1/5.

We restart our considerations for variable lengths with Equation B3. First, except for edge effects, the mean square length is the same for all \( k \) sub-periods. With this approximation it follows:

\[
\text{Var}(\bar{x}) = \frac{k+1}{m^2} \left( \frac{\mu^2}{k} \right).
\]

(B6)

Equation B6 shows that we need to know the mean square length of the inhomogeneities, which can be concluded from the frequency distribution of lengths. There are \( \binom{m-1}{k} \) possibilities to distribute \( k \) breaks in a time series of length \( m \). The relative frequency of a sub-period’s length is then given by,

\[
p(l) = \frac{\binom{m-l-1}{k-1}}{\binom{m-1}{k}}.
\]

(B7)

The nominator of Equation B7 gives the total number of possibilities for a given length \( l \), while the denominator gives the total number of combinations. The nominator of Equation B7 becomes plausible by the following consideration. We have \( k \) breaks (generating \( k + 1 \) sub-periods) within a time series of length \( m \). If a sub-period has the length \( l \), the remaining part (of length \( m - l \)) has to be shared by \( k \) sub-periods. Thus, we have to put \( k - 1 \) breaks into a time series of length \( m - l \). There are \( \binom{m-l-1}{k-1} \) different possibilities to do so.

We further see that \( l = m - k \) is the maximum length for a subperiod, because this leaves just enough space for the other \( k \) sub-periods. Expressing the mean over \( \mu^2 \) as the frequency weighted sum and inserting the maximum length as upper summation limit, we can rewrite Equation B6,

\[
\text{Var}(\bar{x}) = \frac{k+1}{m^2} \sum_{l=1}^{m-k} p(l) \mu^2.
\]

(B8)

Inserting Equation B7 into B8 we get

\[
\text{Var}(\bar{x}) = \frac{k+1}{m^2} \frac{1}{\binom{m-1}{k}} \sum_{l=1}^{m-k} \binom{m-l-1}{k-1} \mu^2.
\]

(B9)

The sum in Equation B9 can be solved by applying the following identity:

\[
\sum_{l=0}^{m} \binom{l}{j} \binom{m-l}{k-j} = \binom{m+1}{k+1}.
\]

(B10)

Using the solutions for \( j = 1 \) and \( j = 2 \), it follows:

\[
\sum_{l=1}^{m-k} \binom{m-l-1}{k-1} \mu^2 = \frac{2m-k}{k+2} \binom{m}{k+1}.
\]

(B11)

Inserting Equation B11 into B9 we obtain

\[
\text{Var}(\bar{x}) = \frac{k+1}{m^2} \frac{1}{\binom{m-1}{k}} \frac{2m-k}{k+2} \binom{m}{k+1}.
\]

(B12)
The two binomial coefficients can be expressed as one quotient,

\[
\text{Var}(\tau) = \frac{k+1}{m^2} \frac{m}{k+1} \frac{2m-k}{k+2}.
\]  
(B13)

Finally we get

\[
\text{Var}(\tau) = \frac{2m-k}{m (k+2)}.
\]  
(B14)

For \( m \gg k > 1 \) we can further approximate

\[
\text{Var}(\tau) \approx \frac{2}{(k+2)}.
\]  
(B15)

The number of independents is equal to the reciprocal of the result in Equation B15,

\[
m_{\text{in}} \approx \frac{k}{2} + 1.
\]  
(B16)