



To what extent does the detection of climate change in Hungary depend on the choice of statistical methods?

Beatrix Izsák^{1,2} · Tamás Szentimrey³

Received: 25 October 2019 / Accepted: 4 June 2020 / Published online: 24 June 2020
© The Author(s) 2020, corrected publication 2020

Abstract

The trend analysis of meteorological time series has gained prominence in recent decades, the most common method being the so-called ‘linear analytical trend analysis’. Until the mid-1990s, trend analysis was commonly performed on non-homogenized data sets, which frequently led to erroneous conclusions. Nowadays, only homogenized data sets are examined, so it really is possible to detect climate change in long meteorological data sets. In this paper, the methodology of linear trend analysis is summarized, the way in which the model can be validated is demonstrated, and there is a discussion of the results obtained if unjustified discontinuities caused by changing measurement conditions, such as the relocation of stations, changes in measurement time, or instrument change occur. On the basis of an examination of records for the preceding 118 years, it is possible to state that both annual and seasonal mean temperature trends display a significant warming trend. In the case of homogenized data series, the change is significant over the entire territory of Hungary; in the case of raw data series, however, the change is not significant everywhere. The validity of the linear model is tested using the F-test, a task as yet carried out on the entire Hungarian data series, series comprising records for over 100 years. Furthermore, neither has a comparison been made of the trend data for raw data series and the homogenized data series with the help of information on station history to explore the causes of inhomogeneity.

Keywords Climate change · Linear trend · Homogenization · Interpolation

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13137-020-00154-y>) contains supplementary material, which is available to authorized users.

✉ Beatrix Izsák
izsak.b@met.hu

¹ Unit of Climatology, Hungarian Meteorological Service, Kitaibel Pál u.1, Budapest 1024, Hungary

² ELTE Faculty of Science, Doctoral School of Earth Sciences, Budapest, Hungary

³ Varimax Limited Partnership, Budapest, Hungary

1 Introduction

The global average temperature has increased at an unprecedented pace over the past century. It has been shown (e.g., Bloomfield 1992; Gao and Hawthorne 2006; Keller 2009) that this increase is statistically significant (IPCC 2013; Foster and Rahmstorf 2011). Increases in temperature have become apparent in many places, both on local and regional scales. The problem is that the data on regional and local temperatures obtain a greater degree of noise than the global average, due to the higher level of averaging in the case of global figures. It is therefore not always clear whether a regional or local warming signal, although apparent, can be considered statistically significant.

Long term instrumental climate observations are available to study the warming signal. The value of these datasets, however, strongly depends on their homogeneity. Results from the homogenization of instrumental climate records indicate that the typical size of inhomogeneities is often of the same order as the climate change signal itself during the twentieth century (Auer et al. 2007; Menne et al. 2009; Brunetti et al. 2006; Caussinus and Mestre 2004; Della-Marta et al. 2004).

Although homogenization began in the 1980s (Alexandersson 1986), the first homogenization seminar was held, in Budapest, in 1996 (Szalai 1996). Since then, homogenized datasets have been produced in many countries and numerous software programs have been published in meteorology (Aguilar 2003; Szentimrey 1998). In recent decades, not only has the need for homogenization been accepted, but it is generally considered that only via these climate time series is it worth studying climate, and especially climate change (Yosef et al. 2019; Mamara et al. 2016; Caussinus and Lyazrhi 1997; Trewin 2013; Zhang et al. 2000).

The purpose of this study is to show the difference between the results of trend analysis performed on raw and homogenized data sets and to uncover station history information (META data) that may be responsible for any inconsistencies found during homogenization (Delvaux et al. 2019; Jones et al. 2004). A further aim in the present work is to determine which kinds of data series the linear model can accept. This type of analysis has previously been conducted on only a small number of raw data series in Hungary (Szentimrey 1989), but in this study statistical tests are performed for annual and seasonal mean temperature values from 25 Hungarian meteorological stations, and not only are they performed on raw, but also on homogenized data sets.

2 Materials and methods

2.1 Data

For the purposes of this study, daily average temperature data for 25 Hungarian stations (Fig. 1) were used covering the period 01.01.1901 to 31.12.2018. As a first step, representative time series had to be generated from the raw

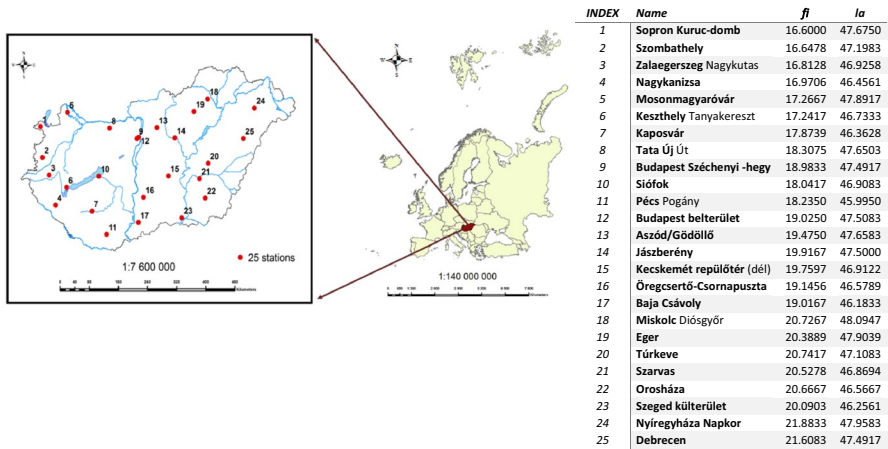


Fig. 1 Location of the stations: name, geographical coordinates where la is the latitude and fi is the longitude (EPSG:4326;WGS 84)

measurements. Homogenization removes inhomogeneities from the stations' data series, acts as a form of quality control, and fills the gaps. In order to calculate a national average, the station data series are interpolated to a relatively dense regular grid. The annual and seasonal averages calculated from the gridpoints thus obtained can rightly be called a national average. In the course of gridding, these raw and homogenized station values are interpolated to a regular grid at a resolution of 0.1 degree.

2.2 Homogenization, data completion, quality control in MASHv3.03

Climate studies, in particular those related to climate change, require long, high-quality, controlled, data sets which are both spatially and temporally representative. Changing the context in which the measurements were taken, for example relocating the station, or a change in the frequency of measurements, or in the instruments used may result in an unduly fractured time series (Klein Tank et al. 2002; Xu et al. 2013). Data errors and inhomogeneities are eliminated and data gaps are filled in using the MASH (Multiple Analysis of Series for Homogenization; Szentimrey 2013) homogenization procedure.

What kind of software is employed for homogenization is of great importance, because if not just inhomogeneities are removed from the data series, but also the process unintentionally alters the signal of climate change, the result will be misleading. Thanks to the mathematical model, using the MASH software, it is possible to detect climate change in the homogenized data set (Venema et al. 2012; Szentimrey 2006a; Peterson et al. 1998).

2.2.1 Additive model of monthly mean temperature series in MASH

If the data series are normally distributed (e.g. temperature), then the additive model can be used (Szentimrey 2006b). In the case of relative methods, a general form of additive model for additional monthly series belonging to the same month in a small climate region can be expressed as follows,

$$X_j(t) = \mu(t) + E_j + IH_j(t) + \varepsilon_j(t) \quad (j = 1, 2, \dots, N; t = 1, 2, \dots, n), \quad (1)$$

where $\mu(t)$ is the common and unknown climate change signal, E_j represents the spatial expected value, $IH_j(t)$ the inhomogeneity signals and $\varepsilon_j(t)$ normal white noise series. The type of inhomogeneity $IH(t)$ is in general a 'step-like function' with unknown break points T and shifts $IH(T) - IH(T + 1) \neq 0$, and it is generally assumed that $IH(n) = 0$. The normally distributed vector variables $\varepsilon(t) = [\varepsilon_1(t), \dots, \varepsilon_N(t)]^T \in N(\mathbf{0}, \mathbf{C})(t = 1, \dots, n)$ are totally independent in time. The spatial covariance matrix \mathbf{C} describes the spatial structure of the series (Szentimrey 2014).

2.2.2 The main features of MASHv3.03

Advantages of MASHv3.03 in the homogenization of monthly series:

- It is a relative homogeneity test procedure.
- It is a step-by-step iteration procedure: the role of series (candidate, reference) changes step by step in the course of the procedure (Supplementary Sects. 1, 2, 3).
- An additive or multiplicative model can be used depending on the distribution.
- It includes quality control and missing data completion.
- It provides the homogeneity of the seasonal and annual series as well.
- Metadata (probable dates of break points) can be used automatically.
- The homogenization results and the metadata can be verified.

In the homogenization of daily series:

- The procedure is based on the detected monthly inhomogeneities.
- It includes quality control and the completion of missing data in daily data.

2.3 MISH v1.03 software

2.3.1 Linear meteorological model for expected values

In the statistical modeling of the meteorological elements it is necessary to assume that the expected values of the variables will change in space and in time alike. The spatial change means that the climate is different in different regions.

The temporal change is the result of any changes in global climate. Consequently, in the case of the linear modeling of expected values it is assumed that

$$E(Z(s_i, t)) = \mu(t) + E(s_i) \quad (i = 0, \dots, M) \tag{2}$$

where the location vectors \mathbf{s} represent the elements of the given space domain D and t is the time, $\mu(t)$ is the temporal trend, that is, the climate change signal and $E(\mathbf{s})$ is the spatial trend (Szentimrey and Bihari 2014).

2.3.2 Additive (linear) interpolation formula

Assuming a linear model, the appropriate additive meteorological interpolation formula is as follows,

$$\hat{Z}(s_0, t) = \lambda_0 + \sum_{i=1}^M \lambda_i \cdot Z(s_i, t)$$

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

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

$$RMSE(s_0) = \sqrt{E \left(\left(Z(s_0, t) - \hat{Z}(s_0, t) \right)^2 \right)},$$

and by the representativity value: $REP(s_0) = 1 - \frac{RMSE(s_0)}{D(s_0)}$.

The optimal interpolation parameters $\lambda_0, \lambda_i (i = 1, \dots, M)$ minimize the root-mean-square error and these are known functions of the local statistical parameters (expectations, standard deviations) and the stochastic connections (correlations), which are climate statistical parameters in meteorology.

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

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

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

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

Consequently the unknown statistical parameters are the spatial trend differences $E(\mathbf{s}_0) - E(\mathbf{s}_i) (i = 1, \dots, M)$, the standard deviation ratios $D(\mathbf{s}_0)/D(\mathbf{s}_i) (i = 1, \dots, M)$ and the correlation system \mathbf{r}, \mathbf{R} . In essence these parameters are climate parameters which in fact means we could interpolate optimally if we knew the climate (Szentimrey et al. 2011; Szentimrey et al. 2014).

2.3.3 The main features of MISHv1.03

The software version MISHv1.03 (Meteorological Interpolation based on Surface Homogenized Data Basis; Szentimrey and Bihari 2014) consists of two units, the modeling and the interpolation systems. The interpolation system can be applied to the results of the modeling system (Supplementary Sect. 4). Below, there is a summary of these two units of the software developed.

The modeling subsystem for climate statistical (local and stochastic) parameters:

- This is based on long homogenized data series and supplementary deterministic model variables. The model variables may include such elements as height, topography, distance from the sea etc. There is neighbourhood modeling, correlation model for each grid point.
- It is also a benchmark study, a cross-validation test for interpolation error or representativity.
- It should be noted that the modeling procedure must be executed only once before the interpolation applications.

The interpolation subsystem:

- Additive (e.g. temperature) or multiplicative (e.g. precipitation) model and interpolation formula can also be used, depending on the climate elements.
- Daily or monthly values and the means from a number of years can be interpolated.
- Just a few predictors are sufficient for the interpolation, and no problem arises if the greater part of daily precipitation predictors is equal to 0.
- Representativity is also modelled.
- Supplementary background information (stochastic variables) e.g. satellite, radar, forecast data can also be used.
- Data series completion, that is, missing value interpolation, completion for monthly or daily station data series is possible.
- Interpolation, the gridding of monthly or daily station data series for given predictand locations is made possible. In case of gridding the predictand locations are the nodes of a relatively dense grid.

2.4 Availability of MASHv3.03 and MISHv1.03 software

Both pieces of software can be downloaded from the website of the Hungarian Meteorological Service. www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/.

2.5 Linear trend estimation

In this chapter the linear model will be described. The use of the α -confidence level estimation in preference to the usual trend coefficient is explained in relation to its use in the description of the change over the total period.

2.5.1 General case

In a discrete case, the following general model for the meteorological time series may be described as:

$$x(t) = m(t) + \varepsilon(t) \quad t = 1, 2, \dots, n \tag{3}$$

where the trend function of the temporal changes of the measurements $x(t)$ is denoted by $m(t)$ and is accompanied by the noise $\varepsilon(t)$; in situations where each element is identically distributed, their expected value is 0, their standard deviation is σ , and they are totally independent. For the investigation of meteorological time series, a more specialized model than (3) is employed, namely the linear model (Szentimrey 1989).

2.5.2 Linear model

In the linear model, measurements are approximated as the following function of time (Dévényi and Gulyás 1988; Mudelsee 2019; Sneyers 1990):

$$x(t) = c_1 + c_2 \cdot t + \varepsilon(t) \quad t = 1, 2, \dots, n \tag{4}$$

where the elements of series $\varepsilon(t)$ are identically distributed and they are totally independent, $E(\varepsilon(t))=0$ and $D(\varepsilon(t))=\sigma$ ($t=1,2,\dots,n$).

Estimation:

$$\hat{c}_1 = \bar{x} - \hat{c}_2 \bar{t} \quad \hat{c}_2 = \frac{\sum_{t=1}^n (x(t) - \bar{x}(t))(t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}$$

where

$$\bar{x} = \frac{\sum_{t=1}^n x(t)}{n} \quad \text{and} \quad \bar{t} = \frac{\sum_{t=1}^n t}{n}$$

Estimation of change over the total period: $\hat{c}_2 \cdot (n - 1)$.

A couple of texts on mathematical statistics provide the linear trend estimation and its properties described below (Móri and Székely 1986; Wooldridge 2013).

Properties:

- Generally: best linear unbiased estimator (BLUE)
- In the case of normal distribution ($\varepsilon(t) \in N(0;\sigma^2)$):

$$\hat{c}_2 \in N\left(c_2; \frac{\sigma^2}{\sum_{t=1}^n (t - \bar{t})^2}\right);$$

$$S^2 = \frac{1}{n - 2} \sum_{t=1}^n (x(t) - (\hat{c}_1 + \hat{c}_2 \cdot t))^2 \in \frac{\sigma^2}{n - 2} \chi_{n-2}^2$$

and \hat{c} and S are independent.

$$\frac{\hat{c}_2 - c_2}{S} \sqrt{\sum_{t=1}^n (t - \bar{t})^2} \in \text{Student}_{n-2} \quad \text{where } (S = \sqrt{S^2}). \quad (5)$$

If a series of data is investigated for which the noise component is not normally distributed and/or the temporal independence is violated, other procedures should be considered, as described in Mudelsee (2019). However, in the present case this is not an issue, since the assumptions (Dévényi and Gulyás 1988; Spinoni et al. 2015; Jones et al. 2004) are acceptable for annual, seasonal, and monthly average temperature series as well, both in terms of normality and independence. The entire data set is not examined together in the case of trend estimation, but the appropriate seasons or months are taken separately, as is the case in the homogenization and interpolation.

2.5.3 Significance, definition of the α -confidence level estimation

In this study it is assumed that the noise is normally distributed, which is acceptable in the case of annual, seasonal and monthly average temperatures. The hypothesis test can be used to verify whether any change is significant. On the basis of property (5) an α -confidence level confidence interval $(\hat{c}_{2_{\alpha 1}}, \hat{c}_{2_{\alpha 2}})$ can be given for c_2 . The proposition can be accepted that $c_2 \neq 0$, that is, a trend can be detected from the data series if the confidence interval does not include 0. So in this case the null hypothesis is rejected, and there is a change. If the confidence interval contains only positive values, then a significant increase has taken place, if only negative, a decrease. If the extent and direction of climate change over a larger area (in this case the whole of Hungary) is to be investigated, it is useful to calculate the α -confidence level estimate and present these values numerically.

1. **Definition** Let the α -confidence level estimation $(\hat{c}_{2_{\alpha}})$ be:

- $\hat{c}_{2_{\alpha}} = 0$ if $0 \in (\hat{c}_{2_{\alpha 1}}, \hat{c}_{2_{\alpha 2}})$;
- $\hat{c}_{2_{\alpha}} = \hat{c}_{2_{\alpha 1}}$ if $0 < \hat{c}_{2_{\alpha 1}}$;
- $\hat{c}_{2_{\alpha}} = \hat{c}_{2_{\alpha 2}}$ if $\hat{c}_{2_{\alpha 2}} < 0$.

The estimation of significant change over the total period may then be calculated: $\hat{c}_{2_{\alpha}}(n-1)$. If the trend coefficient is not displayed but the α -confidence level estimation is, then at the given confidence level it is possible to be certain that throughout the given period this amount of change occurred.

2.6 Spatial interpolation of the temporal trend

After calculating the change over the total period (Fig. 2), the resulting values are interpolated into a dense, regular grid to provide an accurate nationwide map of the trend values in space. For this purpose, the MISH program system is used

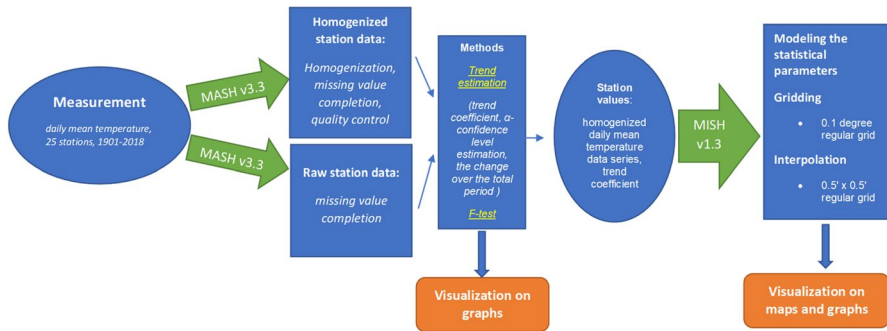


Fig. 2 A general overview of the main elements of the applied statistical data analysis processes and results

(Szentimrey and Bihari 2014), a program developed specifically for the interpolation of meteorological elements. In this case, the trend values are interpolated, since it is clearly observable that the same result is obtained by interpolating station series and then calculating the trend as would be if the station trends were interpolated. Of course, this is only true if the interpolation is based on an adequate mathematical formula, e.g. MISH (Szentimrey et al. 2011). In other cases, of necessity only the station values would be displayed on the map (Vincent et al. 2012).

Because MISH software was specifically developed to interpolate meteorological elements, its use is internationally accepted. The CARPATCLIM database has been developed in collaboration between 9 European countries, and is a database in which homogenization was performed by MASHv3.03 and interpolation by MISHv1.03 software (Szalai et al. 2013). Trend analyses were also conducted on this database, representative both in space and in time (Spinoni et al. 2015).

In this work the seasonal, annual average mean temperature trends were calculated for 25 Hungarian stations over the period of 1901–2018. In order to display trend values on maps, interpolation is not performed with MISH software to a 0.1×0.1 degree grid point, but at a much denser resolution, namely $0.5' \times 0.5'$, taking advantage of the modeling element of MISH. Because of this dense resolution it is possible to interpolate to more than 200,000 Hungarian gridpoints. In order to show how the statistics are misrepresented by the inhomogeneous data set, the same statistics as for the homogenized data set were calculated a number of times. The raw data sets were only filled in by the MASH program and the data sets were not homogenized and quality controlled. For both raw and homogenized data series the change over the total period was calculated using the trend coefficient estimate. In this study, the confidence level (α) is 0.9 for all statistics.

2.7 Trend estimation in general cases

Approaching the trend function with a linear function, the change over the total period is the sum of small decreases or increases of equal magnitude each year. However, it is clear that the situation is not so simple, and that the climate is highly

unlikely to be changing to an identical extent every year. However, if the aim is to determine the change over the total period, the linear model is simple and readily applicable. In order to be able to use the results, i.e. to determine whether the linear model is acceptable, the validity of the assumption must be checked (4). The question arises of how linear model estimation can be interpreted in the general case (3), and in this case, how the results obtained should be interpreted. Szentimrey showed that linear analytical trend analysis can be applied to the more general case (3), but in this case the vector \hat{c} is used to estimate the coefficient vector of the projection of the trend function onto a linearly independent function system (Szentimrey 1989). A special case is shown in Fig. 3. The grey dots represent the observations, the theoretical trend function is shown by the blue line, the projection of the theoretical trend function by the black line. The estimation based on the linear model is represented by the red dashed line.

Those properties of the estimation which are good are also true of the general case, in which statements do not refer to the trend function, but to its projection. In this general case, it is possible to apply the tests given for the linear model and construct a confidence interval for the trend coefficient.

2.8 F-test

The difference between the trend function and the projection can be checked with an F-test. Assuming that noise is normally distributed, which is acceptable for annual, seasonal and monthly average temperatures, the test can be given as follows (Szentimrey 1989):

The null hypothesis:

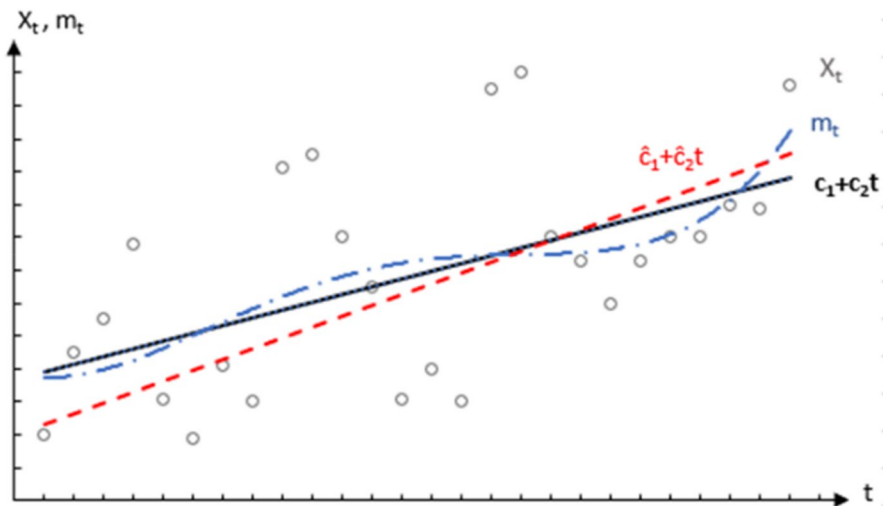


Fig. 3 Observation (x_t), trend function (m_t), the projection of the trend function ($c_1 + c_2 t$) and the estimation ($\hat{c}_1 + \hat{c}_2 t$)

$$H_0 : \exists \mathbf{c} = (c_1, c_2), \quad m(t) \equiv c_1 + c_2 \cdot t \quad t = 1, \dots, n$$

In this case, the test statistics (PS) can be written as follows:

$$PS = \frac{[n/2] - 1}{n - [n/2] - 1} \left(\frac{\sum_{t=1}^n (x(t) - (\hat{c}_1 + \hat{c}_2 \cdot t))^2}{\frac{1}{2} \sum_{t=1}^{[n/2]} (\delta(t) - \bar{\delta})^2} - 1 \right) \tag{6}$$

where $[n/2]$ is the integer part of $n/2$, $\delta(t) = x(2t) - x(2t - 1)$ where $t = 1, \dots, [n/2]$, $\bar{\delta} = \frac{\sum_{t=1}^{[n/2]} \delta(t)}{[n/2]}$.

It can be proven that if the null hypothesis is true, then the PS statistics follow an F distribution with parameters $n - [n/2] - 1, [n/2] - 1$ (Szentimrey 1989).

3 Results and discussion

3.1 Annual trend

In the following section the results are presented, rather representatively than exhaustively. Figure 4a, b shows the change in the mean annual temperature over the total period, the values are in °C, and the trend coefficient was interpolated using MISH for the whole territory of Hungary. Comparing the two maps, a completely different picture is obtained. Once the endpoints of the confidence interval are calculated, the question of whether the change is significant on the given confidence level is addressed (1. Definition). If it is not, no trend is given, whereas in the case of a significant change, the lower absolute value of the endpoints of the confidence interval will be the α -confidence level estimate (\hat{c}_α). If the change over the total period is calculated with the α -confidence level estimate, it is possible to say that with a probability α in the last 118 years, such-and-such an amount of change happened (Fig. 4c, d). While the change is significant in the case of homogenized data sets at the 0.9 confidence level, almost a quarter of the raw data series does not show any significant change at the 0.9 confidence level. The detailed results of trend estimation can be seen in “Appendix”. The average difference between the homogenized and raw trend (estimation of change over the total period) is generally 0.3 °C.

To understand why such different results are obtained, it is necessary to look at station trends (Fig. 5). It can be clearly seen that in the case of homogenized data series there is a significant increase of temperature for each series at the 0.9 confidence level, while in the case of a quarter of the raw data series no change can be detected, or if so, then there are significant differences at the 0.9 confidence level.

The greatest differences can be seen in the cities, where homogenization with MASH gave the highest value of inhomogeneities. There is no significant change at the 0.9 confidence level in the raw data series at stations 6, 11, 16, 21 and 23, whereas the homogenized data series do show a change over the entire period of about 0.8 °C at the 0.9 confidence level. Which corrections were made during homogenization and the available station history data are shown. It was already

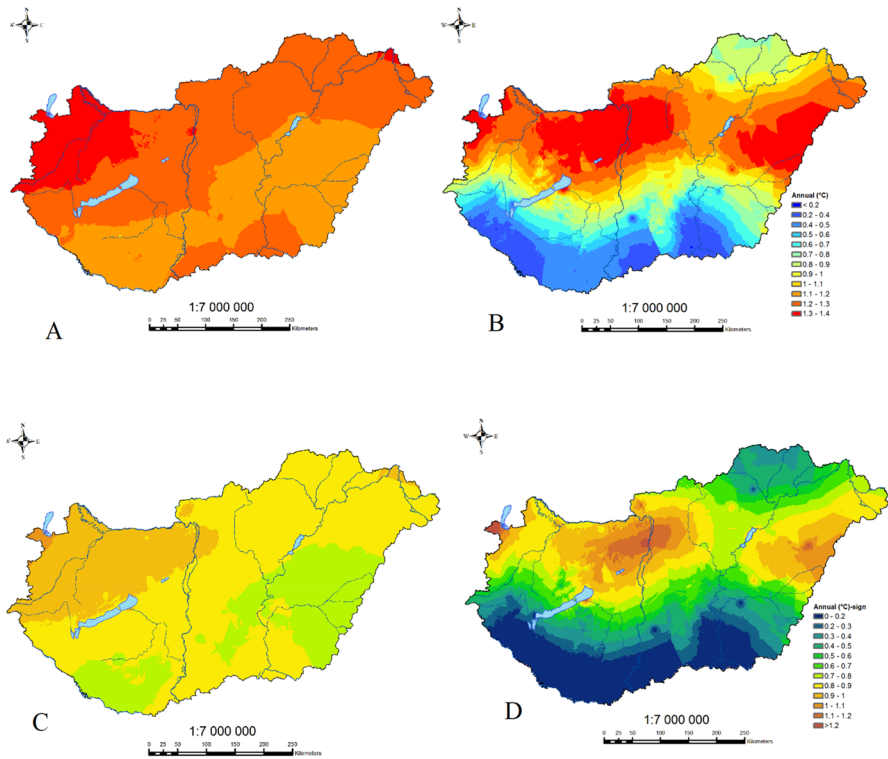


Fig. 4 Annual mean temperature: Estimation of change over the total period $((n - 1) \hat{c}_2)$, 1901–2018 ($^{\circ}\text{C}$) homogenized series (a) and raw series (b); change over the total period, $\alpha=0.9$ confidence level estimation $\hat{c}_{2_{\alpha}}(n - 1)$, 1901–2018 ($^{\circ}\text{C}$), homogenized series (c) and raw series (d)

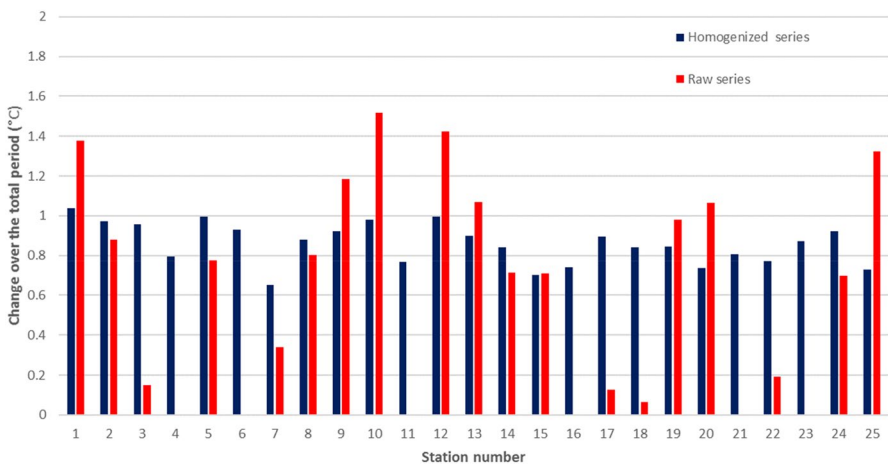


Fig. 5 Station data of annual mean temperature, homogenized series and raw series Change over the total period, $\alpha=0.9$ confidence level estimation $\hat{c}_{2_{\alpha}}(n - 1)$, 1901–2018 ($^{\circ}\text{C}$)

clear that META data alone do not justify breakpoints, and that records are often incomplete (Peterson et al. 1998). The best known inhomogeneity is the urban heat island effect. On the other hand, with the advent of aviation, meteorological stations have often been relocated from cities to nearby, typically cooler, airports (Vincent et al. 2012; Trewin 2010). In this study, however, an example of precisely the opposite effect is present: at one of the Hungarian stations, the airport is located in a warmer micro-climate than that of its previous site. Other non-climatic changes can also be caused by changes in measurement methods: change in measurement time, or instrument change (Begert et al. 2005).

As became apparent in the course of this investigation, one of the most common causes of inhomogeneity is that the station has been moved from indoors to outdoors or vice versa and thus the measurement conditions have changed significantly. The second most common cause is that a methodological change has taken place, namely, the daily average temperature has been calculated from other data because K3 stations (K3 = "limatological station" with 3 measurements/day) became K4 (K4 = "climatological station" with 4 measurements/day), K8 (K8 = "climatological station" with 8 measurements/day), S1 (S1 = "synoptic station" 144 measurements/day) or S2 (S2 = "synoptic station with observer" 24 measurements/day) stations. In this case the extra night time measurement or the hourly or 10 min measurement were significantly different from the average temperature calculated from the 3 measurements/day previously used.

Nagykanizsa (station 4): The raw data series show no significant change (Fig. 6a), whereas the homogenized series show a significant increase of 0.8 °C at the 0.9 confidence level. This was mainly due to the relocation to the outskirts in 1951 and the change from K3 to K8 status. The station was subsequently relocated in 1957, 1958, 1959, 1960 and 1979 but these relocations did not cause as much disruption as the 1951 change in methodology and relocation.

Keszthely (station 6): The homogenized data sets show a warming of 0.93 °C at the 0.9 confidence level (Fig. 6b), while the raw data sets show no significant change at the 0.9 confidence level. The station was moved in 1962, 1966, 1995, 2000 and 2001. The biggest discontinuity was caused by the relocation to the outskirts in 1995, whereas prior to that it was at the garden of the Academy and College of Agriculture, while since 1995 it has been operating at Tanyakereszt. In the meanwhile, measurements were carried out at FlyBalaton Airport for 1 year, and when this ceased it returned to Tanyakereszt. Because homogenization adjusts to the present, MASH automatically downgraded the city centre homogenized data series.

Pécs (station 11): In Pécs, the migration to Pécs Pogány caused the highest detected inhomogeneity. Before that, measurements were made at the College of Education, the Erzsébet University, and then in the garden of a nunnery called Notre Dame in downtown Pécs. In 1956, the measurements were moved to Pogány because an airport was built there. The transition from K3 to K8 status in 1956 was also a key turning point. From 1969 onwards it was an S2 station. It can be seen that while the raw data series show no change at the 0.9 confidence level in the homogenized data series, the total change over the entire period is estimated at 0.77 °C at the 0.9 confidence level (Fig. 6c).

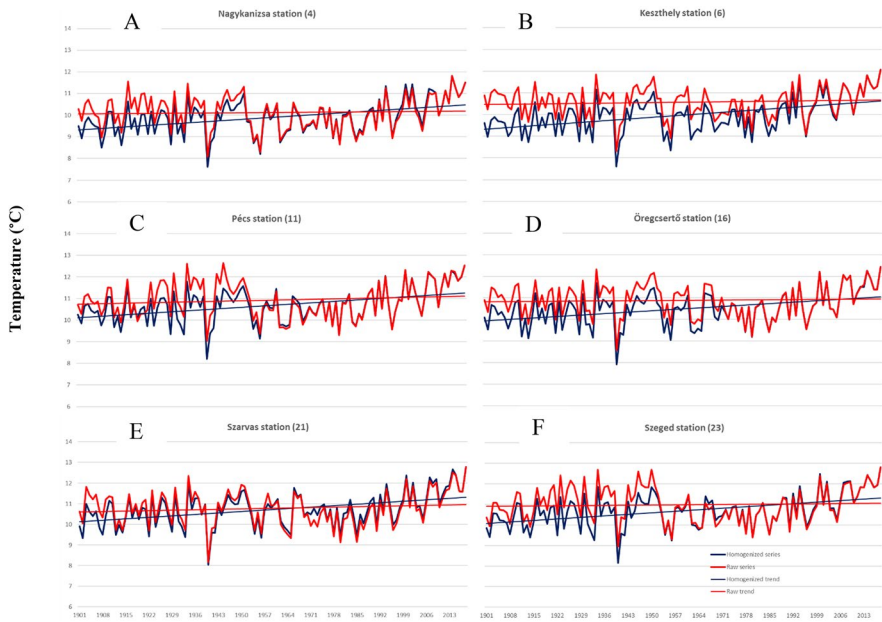


Fig. 6 Station data of annual mean temperature, homogenized series and raw series and the fitted trend-lines at Nagykanizsa (a), Keszthely (b), Pécs (c), Öregcsertő (d), Szarvas (e), Szeged (f), 1901–2018

Öregcsertő-Csornapuszta (station 16): Measurements were taken within the town of Kalocsa until 1961, then in 1961–1992 the station moved to the state economy at a considerable distance from the town center. In 1966 the status of the station was changed from K3 to K4. After that the station moved to Öregcsertő and then to Csornapuszta in 2006, both of which are in the outskirts. $0.74\text{ }^{\circ}\text{C}$ is an estimate of the change over the total period for homogenized series at the 0.9 confidence level (Fig. 6d). The significant difference between homogenized and raw data sets is due to those changes that took place in the 1960s.

Szarvas (station 21): From 1901 the measurements were made at the garden of the local High School until 1928. At that time it was moved to an Economics School, and in 1936 to another location in the outskirts. The Hungarian Meteorological Service's research station from 1968, this location saw a change in status to K4. There was another move in 1975, and again in 1998 its status became an S1 station. It appears that in the course of homogenization, the raw measurements were shifted towards lower values before 1968 and higher ones thereafter. At this station, not only the relocations but also the methodological changes contributed to the fact that the two trend values are very different (Fig. 6e): in the case of homogenized data series it is $0.81\text{ }^{\circ}\text{C}$ at the 0.9 confidence level, while in the case of raw data series it is $0\text{ }^{\circ}\text{C}$ at the 0.9 confidence level.

Szeged (station 23): In common with other stations, where there is no significant change in temperature for the raw data series at the 0.9 confidence level, the Szeged station has moved several times. Until 1926, the measurements were made in the

garden of the Piarist High School, and then until 1951 they were made at the University. A significant degree of inhomogeneity was caused by the relocation from location within the city to the outskirts in 1951. In this case, too, the station moved to an airport and at the same time switched from K3 to K8 status. The station also moved within the airport, but the measurement program changed significantly with the change from K8 to S1 status in 2004. The homogenized data series displayed a change of 0.87 °C over the total period at the 0.9 confidence level (Fig. 6f), while the raw data series do not show any trend due to a less extreme diurnal pattern of temperature variation.

The results of three stations showed that the rate of warming up is significantly higher for the raw data series than for the homogenized series: those at Budapest, Siófok and Debrecen. In the case of the center of Budapest (station 12) the effect of the urban heat island effect can be clearly detected in the raw data series (Fig. 7a), while at Siófok (10) the relocation of the station from the city center to the waterfront causes a break in the data series (Fig. 7b). In the case of Debrecen (station 25), moving to the outskirts did not result in cooling, and in fact higher values were measured. The data from the period between 1901–1950 was collected mainly in Debrecen-Pallag (Fig. 7c), which is in the northern part of the city, in a location surrounded by forests, though the station was then moved to the airport, which is the south part of the city, where there is extensive pasture land around the airport. It is widely known that the climate in a forest is cooler than in such open, flat terrain.

3.2 Seasonal trend

Figure S1 shows the seasonal trends for cases in which it can be clearly seen that the warming is stronger in spring and summer than in autumn and winter. In the case of raw data series, the change is not significant in several cases at the 0.9 confidence level. Figure S2 shows the amount of change in spring over 118 years, on the basis of an α -confidence level estimate. As in Hungary the average spring temperature in the homogenized data series increases the most, there is a striking contradiction in the case of raw data series. In this case, no significant trend is obtained in more than one third of the area. As with spring, the summers also display strong warming, with the trend for raw and homogenized data series shown in Fig. S3. Autumn (Fig. S4) and winter (Fig. S5) show less warming in the homogenized data sets, as the raw data sets do not show significant changes at the 0.9 confidence level in as many places as the homogenized data sets.

3.3 Cooling down detected when checking raw monthly values

An examination of the trend of monthly average temperatures shows that they display a greater degree of variability; for example, in December there is no significant change in the homogenized data series at the 0.9 confidence level, but the raw data series display a significant negative trend at the 0.9 confidence level, and this is also true when a station was moved to the outskirts. That station was the one at Pécs (11), whose raw and homogenized December data are shown in Fig. 8. It is clear to what

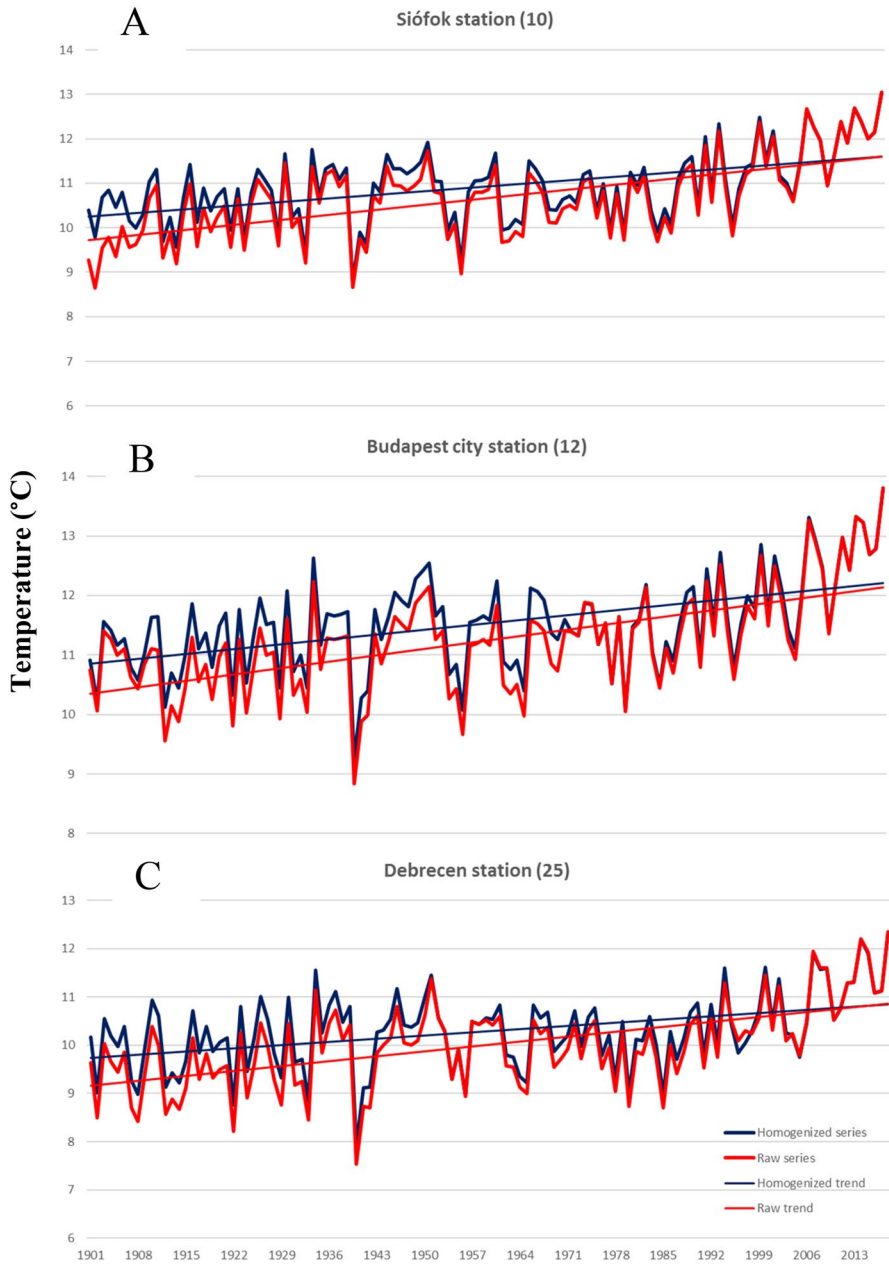


Fig. 7 Station data of annual mean temperature, homogenized series and raw series and the fitted trend-lines at Siófok (a), Budapest (b), Debrecen (c), 1901–2018

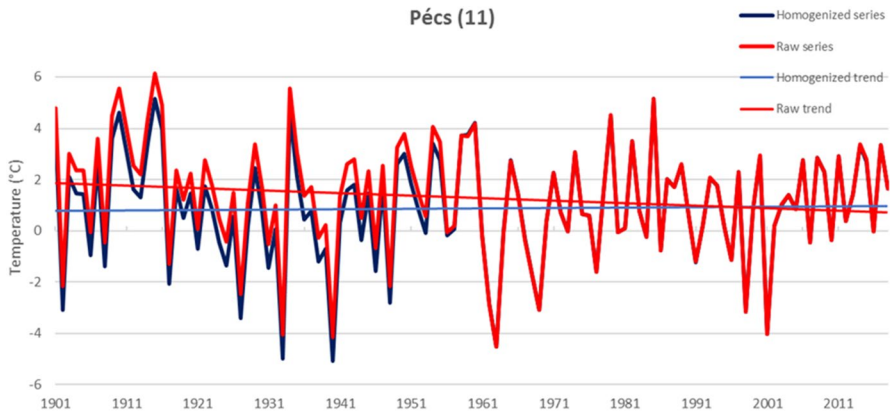


Fig. 8 December mean temperature of Pécs station, homogenized series and raw series and the fitted trendlines, 1901–2018 (°C)

extent inhomogeneity was caused by moving the station from the city in 1956. If this inhomogeneity were not removed from the raw data series, we would falsely draw the conclusion that climate change at Pécs is in the opposite direction of the national or global trend. Of course, this is not the case, which is why homogenization has to be performed before climate studies are carried out and the statistics determined.

3.4 F-test

While the PS test statistics provided by Formula (6) fall below the critical value of 1.53 for a significance level of 0.1 in the 118-year homogenized time series of mean annual temperatures for 25 national stations, the same is not true for the raw data series. In this case, data errors and inhomogeneities make it impossible to examine only climate change. The test statistical values are shown in Fig. 9.

After applying the F-test to seasonal temperature values it was found that the summer values show a very different pattern. The PS values of seasonal mean temperatures for homogenized and raw series are shown in Fig. 10. Autumn and winter test statistics show that the linear model is acceptable with the raw and the homogenized data series as well. However, in the case of spring there are two stations where the linear model is not acceptable. In the case of summer values, most of the test statistics seem to be higher than the critical value.

Figure S6 displays the homogenized seasonal mean temperature values. The scattering of summer averages is typically below and above the trend line fitted to the homogenized data series, and the summer standard deviation is the smallest of any of the seasons, so the difference between the trend function and its projection can be relatively large in any case. Here the reader is referred back to Sects. 2.7 and 2.8: according to the F-test, the linear model is unacceptable for summer average temperatures. A higher order, linearly independent function system might be more appropriate, as it would allow the differences between the

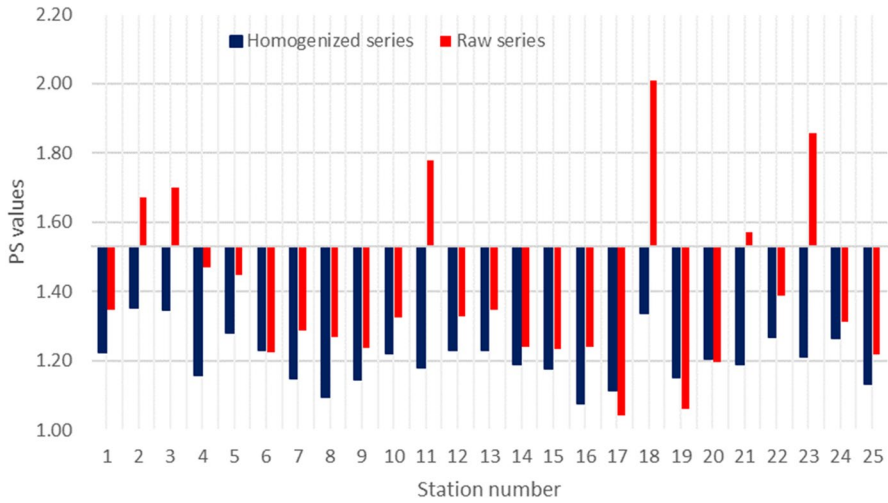


Fig. 9 Station data of PS values, homogenized series and raw series, annual mean temperature, The critical value of 1.53 for a significance level of 0.1

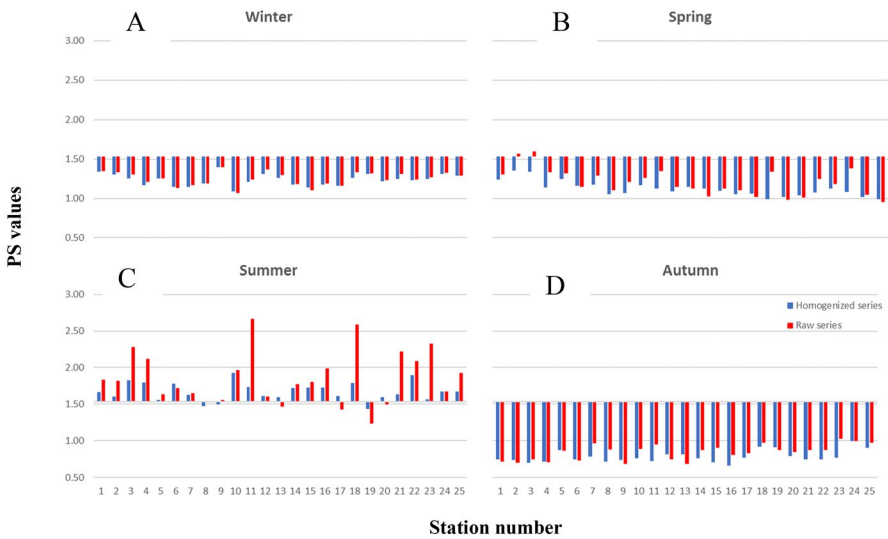


Fig. 10 Station data of PS values, homogenized series and raw series, seasonal mean temperature Winter (a), Spring (b), Summer (c), Autumn (d)

trend function and its projections to the function system to be kept to a minimum. Returning, however, to the general case, it is still possible to use the estimate for change over the total period, in which case the statements made here do not refer to the trend function, but to its projection.

4 Conclusion

On the basis of the results obtained, it is possible to say that only homogenized and quality controlled data sets should be used for climate change studies. It is also reasonable to state that the linear model is acceptable at a significance level of 0.1 for the Hungarian homogenized annual data series. This is also true for spring, winter and autumn trends, but it has been found that this type of estimation of change is unacceptable for summer data. Finding an applicable higher order function system to estimate the summer trend would be an interesting topic to be investigated in a future study. The way in which inhomogeneities can misrepresent statistics has also been demonstrated, even to the extent that the opposite result is obtained after removing discontinuities. It is important to note that the standard deviation of the trend coefficient estimation in the present work depends on the variance of the given meteorological element and on the sample size, i.e. the length of the period. So the longer data series at researchers' disposal, the better the chances of correctly characterizing the direction and degree of climate change. Changes in Hungary have been shown on maps, including only those values where a significant trend was obtained as a result of the hypothesis test. In this way, it is possible to show actual changes in space, and it is also very important to know in which months and seasons and which part of the country the temperature is rising or falling (Rebetez and Reinhard 2008).

In the present study, it has been demonstrated that the detection of climate change strongly depends on the choice of statistical methods. New, more adequate methods can therefore improve our knowledge and understanding of the climate. This, in turn, requires a knowledge and understanding of mathematical statistics.

Acknowledgements Open access funding provided by Eötvös Loránd University.

Author contributions The mathematics were developed by TS, and these methods applied by BI to the meteorological analyses. BI prepared the manuscript with contributions from all co-authors. The authors applied the SDC approach for the sequence of authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Table 1

1. Estimation of change over the total period: $\hat{c}_2 \cdot (n - 1)$

2. The lower endpoint of the confidence interval, change over the total period: $\hat{c}_{2_{\alpha 1}} \cdot (n - 1)$
3. The upper endpoint of the confidence interval, change over the total period: $\hat{c}_{2_{\alpha 2}} \cdot (n - 1)$
4. The α confidence level estimation of change over the total period: $\hat{c}_{2_{\alpha}} (n - 1)$

Station number	$\hat{c}_2 \cdot (n - 1)$	$\hat{c}_{2_{\alpha 1}} \cdot (n - 1)$	$\hat{c}_{2_{\alpha 2}} \cdot (n - 1)$	$\hat{c}_{2_{\alpha}} \cdot (n - 1)$
1	1.4184	1.0395	1.7973	1.0395
2	1.3454	0.9737	1.7171	0.9737
3	1.3327	0.9573	1.7082	0.9573
4	1.1590	0.7957	1.5224	0.7957
5	1.3734	0.9954	1.7515	0.9954
6	1.2968	0.9292	1.6644	0.9292
7	1.0231	0.6518	1.3944	0.6518
8	1.2544	0.8814	1.6275	0.8814
9	1.2918	0.9215	1.6621	0.9215
10	1.3427	0.9798	1.7056	0.9798
11	1.1491	0.7696	1.5285	0.7696
12	1.3523	0.9945	1.7102	0.9945
13	1.2680	0.9001	1.6360	0.9001
14	1.2160	0.8428	1.5891	0.8428
15	1.0840	0.7003	1.4677	0.7003
16	1.1237	0.7420	1.5054	0.7420
17	1.2721	0.8943	1.6499	0.8943
18	1.2307	0.8411	1.6204	0.8411
19	1.2154	0.8460	1.5848	0.8460
20	1.1296	0.7366	1.5226	0.7366
21	1.1971	0.8073	1.5868	0.8073
22	1.1653	0.7717	1.5589	0.7717
23	1.2553	0.8710	1.6397	0.8710
24	1.3166	0.9212	1.7120	0.9212
25	1.1026	0.7301	1.4750	0.7301
Average	1.2366	0.8598	1.6135	0.8598

Annual mean temperature, trend estimation for homogenized data series (n = 118, $\alpha = 0.9$)

Table 2

1. Estimation of change over the total period: $\hat{c}_2 \cdot (n - 1)$
2. The lower endpoint of the confidence interval, change over the total period: $\hat{c}_{2_{\alpha 1}} \cdot (n - 1)$
3. The upper endpoint of the confidence interval, change over the total period: $\hat{c}_{2_{\alpha 2}} \cdot (n - 1)$
4. The α confidence level estimation of change over the total period: $\hat{c}_{2_{\alpha}} (n - 1)$

Station number	$\hat{c}_2 \cdot (n - 1)$	$\hat{c}_{2_{a1}} \cdot (n - 1)$	$\hat{c}_{2_{a2}} \cdot (n - 1)$	$\hat{c}_{2_a} \cdot (n - 1)$
1	1.7666	1.3764	2.1567	1.3764
2	1.2718	0.8800	1.6636	0.8800
3	0.5579	0.1514	0.9645	0.1514
4	0.1362	-0.2593	0.5318	0.0000
5	1.1712	0.7769	1.5655	0.7769
6	0.2146	-0.1465	0.5758	0.0000
7	0.7242	0.3381	1.1103	0.3381
8	1.1968	0.8021	1.5915	0.8021
9	1.5605	1.1827	1.9382	1.1827
10	1.8860	1.5152	2.2568	1.5152
11	0.3694	-0.0568	0.7956	0.0000
12	1.7865	1.4222	2.1509	1.4222
13	1.4521	1.0688	1.8355	1.0688
14	1.0982	0.7142	1.4822	0.7142
15	1.0986	0.7090	1.4882	0.7090
16	0.1262	-0.2725	0.5250	0.0000
17	0.4997	0.1279	0.8714	0.1279
18	0.5170	0.0665	0.9675	0.0665
19	1.3436	0.9783	1.7089	0.9783
20	1.4567	1.0666	1.8467	1.0666
21	0.3567	-0.0678	0.7813	0.0000
22	0.5959	0.1937	0.9980	0.1937
23	0.1643	-0.2731	0.6018	0.0000
24	1.1110	0.6994	1.5225	0.6994
25	1.7010	1.3246	2.0773	1.3246
Average	0.9665	0.5727	1.3603	0.6158

References

- Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J.: Guidelines on climate metadata and homogenisation. WCDMP-No 53, WMO/TD No. 1186, p. 52. World Meteorological Organization (2003)
- Alexandersson, H.: A homogeneity test applied to precipitation data. *J. Climatol.* **6**(6), 661–675 (1986)
- Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., Nieplova, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region. *Int. J. Climatol.* **27**(1), 17–46 (2007). <https://doi.org/10.1002/joc.1377>
- Begert, M., Schlegel, T., Kirchhofer, W.: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int. J. Climatol.* **25**, 65–80 (2005). <https://doi.org/10.1002/joc.1118>
- Bloomfield, P.: Trends in global temperature. *Clim. Change* **21**(1), 1–16 (1992). <https://doi.org/10.1007/BF00143250>
- Brunetti, M., Maugeri, M., Monti, F., Nanni, T.: Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *Int. J. Climatol.* **26**(3), 345–381 (2006)

- Caussinus, H., Lyazrhi, F.: Choosing a linear model with a random number of change-points and outliers. *Ann. Inst. Stat. Math.* **49**(4), 761–775 (1997)
- Caussinus, H., Mestre, O.: Detection and correction of artificial shifts in climate series. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **53**(3), 405–425 (2004)
- Della-Marta, P., Collins, D., Braganza, K.: Updating Australia's high-quality annual temperature dataset. *Aust. Meteorol. Oceanogr. J.* **53**(2), 75–93 (2004)
- Delvaux, C., Ingels, R., Vrábel, V., Journée, M., Bertrand, C.: Quality control and homogenization of the Belgian historical temperature data. *Int. J. Climatol.* **39**(1), 157–171 (2019)
- Dévényi, D., Gulyás, O.: *Matematikai statisztikai módszerek a meteorológiában*. Tankönyvkiadó (1988)
- Foster, G., Rahmstorf, S.: Global temperature evolution 1979–2010. *Environ. Res. Lett.* **6**(4), 044022 (2011)
- Gao, J., Hawthorne, K.: Semiparametric estimation and testing of the trend of temperature series. *Econom. J.* **9**(2), 332–355 (2006)
- IPCC: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge (2013)
- Jones, D., Collins, D., Nicholls, N., Phan, J., Della-Marta, P.: A new tool for tracking Australia's climate variability and change. *Bull. Aust. Meteorol. Oceanogr. Soc.* **17**, 65–69 (2004)
- Keller, C.F.: Global warming: a review of this mostly settled issue. *Stoch. Environ. Res. Risk Assess.* **23**(5), 643–676 (2009)
- Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Miletta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A.F.V., Forland, E., Mietet, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V., Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.* **22**(12), 1441–1453 (2002). <https://doi.org/10.1002/joc.773>
- Mamara, A., Argiriou, A.A., Anadranistakis, M.: Recent trend analysis of mean air temperature in Greece based on homogenized data. *Theor. Appl. Climatol.* **126**(3–4), 543–573 (2016)
- Menne, M.J., Williams, C.N., Vose, R.S.: The U.S. historical climatology network monthly temperature data, version 2. *Bull. Am. Meteorol. Soc.* **90**(7), 993–1008 (2009)
- Móri, F.T., Székely, J.G. (eds.): *Többszintű statisztikai analízis*. Műszaki Könyvkiadó, Budapest (1986)
- Mudelsee, M.: Trend analysis of climate time series: a review of methods. *Earth Sci. Rev.* **190**, 310–322 (2019)
- 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., Szentimrey, T., Salinger, J., Førland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol.* **18**(13), 1493–1517 (1998)
- Rebetez, M., Reinhard, M.: Monthly air temperature trends in Switzerland 1901–2000 and 1975–2004. *Theor. Appl. Climatol.* **91**(1), 27–34 (2008)
- Sneeyers, R. (1990): On the statistical analysis of series of observations. Technical note 143, WMO-No. 415. World Meteorological Organization (WMO) (1990)
- Spinoni, J., Szalai, S., Szentimrey, T., Lakatos, M., Bihari, Z., Nagy, A., Németh, Á., Kovács, T., Mihic, D., Dacic, M., Petrovic, P., Kržič, A., Hiebl, J., Auer, I., Milkovic, J., Štepanek, P., Zahradníček, P., Kilar, P., Limanowka, D., Pyrc, R., Cheval, S., Birsan, M.-V., Dumitrescu, A., Deak, G., Matei, M., Antolovic, I., Nejedlík, P., Štastný, P., Kajaba, P., Bochníček, O., Galo, D., Mikulová, K., Nabyvanets, Y., Skrynyk, O., Krakovskaya, S., Gnatiuk, N., Tolasz, R., Antofie, T., Vogt, J.: Climate of the Carpathian Region in the period 1961–2010: climatologies and trends of 10 variables. *Int. J. Climatol.* **35**(7), 1322–1341 (2015)
- Szalai, S., Auer, I., Hiebl, J., Milkovich, J., Radim, T., Štepanek, P., Zahradníček, P., Bihari, Z., Lakatos, M., Szentimrey, T., Limanowka, D., Kilar, P., Cheval, S., Deak, Gy., Mihic, D., Antolovic, I., Mihajlovic, V., Nejedlík, P., Štastný, P., Mikulova, K., Nabyvanets, I., Skryryk, O., Krakovskaya, S., Vogt, J., Antofie, T., Spinoni, J.: *Climate of the Greater Carpathian Region. Final technical report, European Commission - JRC* (2013)
- Szalai, S. (ed.): *Proceedings of the First Seminar for Homogenization of Surface Climatological Data. Seminar for Homogenization of Surface Climatological Data 1st session (6–12 October 1996)*

- Budapest, Hungary 1996, p. 141. Hungarian Meteorological Service. https://library.wmo.int/index.php?lvl=notice_display&id=17285#.Xk5i6kraUk. Accessed 21 Feb 2020
- Szentimrey, T.: A lineáris analitikus és a periodikus trendvizsgálat néhány elvi-módszertani kérdése. *Időjárás* **93**(2–3), 151–156 (1989)
- Szentimrey, T.: Multiple analysis of series for homogenization (MASH). In: Szalai, S., Szentimrey, T., Szinell, C. (eds.) *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data. Seminar for Homogenization of Surface Climatological Data, 2nd session (9–13 November 1998)*, Budapest, Hungary, WCDMP-No. 41, pp. 27–46. World Meteorological Organization (WMO) (1999)
- Szentimrey, T.: Development of MASH homogenization procedure for daily data. In: Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S. (eds.) *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases. Seminar for Homogenization and Quality Control in Climatological Databases 5th session (29 May–2 June 2006; Budapest, Hungary)*. WMO/TD-No. 1493; WCDMP-No. 71, pp. 123–130. World Meteorological Organization (WMO) (2006a)
- Szentimrey, T.: An overview on the main methodological questions of homogenization. In: Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S. (eds.) *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases. Seminar for Homogenization and Quality Control in Climatological Databases 5th session (29 May–2 June 2006; Budapest, Hungary)*. WMO/TD-No. 1493; WCDMP-No. 71, pp. 1–6. World Meteorological Organization (WMO) (2006b)
- Szentimrey, T.: Theoretical questions of daily data homogenization. *Időjárás* **117**(1), 113–122 (2013)
- Szentimrey, T. (2014): Manual of homogenization software MASHv3.03. In: Hungarian Meteorological Service, Budapest, Hungary (2014)
- Szentimrey, T., Bihari, Z.: *Manual of Interpolation Software MISHv1.03*, p. 89. Hungarian Meteorological Service, Budapest (2014)
- Szentimrey, T., Bihari, Z., Lakatos, M., Szalai, S.: Mathematical, methodological questions concerning the spatial interpolation of climate elements. *Időjárás* **115**(1–2), 1–11 (2011)
- Szentimrey, T., Bihari, Z., Lakatos, M.: Mathematical questions of spatial interpolation of climate variables. In: Lakatos, M., Szentimrey, T., Marton, A. (eds.) *Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques in Climatology and Meteorology 2014*. WCDMP-No. 84, pp. 107–114. World Meteorological Organization (WMO) (2014)
- Trewin, B.: Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Clim. Change* **1**, 490–506 (2010). <https://doi.org/10.1002/wcc.46>
- Trewin, B.: A daily homogenized temperature data set for Australia. *Int. J. Climatol.* **33**(6), 1510–1529 (2013)
- Venema, V.K.C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., Brandsma, T.: Benchmarking homogenization algorithms for monthly data. *Clim. Past* **8**(1), 89–115 (2012)
- 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.* **117**, D18110 (2012). <https://doi.org/10.1029/2012jd017859>
- Wooldridge, J.M.: *Introductory Econometrics: A Modern Approach*, 5th edn. South-Western, Cengage Learning, Boston (2013)
- 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.* **118**, 9708–9720 (2013). <https://doi.org/10.1002/jgrd.50791>
- Yosef, Y., Aguilar, E., Alpert, P.: Changes in extreme temperature and precipitation indices: using an innovative daily homogenized database in Israel. *Int. J. Climatol.* **39**(13), 5022–5045 (2019)
- Zhang, X., Vincent, L.A., Hogg, W.D., Niitsoo, A.: Temperature and precipitation trends in Canada during the 20th century. *Atmos. Ocean* **38**(3), 395–429 (2000)