Post-print of: Int. J. Climatol., 17 pp, https://doi.org/10.1002/joc.8080

Data quality control and homogenization of daily precipitation and air temperature (mean, max and min) time series of Ukraine

Oleg Skrynyk[1*], Vladyslav Sidenko[1], Enric Aguilar[2], Jose Guijarro[3], Olesya Skrynyk[1,4], Liudmyla Palamarchuk[1], Dmytro Oshurok[1], Valeriy Osypov[1], Volodymyr Osadchyi[1]

[1] *Ukrainian Hydrometeorological Institute, Kyiv, Ukraine*

[2] *Center for Climate Change, C3, Geography Department, Universitat Rovira i Virgili, Tarragona, Spain*

[3] *State Meteorological Agency (AEMET), Balearic Islands Office, Palma de Mallorca, Spain*

[4] *National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine*

[*] Corresponding author, 37 Nauky Ave, Kyiv, 03028, Ukraine. skrynyk@uhmi.org.ua

Abstract

In this paper we present the results of quality control and homogenization procedures applied to long time series of daily atmospheric precipitation sums (Rr) and daily mean (Tm), maximum (Tx) and minimum (Tn) air temperature collected in Ukraine. The daily data from 178 meteorological stations covering the period of 1946-2020 were analyzed. In order to perform a thorough quality assurance check, we used the R package INQC, while the Climatol homogenization software was used to detect and remove breaks from the time series. The INQC quality assurance tests revealed a relatively small number of erroneous records (around 0.01% for each variable) and suspicious values (up to 0.09%). The application of Climatol resulted in 195, 296, 355 and 359 break points, detected for Rr, Tm, Tx, and Tn, respectively. These quantities coincide roughly with the results of the HOMER homogenization procedure applied to monthly time series for the same stations and almost the same period (performed in the previous works of the authors). To verify the homogenization results, statistical comparison of the raw and homogenized time series was performed. The verification demonstrated that the quality control and homogenization procedures detected and removed errors and breaks very well, and air temperature and precipitation fields after the homogenization are more self-consistent compared to the original raw data.

KEYWORDS
daily atmospheric precipitation, daily air temperature, time series, quality control, homogenization, INQC, Climatol, Ukraine

1. INTRODUCTION

A considerable part of modern climate studies rely on direct application of empirical data collected through meteorological measurements and/or observations (Hartmann et al., 2013). Therefore, the creation of high-quality digital datasets of station time series (the longer, the better) with different time resolutions is extremely important for deepening and improving our knowledge about climate, its change and variability. However, the time series of climate data are often affected by some quality control (QC) issues (missing values, outliers, errors, etc.) and artifacts (sharp breaks or gradual trends) (Aguilar et al., 2003; WMO, 2018). The latter, usually referred to as climatological inhomogeneity or station signal, can arise as a result of various factors unrelated to climate change. The nature of their appearance is artificial and consists in changing measurement methods, changing the observer or equipment; changes in the environmental conditions around the stations or with the relocation of the station (Trewin, 2010). It is clear that such undesirable effects (QC issues and station signals) must be eliminated as much as possible before further use of the data (Aguilar et al., 2003). Consequently, the QC and homogenization of raw data are important steps in climatological research, because the direct analysis of meteorological information without the above mentioned preliminary processing can lead to incorrect results and wrong conclusions (e.g., Toreti et al., 2011).

There are numerous papers where results of QC and homogenization procedures applied on the global, national, regional or local level to station time series of various climate variables with the monthly (e.g., Li et al., 2012; Mamara et al., 2013; Mamara et al., 2014; Vertačnik et al., 2015; Prohom et al., 2015; Laapas and Venäläinen, 2017; Mamara et al.,2017; Kolendowicz et al., 2018; Menne et al., 2018; Aruffo and Di Carlo, 2019; Coll et al., 2020; Kessabi et al., 2022), daily (e.g., Kuglitsch et al., 2009; Trewin, 2012; Xu et al., 2013; Spinoni et al., 2014; Hewaarachchi et al., 2017; Squintu et al., 2018; Yosef et al., 2018; Fioravanti et al., 2019; Randriamarolaza et al., 2021; Dijkstra et al., 2021; Mateus and Potito, 2021) and even hourly (Dumitrescu et al., 2020) time resolutions are presented. For modern climatological research, time series with daily time resolution are specially precious. For instance, they are used to study climate extremes, since extreme weather phenomena are characterized by a short duration (from several hours to several days) (e.g., Moberg et al., 2006; Mistry, 2019; Yosef et al., 2019; Randriamarolaza et al., 2021).

Number of studies on the national level were performed in Ukraine to carry out QC and detection and adjustment of inhomogeneities in time series of the essential climate variables (ECV) such as precipitation and air temperature (Osadchyi et al., 2017; Skrynyk et al., 2018; Skrynyk et al., 2019; Osadchyi et al., 2022). However, only monthly data were analyzed by means of the HOMER software.

The main objective of this paper is to present results of quality assurance and homogenization procedures applied to long station time series of daily atmospheric precipitation and air temperature (mean, maximum and minimum) collected in Ukraine. For all four considered climate variables, data for 178 stations over the period of 1946-2020 were analyzed. The list of stations was the same as in the previous studies performed at the monthly scale, while the period under study was slightly extended to include the recent years. In order to achieve the objective, we used the state-of-the-art and well established software, INQC for detection of erroneous values and Climatol for break detections and shift adjustments. Our research can be considered as the next logical and consistent step in the study of climatological inhomogeneities in the national climate time series. The obtained results make it possible to carry out a comprehensive multi-scale comparative analysis of inhomogeneities detected in the time series at different time scales and by means of different software. Also, homogenized databases of long daily series of meteorological observations can be used as a solid basis for specialized geostatistical interpolation/downscaling and creation of gridded databases, and research on changes in climate extremes.

2. DATA AND METHODS

2.1. Ukrainian daily atmospheric precipitation and air temperature (mean, max and min) time series (1946-2020)

The daily data of the considered essential climate variables (Rr, Tm, Tx, and Tn) were obtained from the Central Geophysical Observatory (CGO), which is an observation institution of the Ukrainian Weather Service. It should be recalled that in Ukraine daily mean air temperature, Tm, is defined as an average of all sub-daily values measured on this day. The majority of the data were provided in digital form as Excel spreadsheets and an ACCESS data file. The list of 178 meteorological stations for all variables (Rr, Tm, Tx, and Tn) was exactly the same as in the previous works on quality check and homogenization of monthly data of Ukraine (Osadchyi et al., 2017; Skrynyk et al., 2018). It is also provided as supplementary material (SM1) to this paper. The station locations on the territory of Ukraine are shown in Figure 1. Similarly to (Osadchyi et al., 2017; Skrynyk et al., 2018), the year of 1946 was selected as the earliest year of the data due to an extremely large percentage of missing values in the time series during World War II (1939-1945). It is well known that a large number of synchronous missing data/periods is a serious obstacle for any relative homogenization procedure. After a preliminary processing and analysis of the provided digital datasets, the corresponding time series were transformed to the formats suitable for the QC and homogenization software. The preliminary analysis also revealed a fairly large number of missing values in the time series of all four variables, with the largest quantity in atmospheric precipitation observations. The two most problematic periods are 1946-1960 for the many time series collected over the whole country, and 2015-2020 for 26 stations located in the Crimea and Eastern parts of Lugansk and Donetsk regions. Unfortunately, data for the last six years from those stations are not directly available to Ukrainian scientists.

For the first period, 1946-1960, in order to complete the missing values, data rescue (digitization from paper sources) was performed by the authors during the recent couple of years. The paper sources of the daily data records (special meteorological tables and reports) were also provided by CGO. As a result of the digitization activity, 1,651,930 daily values of the climate variables were rescued and introduced for further analysis. We planned to continue the digitization until all available in the paper form information would be rescued. However, due to some circumstances (the Russian aggression and subsequent sharp reduction of financing, temporal evacuation from Kyiv of several research group members, etc.) we had to stop this work and a certain amount of values is still missing.

For the second period, 2015-2020, we tried to find data on the Internet. Fortunately, 3-hourly data (with measurements performed at 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00 UTC) for 20 stations (mainly, from the Crimea) covering 2015-2020 were found and downloaded from (Meteorological data, 2005) and then used to calculate corresponding daily values. Data from the other inaccessible climatological stations were assumed to be missing. Calculation of daily mean air temperature and daily sum of atmospheric precipitation from the sub-daily data is rather straightforward and was performed according to definitions of these parameters. Daily min/max air temperature was calculated as a minimal/maximal value of eight measurements for a particular day. Such approach can introduce some errors to Tn and Tx data, but in our case, the raw errored data are still better than missing values.

The general visualization of the data completeness in the time series is presented in Figure 2 in the form of a heat map. Here, the completeness is defined as the percentage of available values for each station and year. As can be seen from the figure, most of the missing values are still concentrated in the period of 1946-1960 (15 years). The time period 1961-2020 has, on average, a satisfactory and high amount of series completeness (more than 80-90%). It should be noted that stations with time series of almost 100% completeness are uniformly distributed over the territory of Ukraine, allowing to perform a relative homogenization by means of the Climatol software without problems. Nevertheless, the results of this study for the periods of 1946-1960 and 2015-2020 (for the Crimea stations) should be used with some caution.

2.2. Quality assurance procedure and the INQC package

Time series of meteorological observations and measurements almost always contain errors of various types and origins, which can significantly affect results of the data analysis. That is why it is necessary to apply a quality control procedure to identify and remove errors, outliers, etc.

The quality control procedure in our study was performed in two stages. The first one is a preliminary “manual” control using the built-in tools of the Excel spreadsheet and visual inspection of the digitized series. Such “manual” QC check aims to clear the series of various types of typographical errors that may have occurred during the digitization of the data. Another reason for performing this preliminary quality control check is that the data must be cleaned of non-numeric values in order to let the software be used in the subsequent steps of the QC procedure to function properly.

The next, main phase of quality control is the analysis of time series using the INQC software (Indecis Quality Control of Climatological Daily Time Series, https://CRAN.R-project.org/package=INQC). INQC is a software product that runs in the R environment and is specifically designed to perform quality control of climatological time series with the daily time resolution. Originally, the software was developed in order to QC the European Climate Assessment & Dataset station data in the frame of the INDECIS project (http://www.indecis.eu/). However, INQC's functions can be used to deal with quality problems in any other climatological data set with the daily time resolution. The package includes more than 20 fully parameterizable quality assurance functions/tests for detecting false/erroneous or suspicious values in a time series of basic climate variables such as mean, maximum and minimum air temperatures, atmospheric precipitation and pressure, relative humidity, wind speed, etc. INQC works by applying a series of tests to the data, detecting spurious values (e.g. negative precipitation or temperatures above/below certain set limits), suspicious values (e.g. extreme values that fall out of the frequency distribution of analyzed meteorological variables) and cumulatively suspicious values (which are repeated too many times in separate short periods). The list and brief description of the INQC functions, which were used to quality control the time series, is presented in Table 1. The set of functions and, accordingly, the tests applied for different meteorological variables can be different. At this stage, quality control was carried out for the entire time period of the study, namely for the years 1946-2020 for 178 stations.

During the quality control performed by means of the INQC software, each value in the time series is assigned an additional value based on the system of integer quality control flags (Table 2), which provide information about the status of the checked data for further analysis and decision-making regarding possible correction or removal of false/erroneous/suspicious value. The flags might be produced by means of different INQC tests/functions. For instance, erroneous values (flag 1) might be detected by functions 'weirddate' or 'physics'; almost certain error (flag 2) by 'jumpQUANT' or ‘newfriki’; outlier suspect (flag 3) by ‘IQRoutliers’; collectively suspect (flag 4) might come from ‘flat’ or ‘toomany’ (see Table 1 for function descriptions).

2.3. Homogenization software Climatol

In order to detect and remove station signals in the time series of daily mean, maximum and minimum air temperatures and daily sums of atmospheric precipitation of Ukraine, the software Climatol (Climate Tools: Series Homogenization and Derived Products, https://CRAN.R-project.org/package=climatol) (Guijarro, 2018) was selected. Climatol has been successfully used at the national and international levels (e.g., Azorin-Molina et al., 2019; Coll et al., 2020; Dumitrescu et al., 2020; Kessabi et al., 2022) for the homogenization of time series of different meteorological variables with the sub-daily, daily or monthly time resolution.

There are several reasons why the Climatol software was used to homogenize the daily time series of the essential climate variables of Ukraine: (1) it is capable to perform homogenization of data with a daily time resolution (e.g., Azorin-Molina et al., 2019); (2) the software has been well evaluated and verified along with other similar programs/algorithms and it has shown good results (Venema et al., 2012; Guijarro et al., 2019); (3) it can be applied automatically what significantly facilitates homogenization of large datasets; (4) uncertainties of the Climatol adjustment algorithm have been evaluated and quantified what can provide an assessment of the added value of the Climatol homogenization (Skrynyk et al., 2020); (5) it has lower (compared to other similar software) requirements for the completeness of the time series and allows to perform their homogenization even in case of a relatively large number of missing data.

The first step of the Climatol algorithm is the data normalization and filling in missing values through an iterative procedure in which the main statistical properties of time series, namely means and standard deviations, are recalculated at every iteration until their stationary/stable values are obtained. When this condition is met, the calculation of anomaly series is carried out by subtracting the corresponding composite reference series (weighted average of a prescribed number of the nearest available normalized data). This allows to evaluate the quality of the series by removing outliers exceeding the user-defined threshold value (Climatol control parameter dz.max), as well as to detect inhomogeneities in the series using the Standard Normal Homogenization

Test (SNHT) (Alexandersson, 1986; Alexandersson and Moberg, 1997; Khaliq and Ouarda, 2007; Toreti et al., 2011). Typical threshold value of dz.max for outlier removal is 5 standard deviations. In case the distribution is skewed, outliers can be removed more flexibly using an additional (dz.min) parameter, which by default is equal to dz.max. The default SNHT threshold values for both detection levels (windowed, snht1, and performed for the whole series, snht2) are the same snht1=snht2=25. However, these values are not universal, and might be changed depending on the properties of the studied series. The smaller the value of the SNHT test, the more homogeneous it is considered to be. The SNHT test is designed to detect single breakpoints. Therefore, to avoid possible errors in detecting inhomogeneities, this procedure is repeated iteratively. Every time SNHT exceeds the threshold value, the series is divided into two segments at the point with the largest value. After that, the SNHT analysis is applied again to each segment to avoid possible errors in the detection of break points. In addition to the detection of break points, another part of the homogenization procedure is the adjustment/correction of shifts, whose amplitudes are calculated using an orthogonal (type II) linear regression model. More comprehensive description of the theoretical aspects and mathematical basis of the Climatol method can be found in the user manual (Guijarro, 2018).

In our study, the homogenization of the daily time series was carried out using the following practical steps: (1) preparation and reading of input files; (2) obtaining monthly data from daily values; (3) running Climatol in an exploratory mode at the monthly level to determine the parameters for outliers and inhomogeneities detection (dz.max, snht1 and snht2); (4) homogenization (detection of breakpoints and outliers) at the monthly level; (5) estimation and automatic infilling of all missing data and correction of inhomogeneities at the daily scale based on the information about breakpoints from the previous stage; (6) export/extraction of the homogenized daily time series for further analysis.

The Climatol homogenization procedure (detection and correction) can be performed directly at the daily time resolution. However, the application of Climatol on the monthly time scale is a more effective way due to a much higher signal-to-noise ratio. As a result of the described procedure, sets of output files are usually obtained. Among them, there are the output text and r-data files, which contain quality controlled and homogenized series of the meteorological variables, series of detected inhomogeneities, detected errors, as well as a file with flags indicating information whether value is original, infilled (originally missing), or adjusted to eliminate inhomogeneities or outliers. Others are PDF files – graphical output with various diagnostic graphics. In addition, the set of output files can be easily controlled by the user with the help of a wide range of built-in Climatol functions.

In addition to the packages described above, we also used the RClimDex software (Zhang et al., 2018) to calculate yearly time series of two climate extreme indices, TN90p (warm nights) and TX90p (warm days), and the least squares method to calculate linear trends (slopes of the regression lines) in time series of either the essential climate variables or the climate extreme indices in the verification procedure of the obtained QC and homogenization results. The ordinary kriging method was used to build maps demonstrating the trend spatial distributions.

3. RESULTS

3.1. Quality control results

The first stage of the QC procedure was carried out for the periods in which the process of manual digitization of the data from paper sources was performed (mainly 1946-1960 or 1946-1975). This QC check revealed 56 rough errors. Their largest numbers were found in the series of maximum (26) and mean air temperature (18), slightly smaller numbers in the series of daily precipitation sums (5) and minimum temperature (7). The found errors were corrected based on the original paper sources (tables/reports).

The results of INQC application to the daily time series of Ukraine are presented in Table 3. As can be seen from the table, the number of the data values that passed quality control successfully is greater than 88% for each variable. The percentage was calculated based on the total quantity of the time series members, including missing values. The numbers of detected true errors (marked by INQC with the integer flag “1”) range from 15 to 502 depending on the variable, however, in percentage rate these values do not exceed 0.01%. Larger numbers of values fell into the categories of probable errors, suspicious outliers, and collectively suspects. In this case, the corresponding percentages reach 0.09%, however, in absolute terms, these numbers are quite significant. For instance, 4,275 collectively suspects were revealed by INQC in the

Rr time series. The percentage of missing data is approximately 10-11.5% for each ECV considered.

Given the circumstances mentioned above and due to quite large numbers of detected suspicious values, it was impossible to check their correctness by comparing with original records in published meteorological tables and reports. However, we checked all detected true errors, which, probably, have the most harmful influence on the daily time series and, for instance, further calculation of climate extreme indices. The summary of our check is presented in Table 4. In addition, the QC procedure revealed a large number (7,305) of errors/suspects in precipitation time series for one of the stations (these cases are not included in Table 4). After comparison with corresponding daily values reporter in paper sources, it was discovered that for this station decimal separator was omitted during the period of 1946-1965. It is worth also noting that the Climatol software is also capable of detecting and removing outliers. Therefore, all suspicious values detected by INQC are subject to one more check on the next stage of the data processing.

3.2. Climatol parameters testing

In order to tune the Climatol software to remove inhomogeneities from the time series as much as possible, it is necessary to choose the proper threshold values for outliers detection/removing (dz.max) and the Standard Normal Homogenization Test (snht1, snht2). As it was mentioned above, the default values of these parameters are not universal and must be changed depending on the properties of the studied series. According to the recommendations (Guijaro, 2018), the best practice to select proper values of the tunable parameters is to use the results of Climatol runs in an exploratory mode. For example, analyzing the results of the Climatol software applied in the exploratory mode to the monthly mean air temperature (namely, the histogram of normalized anomalies shown in Fig. 3a), it becomes clear that the default value dz.max=5 is not the best fit. Using this value will remove a significant amount of data from the time series that are not outliers. In this case, the value dz.max=10 appears to be more correct, since the following groups of values are separated by visible minima. As for the snht1 and snht2 parameters, they can be determined based on the histograms of the maximum windowed SNHT (Fig. 3b) and the maximum global SNHT (Fig. 3c), respectively. According to the recommendations of the Climatol user manual, a visual analysis of these figures suggests that snht1 and snht2 should be set to 55 and 220, respectively. However, only 15 breaks were detected with such parameters. For the data set consisting of 178 time series defined over the period of 75 years (1946-2020), this amount of break points seems to be too small and probably incomplete. Especially if the fact is taken into account that the number of influencing events (station relocations, replacing measuring devices and/or methodology etc.) which were reported in the station metadata is much higher (namely, 221). Performing homogenization with such values of snht1 and snht2 parameters, the results will not be of sufficient quality, and the obtained series will still contain a significant amount of inhomogeneities.

In order to define the optimal values for the tunable Climatol parameters, we performed sensitivity tests on the monthly scale and compared the results with the previously published papers (Osadchyi et al., 2017; Skrynyk et al., 2018; Osadchyi et al., 2022) and the metadata collected from historical descriptions of the considered stations. For each climate variable under study, three Climatol runs were performed with different values of the snht1 and snht2 parameters, which were gradually reduced by 50% from test to test. Their initial values were obtained from the Climatol runs in the exploratory mode. At the same time, the dz.max parameter remained the same for all numerical experiments and its value was also defined from the exploratory Climatol run. The last, fourth test, was performed with the snht1 and snht2 values, which provide results (e.g., numbers of break points) which are comparable and consistent with the corresponding ones obtained previously on the monthly scale. The brief summary of the performed tests is presented in Table 5. The outputs of the last Climatol runs were accepted as the main results of the QC and homogenization procedures applied to the daily time series of ECV (Rr, Tm, Tx, and Tn) of Ukraine.

3.3. Homogenization results

Below we provide some details and statistical information characterizing the homogenization results. Statistical comparison of the raw and corresponding homogenized time series, which can be considered as verification of the performed homogenization, is also presented.

3.3.1. Air temperature data

The percentage of original values in the quality controlled and homogenized time series of daily mean temperature is 53.4%, while the remaining 46.6% are values obtained or changed as a result of the homogenization and quality control procedures. Among them, 11.1% are infilled originally missing data and 36.6% are corrected values. The last group includes both the values changed due to detecting breaks and consequent shift adjusting and the re-infilled values that were previously removed as outliers. In the homogenized time series of Tn and Tx, the numbers of original values are somewhat smaller, 45.4% and 45.6%, respectively. The number of infilled missing values is approximately at the same level as for Tm series, 10.5% and 10.4%, while the numbers of corrected values increased to 44.2% and 44.0%, respectively.

The total number of the detected break points for the daily mean temperature data is 296, which means 1.66 breaks per station on average. In the Tn and Tx time series, 359 and 355 break points were found, respectively (approximately 2 breaks per station for both variables). The distribution of stations in respect to the number of the detected break points is presented in Figure 4. No breakpoints were found in 52 (29.2%), 38 (21.4%), and 33 (18.5%) series of Tm, Tx, and Tn, respectively. All of these three datasets have between 55% and 66% of stations with the number of breakpoints from 1 to 3. In the Tm time series, the largest number of breaks (5) was found in 8 stations. The minimum and maximum air temperature series have one station each with a maximum number of 7 and 8 breaks, respectively.

The time distribution of the detected inhomogeneities is presented in Figure 5. For the Tm data, there were found about 3.9 break points per year, and 4.8 and 4.7 for the datasets of minimum and maximum daily temperatures. The number of years without detected breaks is 5, 6 and 4 for Tm, Tx, and Tn, respectively. The vast majority of years (from 42% to 54%) have between 1 and 5 gaps. In the series of mean temperature, the maximum number of breaks was detected in 1988 (17), in the series of minimum temperature - in 1987 (16), the maximum - in 2014 (13).

Figure 6 shows a spatial distribution of the climatological stations used in the study and numbers of the detected breaks in corresponding time series. As can be seen from these figures, only Tm and Tx have slightly similar spatial patterns of the numbers of the detected breaks.

3.3.2. Atmospheric precipitation data

In the obtained series of daily sums of atmospheric precipitation, the number of original values is considerably larger, compared to air temperature values, and reaches 81.2%. This can be explained by the significantly lower number of the detected breaks compared to the air temperature time series. The remaining 18.8% are infilled missing values (11.1%) and corrected shifts and outliers (7.7%). None of the stations has 100% original series. However, the number of stations with more than 90% of the original data is significantly greater than in the temperature series and reaches 57 stations.

In the series of daily atmospheric precipitation, 195 break points were found, which is approximately 1.1 breaks per station. The distribution of stations according to the number of break points is shown in Fig. 4a. For 83 (46.6%) series no break points were detected. 46.6% of series have 1 to 3 breaks. The remaining 6.7% of series have 4 to 6 breaks. For Rr, on average, there are about 2.6 break points per one year (Fig. 5a). The number of years without detected breaks is 15 (15%). The vast majority of years (54%) have between 1 and 5 gaps. Despite the reasonable assumption that the earlier years of the period under study should have a greater number of heterogeneities, the maximum number of breaks was detected in 2018 (11).

It is interesting to note that the numbers of matches of detected break points with station metadata (within one-year span around reported influencing events) are 8.1, 23.5, 21.3, and 35.7% for Rr, Tm, Tx, and Tn, respectively. While the numbers of matches with break points detected by the HOMER software are 33.2, 14.5, 15.8 and 17.8% for Rr, Tm, Tx, and Tn, respectively. Therefore, despite the rather close total numbers of the detected breaks by means of Climatol and HOMER, their distributions in time and over stations are quite different.

3.3.3. Verification of the QC and homogenization results

In order to verify the results of the quality control and homogenization procedures, we performed statistical comparison of raw and quality controlled and homogenized time series. Based on the general ideas, it is clear that removal of errors and breaks (with subsequent shifts adjustment) should lead to more self-consistent (more homogeneous in time and space) fields of the ECVs. One of the ways to show this is to calculate linear trends of the time series before and after QC/homogenization and compare their statistical and spatial distributions.

In Figure 7, we present box-plots of linear trends (slopes of linear approximation lines) calculated for annual time series of Rr, Tm, Tx, and Tn before and after the data processing. The similar figures were also obtained for the data with the daily time resolution, but they were not included in the text. In the raw time series, all missing values were completed before the trend calculation by means of Climatol run with no detected break points. This procedure was done in order to have comparable time series (with the same length). As can be seen from the figure, the QC and homogenization significantly reduced the width of the trend distributions for all variables and removed the largest trend outliers. Such reduction means that the homogenized data are more self-consistent compared to the raw ones. Therefore, for time series of the climate variable the added value of the performed data processing is noticeable.

Since daily Rr, Tm, Tx, and Tn data are often used to calculate climate extreme indices and detect climate change in extremes (e.g., Randriamarolaza et al., 2021; Sidenko 2022), it is interesting to see how the quality control and homogenization influence the trend calculation in time series of such indices. As an example, we calculated the yearly time series of TN90p (warm nights) and TX90p (warm days) by means of the RClimDex software (Zhang et al., 2018) and computed corresponding linear trends. Their spatial distributions over the domain of Ukraine before and after the QC/homogenization procedures are presented in Figure 8. Figure 9 shows box-plots of the calculated trends characterizing their statistical distributions. As can be seen from Figure 8, for the climate extreme indices the data quality check and homogenization also substantially reduced spatial inhomogeneities (especially in TX90p time series), making regular changes in the climate extremes more reasonable and consistent. However, a certain amount of inhomogeneities (isolated areas/spots in the trend patterns), which are difficult to explain from the physical point of view, is still present. The probable reason for the remaining inhomogeneities might be the relatively large numbers of missing values in the time series (mainly the first fifteen years of the period under study). As can be seen from Figure 9, the box-plots of the calculated trends for TN90p and TX90p were also reduced but not so noticeably as for the time series of the ECVs.

4. CONCLUSION

In this paper, we present the results of the quality assurance and homogenization procedures applied to the long daily time series of atmospheric precipitation and air temperature (mean, max and min) collected in Ukraine. The data of 178 stations which constitute the almost entire modern national monitoring system were processed and analyzed over the 75-year period, 1946-2020. The data sets were provided by the official observational institution of the Ukrainian Weather Service, the Central Geophysical Observatory, in the digital form. However, the series of each considered climate variable contained a relatively large number of missing values. In order to increase time series completeness, data rescue activities were performed which resulted in 1,651,930 digitized daily values (mainly for the period of 1946-1960). Besides, a certain amount of daily data for 2014-2020 for the Crimea meteorological stations were obtained from the sub-daily values downloaded from the free Internet source.

The quality control check, performed by the well established software INQC, revealed the relatively small number of errors (not more than 0.01% for each of ECVs) and the slightly higher number of suspicious data (0.09%). All detected errors were compared with available paper sources and were corrected/confirmed/removed depending on the comparison results.

The homogenization procedure, carried out by means of the well known R package Climatol, detected 195, 296, 355 and 359 breaks in Rr, Tm, Tx and Tn time series, respectively. Such numbers coincide roughly with similar homogenization results obtained with the HOMER software for the same stations and almost the same period but for the data with the monthly time resolution. However, only 33.2, 14.5, 15.8 and 17.8% of the breaks detected by Climatol for Rr, Tm, Tx and Tn data respectively, coincide with ones detected by HOMER. The reasons for such discrepancies should be studied additionally in the future.

In order to verify the QC/homogenization results, we performed statistical comparison of the raw and homogenized time series. The comparison was conducted for linear trends calculated for both ECVs on the yearly scale and two climate extreme indices, TN90p and TX90p. The verification procedure demonstrated that the quality control and homogenization procedures detected and removed errors and breaks very well, and air temperature and precipitation fields after the homogenization are more self-consistent compared to the original raw data.

ACKNOWLEDGEMENTS

The authors are grateful to two anonymous reviewers for careful reading of the manuscript and valuable comments and suggestions they have made.

ORCID

Oleg Skrynyk https://orcid.org/0000-0001-8827-0280

Vladyslav Sidenko https://orcid.org/0000-0002-4143-2913

Enric Aguilar https://orcid.org/0000-0002-8384-377X

Jose Guijarro https://orcid.org/0000-0002-9527-9758

Olesya Skrynyk https://orcid.org/0000-0003-0332-5073

Liudmyla Palamarchuk https://orcid.org/0000-0001-9906-8870

Dmytro Oshurok https://orcid.org/0000-0003-1192-3823

Valeriy Osypov https://orcid.org/0000-0003-4853-8021

Volodymyr Osadchyi https://orcid.org/0000-0002-0428-4827

REFERENCES

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. and Wieringa, J. (2003) WMO guidelines on climate metadata and homogenization. WCDMP No.53, WMO-TD No. 1186. Geneva, Switzerland: WMO

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

Alexandersson, H. and Moberg, A. (1997) Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology*, 17(1), 25–34. https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J

Aruffo, E. and Di Carlo, P. (2019). Homogenization of instrumental time series of air temperature in Central Italy (1930-2015). *Climate Research*, 77(3), 193–204. http://doi.org/10.3354/cr01552

Azorin-Molina, C., Guijarro, J.A., McVicar, T.R., Trewin, B.C., Frost, A.J. and Chen, D. (2019) An approach to homogenize daily peak wind gusts: an application to the

Australian series. *International Journal of Climatology*, 39(4), 2260–2277. https://doi.org/10.1002/joc.5949

Coll, J., Domonkos, P., Guijarro, J., Curley, M., Rustemeier, E., Aguilar, E., Walsh, S. and Sweeney, J. (2020). Application of homogenization methods for Ireland's monthly precipitation records: comparison of break detection results. *International Journal of Climatology*, 40(14), 6169–6188. http://doi.org/10.1002/joc.6575

Dijkstra, F., de Vos, R., Ruis, J. and Crok, M. (2021). Reassessment of the homogenization of daily maximum temperatures in the Netherlands since 1901. *Theoretical and Applied Climatology*, 147(3-4), 1185–1194. http://doi.org/10.1007/s00704-021-03887-4

Dumitrescu, A., Cheval, S. and Guijarro, J.A. (2020). Homogenization of a combined hourly air temperature dataset over Romania. *International Journal of Climatology*, 40(5), 2599–2608. http://doi.org/10.1002/joc.6353

Fioravanti, G., Piervitali, E. and Desiato, F. (2019). A new homogenized daily data set for temperature variability assessment in Italy. *International Journal of Climatology*, 39(15), 5635–5654. http://doi.org/10.1002/joc.6177

Guijarro, J.A. (2018) Homogenization of climatic series with Climatol. Version 3.1.1. Guide. Available at https://www.climatol.eu/homog_climatol-en.pdf. [Accessed 22 November 2022]

Guijarro, J.A., Aguilar, E., Domoncos, P., Sigró, J., Štepánek, P., Venema, V. and Zahradnícek, P. (2019) Benchmarking results of the homogenization of daily essential climatic variables within the INDECIS project. Vienna, Austria: General Assembly of the European Geosciences Union, pp. 7–12, April 2019, EGU2019-10896-1

Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A., Soden, B.J., Thorne, P.W., Wild, M. and Zhai, P.M. (2013) Observations: atmosphere and surface. In: Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York. NY: Cambridge University Press

Hewaarachchi, A.P., Li, Y., Lund, R. and Rennie, J. (2017). Homogenization of daily temperature data. *Journal of Climate*, 30(3), 985–999. http://doi.org/10.1175/jcli-d-16-0139.1

Khaliq, M.N. and Ouarda, T.B.M.J. (2007). On the critical values of the standard normal homogeneity test (SNHT). *International Journal of Climatology*, 27(5), 681–687. http://doi.org/10.1002/joc.1438

Kessabi, R., Hanchane, M., Guijarro, J.A., Krakauer, N.Y., Addou, R., Sadiki, A. and Belmahi, M. (2022). Homogenization and trends analysis of monthly precipitation series in the Fez-Meknes region, Morocco. *Climate*, 10(5), 64. http://doi.org/10.3390/cli10050064

Kolendowicz, L., Czernecki, B., Półrolniczak, M., Taszarek, M., Tomczyk, A.M. and Szyga-Pluga, K. (2018). Homogenization of air temperature and its long-term trends in Poznań (Poland) for the period 1848–2016. *Theoretical and Applied Climatology*, 136(3-4), 1357–1370. http://doi.org/10.1007/s00704-018-2560-z

Kuglitsch, F.G., Toreti, A., Xoplaki, E., Della-Marta, P.M., Luterbacher, J. and Wanner, H. (2009). Homogenization of daily maximum temperature series in the Mediterranean. *Journal of Geophysical Research*, 114(D15). http://doi.org/10.1029/2008jd011606

Laapas, M. and Venäläinen, A. (2017). Homogenization and trend analysis of monthly mean and maximum wind speed time series in Finland, 1959-2015. *International Journal of Climatology*, 37(14), 4803–4813. http://doi.org/10.1002/joc.5124

Li, Q., Peng, J. and Shen, Y. (2012). Development of China homogenized monthly precipitation dataset during 1900–2009. *Journal of Geographical Sciences*, 22(4), 579–593. http://doi.org/10.1007/s11442-012-0948-8

Mamara, A., Argiriou, A.A. and Anadranistakis, M. (2013). Homogenization of mean monthly temperature time series of Greece. *International Journal of Climatology*, 33(12), 2649–2666. http://doi.org/10.1002/joc.3614

Mamara, A., Argiriou, A.A., and Anadranistakis, M. (2014). Detection and correction of inhomogeneities in Greek climate temperature series. *International Journal of Climatology*, 34(10), 3024–3043. https://doi.org/10.1002/joc.3888

Mamara, A., Argiriou, A.A. and Anadranistakis, M. (2017). Homogenization of Precipitation Series in Greece. In: Karacostas, T., Bais, A., Nastos, P. (eds)

Perspectives on Atmospheric Sciences. Springer Atmospheric Sciences. Springer, Cham. https://doi.org/10.1007/978-3-319-35095-0_83

Mateus, C. and Potito, A. (2021). Development of a quality-controlled and homogenized long-term daily maximum and minimum air temperature network dataset for Ireland. *Climate*, 9(11), 158. http://doi.org/10.3390/cli9110158

Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J., Lawrimore, J.H. (2018) The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, 31, 9835–9854. http://doi.org/10.1175/JCLI-D-18-0094.1

Meteorological data [homepage on the internet]. 2005. Available from: http://meteomanz.com/. [Accessed July 22, 2022]

Mistry, M. (2019). A high-resolution global gridded historical dataset of climate extreme Indices. *Data*, 4(1), 41. http://doi.org/10.3390/data4010041

Moberg, A., Jones, P.D., Lister, D., Walther, A., Brunet, M., Jacobeit, J., Alexander, L.V., Della-Marta, P.M., Luterbacher, J., Yiou, P., Chen, D., Klein Tank, A.M.G., Saladié, O., Sigró, J., Aguilar, E., Alexandersson, H., Almarza, C., Auer, I., Barriendos, M. and Begert, M. (2006). Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *Journal of Geophysical Research*, 111(D22). http://doi.org/10.1029/2006jd007103

Osadchyi, V., Skrynyk, O., Radchenko, R. and Skrynyk, O. (2017). Homogenization of Ukrainian air temperature data. *International Journal of Climatology*, 38(1), 497–505. http://doi.org/10.1002/joc.5191

Osadchyi, V., Skrynyk, O., Palamarchuk, L., Skrynyk, O., Osypov, V., Oshurok, D. and Sidenko, V. (2022). Dataset of gridded time series of monthly air temperature (min, max, mean) and atmospheric precipitation for Ukraine covering the period of 1946-2020. *Data in Brief*, 44, 108553. http://doi.org/10.1016/j.dib.2022.108553

Prohom, M., Barriendos, M. and Sanchez-Lorenzo, A. (2015). Reconstruction and homogenization of the longest instrumental precipitation series in the Iberian Peninsula (Barcelona, 1786-2014). *International Journal of Climatology*, 36(8), 3072–3087. http://doi.org/10.1002/joc.4537

Randriamarolaza, L.Y.A., Aguilar, E., Skrynyk, O., Vicente-Serrano, S.M. and Domínguez-Castro, F. (2021). Indices for daily temperature and precipitation in

Madagascar, based on quality-controlled and homogenized data, 1950-2018. *International Journal of Climatology*, 42(1). http://doi.org/10.1002/joc.7243

Sidenko, V.P. (2022). Climatic studies of extreme weather conditions, events and phenomena in Ukraine and the world. *Hydrology, hydrochemistry and hydroecology*, 2(64), 53–71. http://doi.org/10.17721/2306-5680.2022.2.5 (In Ukrainian)

Skrynyk, O.Y., Aguilar, E., Skrynyk, O.A., Sidenko, V., Boichuk, D. and Osadchyi, V. (2018). Quality control and homogenization of monthly extreme air temperature of Ukraine. *International Journal of Climatology*, 39(4), 2071–2079. http://doi.org/10.1002/joc.5934

Skrynyk, O.A., Boichuk, D.O. and Sidenko, V.P. (2019). Detection and removal of inhomogeneity in time series of climatological variables. *Hydrology, hydrochemistry and hydroecology*, 2 (53), 88–100. (In Ukrainian)

Skrynyk, O., Aguilar, E., Guijarro, J., Randriamarolaza, L.Y.A. and Bubin, S. (2020). Uncertainty evaluation of Climatol's adjustment algorithm applied to daily air temperature time series. *International Journal of Climatology*, 41(S1). http://doi.org/10.1002/joc.6854

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., Štepánek, P., Zahradnícek, P., Kilar, P., Limanowka, D. and Pyrc, R. (2014). Climate of the Carpathian Region in the period 1961-2010: climatologies and trends of 10 variables. *International Journal of Climatology*, 35(7), 1322–1341. http://doi.org/10.1002/joc.4059

Squintu, A.A., van der Schrier, G., Brugnara, Y. and Klein Tank, A. (2018). Homogenization of daily temperature series in the European Climate Assessment & Dataset. *International Journal of Climatology*, 39(3), 1243–1261. http://doi.org/10.1002/joc.5874

Toreti, A., Kuglitsch, F.G., Xoplaki, E., Della-Marta, P.M., Aguilar, E., Prohom, M. and Luterbacher, J. (2011). A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *International Journal of Climatology*, 31(4), 630–632. http://doi.org/10.1002/joc.2088

Trewin, B. (2010) Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Climate Change*, 1(4), 490–506. https://doi.org/10.1002/wcc.46

Trewin, B. (2012). A daily homogenized temperature data set for Australia. *International Journal of Climatology*, 33(6), 1510–1529. http://doi.org/10.1002/joc.3530

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. and Marinova, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1), 89–115. http://doi.org/10.5194/cp-8-89-2012

Vertačnik, G., Dolinar, M., Bertalanič, R., Klančar, M., Dvoršek, D. and Nadbath, M. (2015). Ensemble homogenization of Slovenian monthly air temperature series. *International Journal of Climatology*, 35(13), 4015–4026. http://doi.org/10.1002/joc.4265

World Meteorological Organization (2018). *Guide to Climatological Practices*. World Meteorological Organization. WMO-No 100. Geneva, Switzerland. ISBN 978-92-63-10100-6

Xu, W., Li, Q., Wang, X.L., Yang, S., Cao, L. and Feng, Y. (2013). Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *Journal of Geophysical Research: Atmospheres*, 118(17), 9708–9720. http://doi.org/10.1002/jgrd.50791

Yosef, Y., Aguilar, E. and Alpert, P. (2018). Detecting and adjusting artificial biases of long-term temperature records in Israel. *International Journal of Climatology*, 38(8), 3273–3289. http://doi.org/10.1002/joc.5500

Yosef, Y., Aguilar, E. and Alpert, P. (2019). Changes in extreme temperature and precipitation indices: using an innovative daily homogenized database in Israel. *International Journal of Climatology*, 39(13), 5022–5045. http://doi.org/10.1002/joc.6125

Zhang, X., Feng, Y., Chan, R. (2018) Introduction to RClimDex v1.9. Guide. Climate research Division, Environment Canada, Downsview Ontario, Canada

Table 1. The main INQC functions/tests used for quality control of the series. The sign ‘+’ (‘-’) means a test was (was not) applied to a variable

#	Function name	Variables to which it is applied				Function definition
		Rr	Tm	Tx	Tn	
1	weirddate	+	+	+	+	Finds invalid dates (such as 1946.01.32 or 1946.13.01) or years outside the range set by the first and last records in the file
2	duplas	+	+	+	+	Detects duplicate dates.
3	roundprecip	+	-	-	-	Splits data by month and checks if a decimal value is repeated too many times.
4	repeatedvalue	+	-	-	-	Tracks values which repeat too many times and, given the typical decaying distribution of the variable (designed for precipitation) are considered too large to repeat that many times.
5	drywetlong	+	-	-	-	Detects episodes of too many consecutive wet or dry days. Uses a peak over threshold approach and a Pareto distribution fit over the observed sequences.
6	physics	+	+	+	+	Compares the value to a specified threshold, considered to be the limit of physically possible values. In some cases this threshold is observational (e.g. 60°C), in others it is derived from the nature of the variable (e.g. 0 mm).
7	suspecta-cumprec	+	-	-	-	Detects values above a threshold preceded by a given number of “no precipitation days”
8	paretogadget	+	-	-	-	Returns the positions exceeding the value corresponding to a return period based on Pareto distribution and peak over threshold approach
9	jumpABS	+	+	+	+	Labels interdiurnal differences in absolute values considered to be too large.
10	jumpQUANT	+	+	+	+	Labels interdiurnal differences in quantiles considered to be too large.
11	flat	+	+	+	+	Detects consecutive equal values. Can be adapted to detect consecutive equal decimal part of the values
12	newfriki	+	+	+	+	Isolates extreme values which are not continuous in the distribution. If the gap is larger than a pre-set big margin, the value is flagged.
13	IQRoutliers	+	+	+	+	Computes outliers centralized around a day, using a number of days around it and based on the interquartile range. Creates a tolerance interval centered around each day of the year, using all the present values in the empirical distribution for the designed window. Values outside the interval are flagged as outliers.
14	toomany	+	+	+	+	Splits data by month or year and looks if a value is repeated too many times
15	rounding	+	+	+	+	Splits data by month and looks if a decimal value is repeated too many times
16	txtn	-	-	+	+	Compares daily maximum and daily minimum temperature and flags those values where Tx is larger or equal than Tn

Table 2. INQC quality control flags and their meaning

QC Flag	Meaning
0	Passed QC
1	Error
2	Almost certain, error
3	Outlier, suspect
4	Collectively suspect
9	Missing value

Table 3. The number/percentage of the time series members that passed INQC tests or were detected as erroneous or suspicious values

QC	Variable							
	Rr		Tm		Tx		Tn	
	Number of values	%	Number of values	%	Number of values	%	Number of values	%
Passed QC	4,326,684	88.73	4,383,669	89.9	4,361,722	89.45	4,360,942	89.43
Error	15	<0.01	87	<0.01	166	<0.01	502	0.01
Almost certain, error	1,731	0.04	2,162	0.04	1,414	0.03	838	0.02
Outlier, suspect	3,926	0.08	3,236	0.07	3,108	0.06	3,344	0.07
Collectively suspect	4,275	0.09	200	<0.01	238	<0.01	352	0.01
Missing value	539,501	11.06	486,778	9.98	509,484	10.45	510,154	10.46
Total amount	4,876,132	100	4,876,132	100	4,876,132	100	4,876,132	100

Table 4. The number of checked, corrected, confirmed and deleted by INQC true errors

Value	Variable				Total
	Rr	Tm	Tx	Tn	
Checked	15	87	166	502	770
Corrected	14	61	90	389	554
Confirmed	-	26	76	111	213
Deleted	1	-	-	2	3

Table 5. The number of breakpoints detected with different Climatol parameters compared to breakpoints found in the previous studies (Osadchyi et al., 2017; Skrynyk et al., 2018; Osadchyi et al., 2022) and station metadata

<table>
<tr><th rowspan="2">ECV</th><th colspan="4">Tunable Climatol parameters and detected breakpoints in four sensitivity tests</th><th rowspan="2">Detected breakpoints (HOMER)</th><th rowspan="2">Reported influencing events (metadata)</th></tr>
<tr><th>test #1</th><th>test #2</th><th>test #3</th><th>test #4</th></tr>
<tr><td rowspan="2">Rr</td><td>dz.max=11
snht1=35
snht2=80</td><td>dz.max=11
snht1=18
snht2=40</td><td>dz.max=11
snht1=9
snht2=20</td><td>dz.max=11
snht1=13
snht2=10</td><td rowspan="2">187</td><td rowspan="8">221</td></tr>
<tr><td>8</td><td>42</td><td>413</td><td>195</td></tr>
<tr><td rowspan="2">Tm</td><td>dz.max=10
snht1=55
snht2=220</td><td>dz.max=10
snht1=28
snht2=110</td><td>dz.max=10
snht1=14
snht2=55</td><td>dz.max=10
snht1=15
snht2=35</td><td rowspan="2">304</td></tr>
<tr><td>15</td><td>55</td><td>329</td><td>296</td></tr>
<tr><td rowspan="2">Tx</td><td>dz.max=12
snht1=70
snht2=250</td><td>dz.max=12
snht1=35
snht2=125</td><td>dz.max=12
snht1=18
snht2=63</td><td>dz.max=12
snht1=20
snht2=65</td><td rowspan="2">367</td></tr>
<tr><td>8</td><td>84</td><td>445</td><td>355</td></tr>
<tr><td rowspan="2">Tn</td><td>dz.max=8
snht1=55
snht2=280</td><td>dz.max=8
snht1=28
snht2=140</td><td>dz.max=8
snht1=14
snht2=70</td><td>dz.max=8
snht1=16
snht2=32</td><td rowspan="2">377</td></tr>
<tr><td>19</td><td>74</td><td>411</td><td>359</td></tr>
</table>

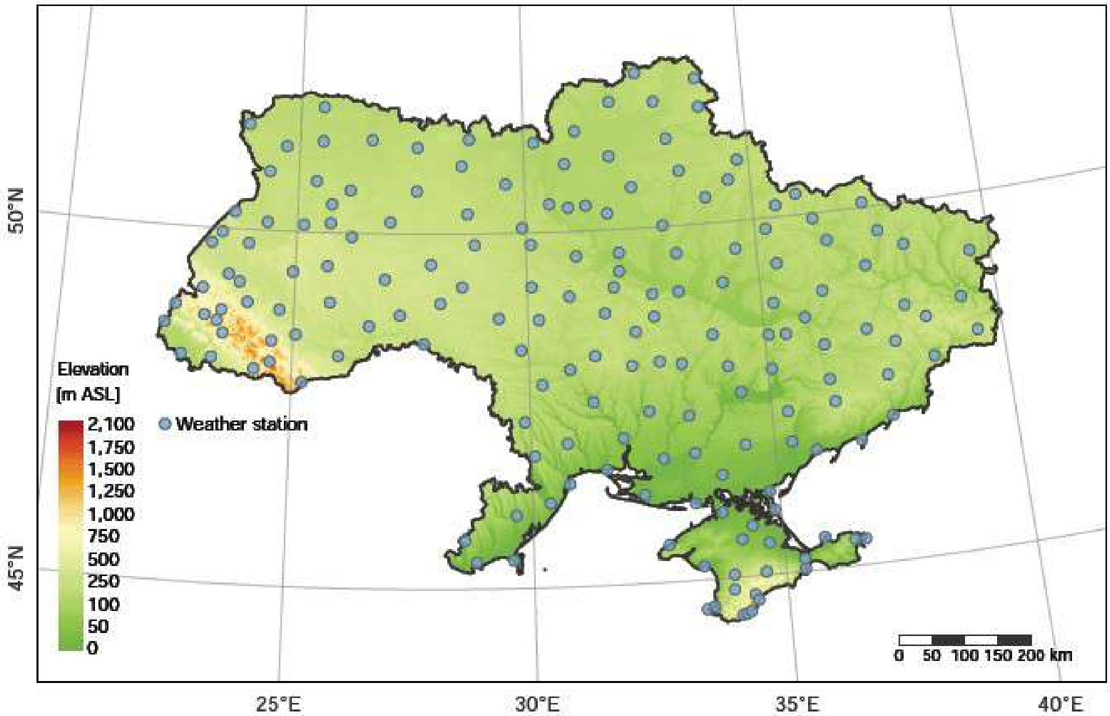

FIGURE 1. Locations of the climatological/weather stations of Ukraine used in the study

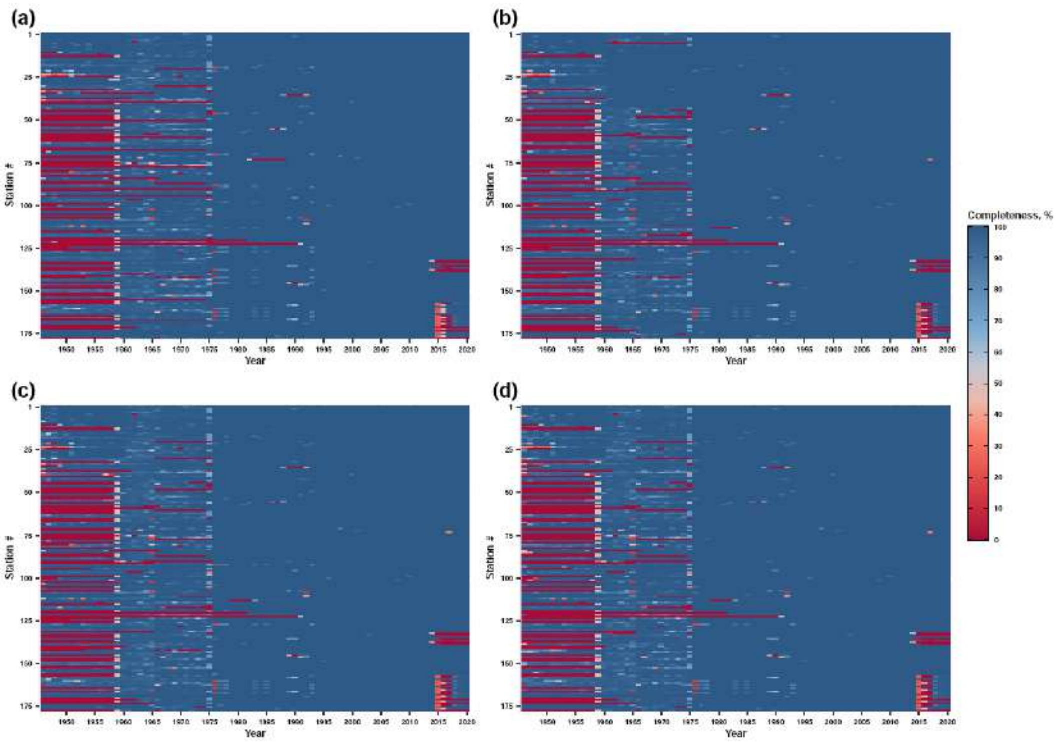

FIGURE 2. Heat maps of the daily time series completeness after the data rescue: (a) atmospheric precipitation; (b) mean air temperature; (c) maximum air temperature; (d) minimum air temperature. Station IDs (#) along the vertical axes are provided according to the station list (SM1).

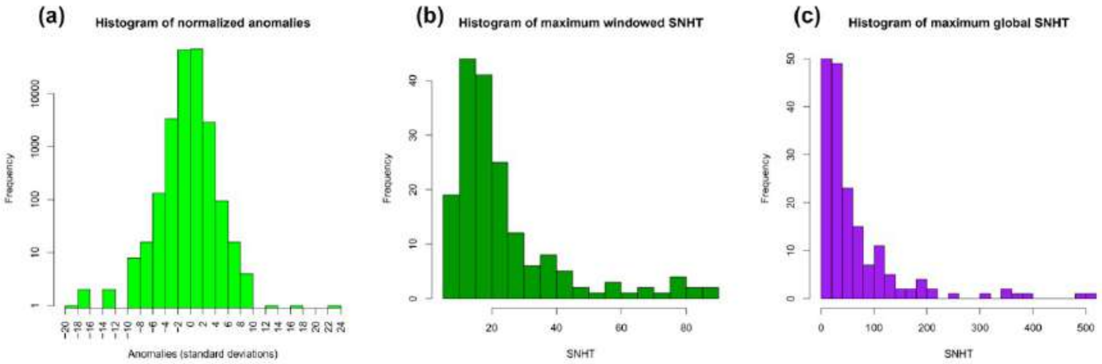

FIGURE 3. The results of Climatol applied in the exploratory mode to the monthly mean air temperature time series.

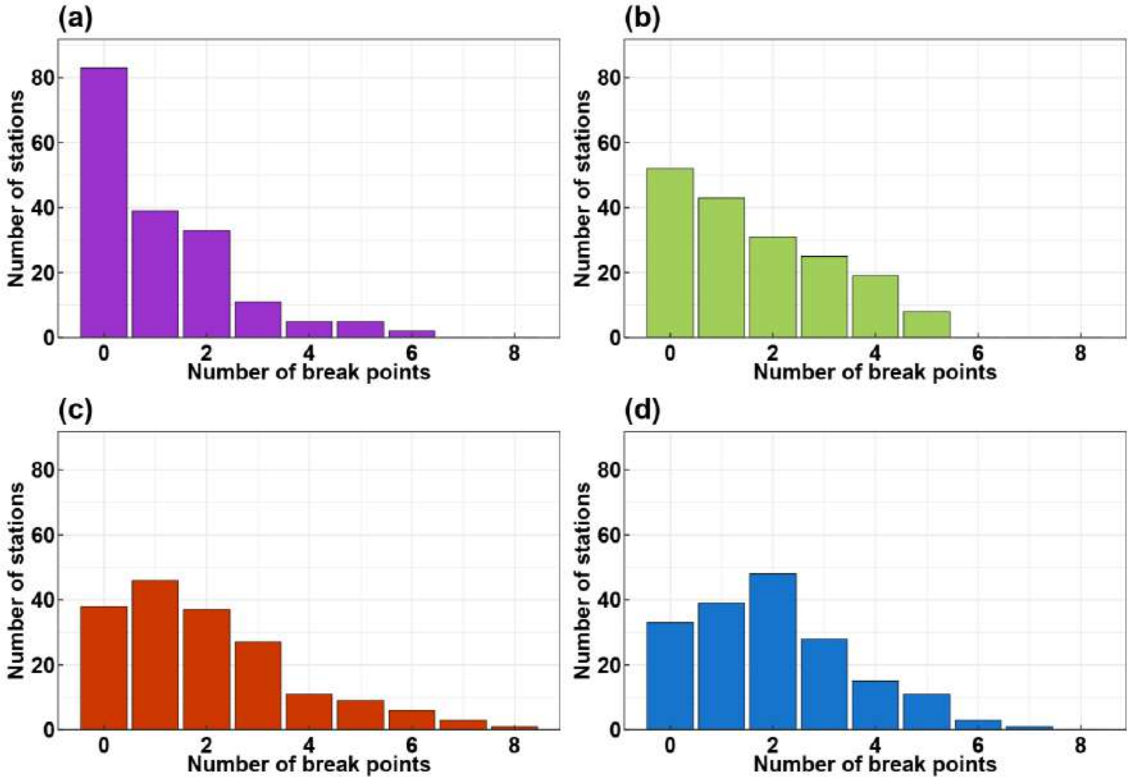

FIGURE 4. Distribution of stations with respect to the number of break points detected in the time series of: (a) Rr, (b) Tm, (c) Tx and (d) Tn.

FIGURE 5. The number of breakpoints per year for the period of 1946-2020 in the time series of: (a) Rr, (b) Tm, (c) Tx and (d) Tn.

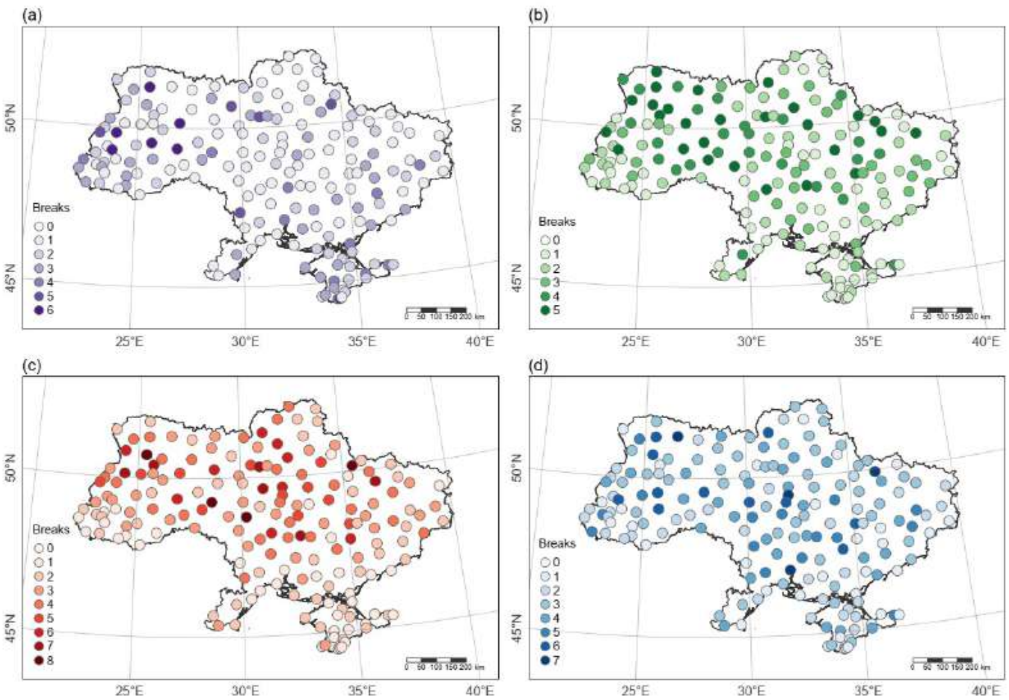

FIGURE 6. Spatial distribution of the climatological stations on the territory of Ukraine and numbers of the detected break points in the corresponding time series: (a) Rr, (b) Tm, (c) Tn and (d) Tx.

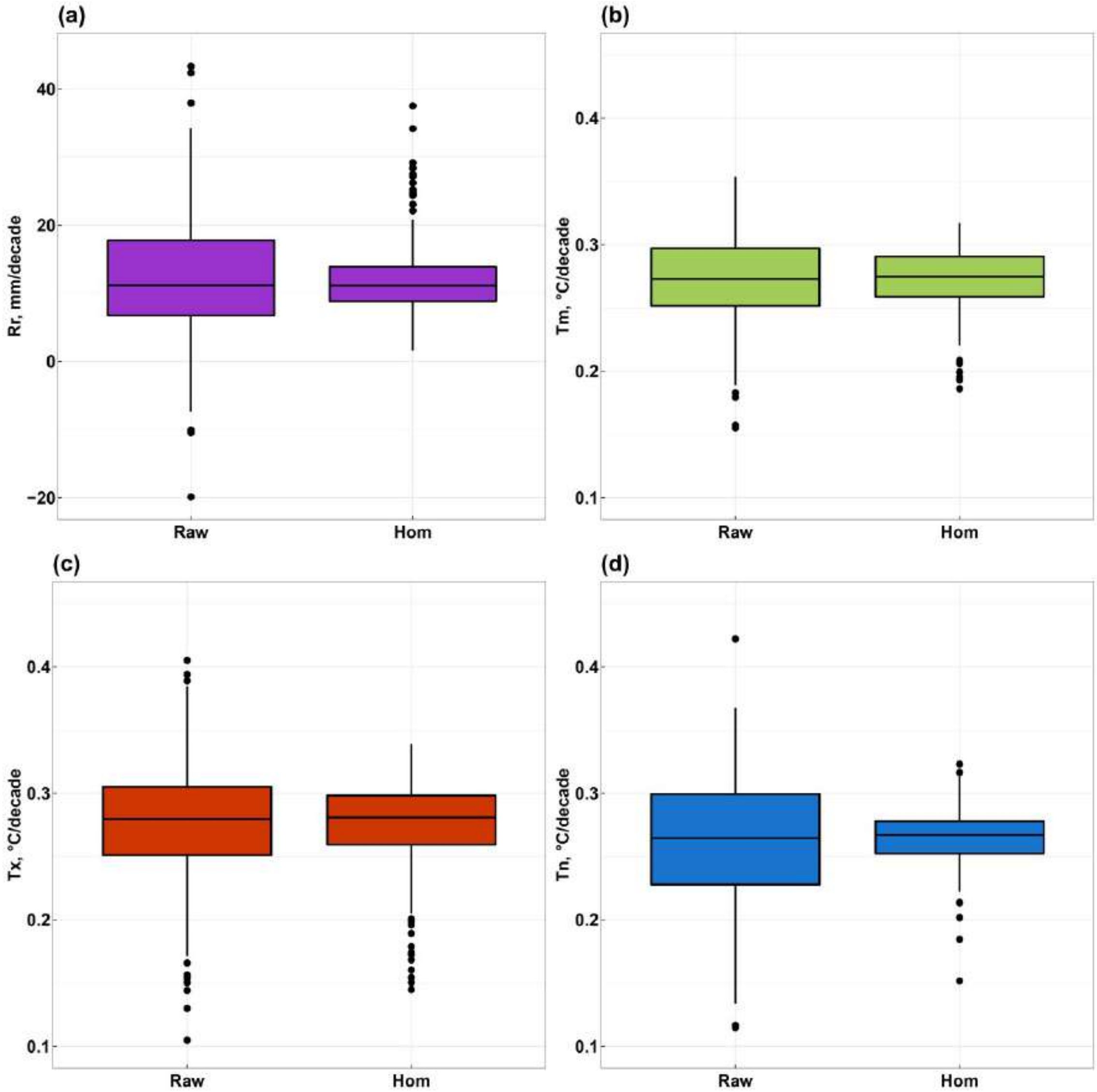

FIGURE 7. Boxplots of the linear regression coefficients (trends) in the raw (left) and homogenized (right) yearly time series: (a) Rr, (b) Tm, (c) Tx and (d) Tn.

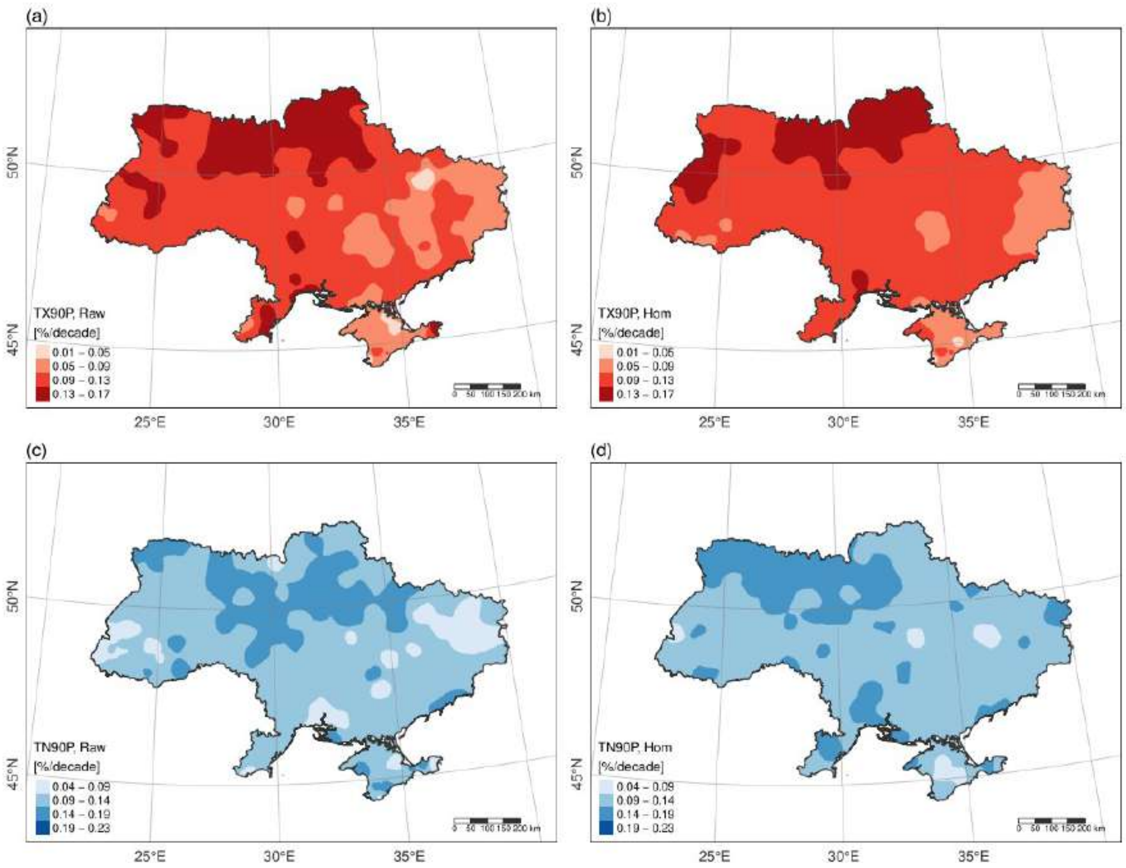

FIGURE 8. Spatial distribution of the calculated trends in TN90p and TX90p before (left) and after (right) QC/homogenization.

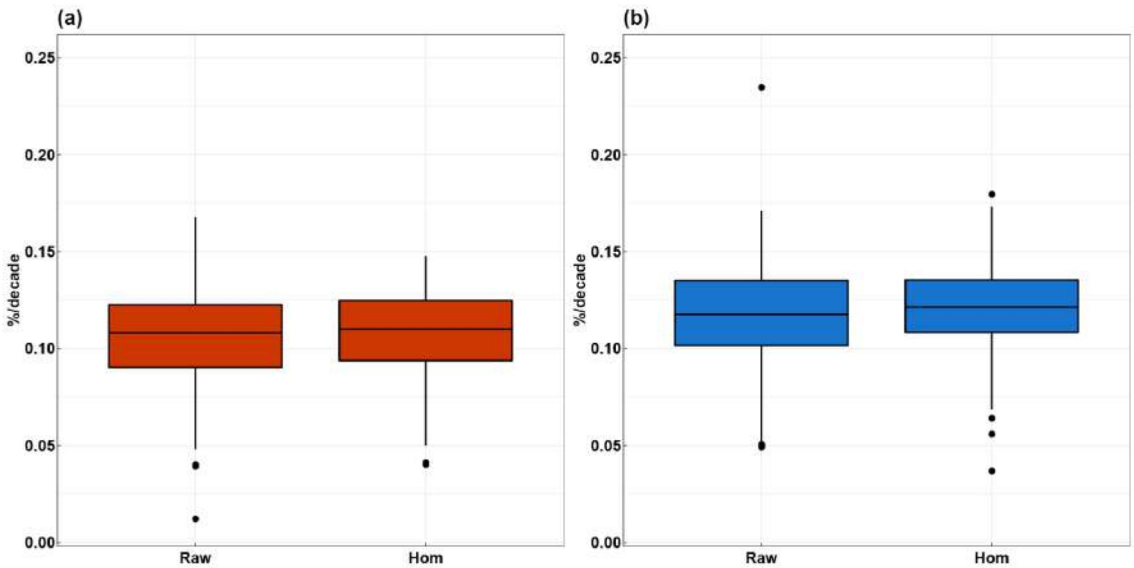

FIGURE 9. Boxplots of the calculated trends in (a) TX90p and (b) TN90p before (left) and after (right) QC/homogenization.