ACMANT homogenization software Manual, version 5.0 subversion 5

(automatic homogenization for additive biases)

2021

Content

1. Introduction

ACMANT (Applied Caussinus – Mestre Algorithm for homogenizing Networks of climatic Time series) is known about its outstanding accuracy among tested automatic homogenization methods (Domonkos et al., 2021). This subversion ACMANTv5_S5 is suitable for the homogenization of temperature, relative humidity, wind speed, sunshine duration, radiation (of any kind) and atmospheric pressure. All these variables can be homogenized either on daily or monthly scales. This subversion of ACMANTv5 is applicable only for automatic homogenization. ACMANTv5 offers metadata use together with the statistical procedure. Further principal properties of ACMANTv5_S5 are:

(i) ACMANTv5 includes an automatic networking procedure, therefore the spatial similarity in the climate of the source area of the input data is not required. An exception will be mentioned in Sect. 2.1. Input datasets may have any size between 4 and 5000 time series.

(ii) ACMANTv5 removes automatically the physically impossible values from the input, and users may define thresholds for climatic outliers. Beyond the control of climatic outliers, ACMANTv5 offers the filtering of spatial outlier values with monthly input, and the filtering of short-term spatial outlier periods with either daily or monthly input.

(iii) Interrupted homogenization of datasets can be continued without repeating the homogenization of the series whose homogenization results have once been produced.

(iv) Input time series may cover varied time periods, and strict rules must be followed in their preparation.

(v) Metadata can be used in automatic mode. The list of metadata dates must be prepared in the prescribed form. If metadata list is not provided, the program will run without metadata use.

(vi) The major part of the theoretical structure of the previous ACMANT version (Domonkos, 2020) has been kept. However, a novelty of ACMANTv5 is the combined time series comparison of time series (Domonkos, 2021) in the first cycle of the homogenization procedure. This provides improved accuracy when spatially systematic biases affect data homogeneity (Domonkos, 2021).

(vii) The software offers default solutions regarding several details of the homogenization, but users may select personalized options. The list of optionally alterable characteristics are as follows:

- Minimum of spatial correlations (default = 0.4);
- Thresholds for climatic outliers (default = thresholds of physically possible range)
- Time series subjected to homogenization (default = all);
- Gap filling;
- Items of homogenization output;

Any input dataset for ACMANTv5 should be free of strikingly large, climatically impossible outlier values, values generated by spatial interpolation or by a previous homogenization procedure in which spatial comparisons were applied. Input datasets

should preferably contain all the existing observed climatic data of the regions under study, as the efficiency and accuracy of homogenization increase with the spatial density of data.

If you have questions, comments or any unexpected experience in using ACMANTv5 or ACMANTv4.3, please do not hesitate to contact with the method developer:

Peter Domonkos
dpeterfree@gmail.com

References:

Domonkos, P**.** 2020: ACMANTv4: Scientific content and operation of the software. 71pp. https://github.com/dpeterfree/ACMANT.

Domonkos, P. 2021: Combination of using pairwise comparisons and composite reference series: a new approach in the homogenization of climatic time series with ACMANT. In: Atmosphere special issue: Application of Homogenization Methods for Climate Records (ed. Domonkos, P.). https://doi.org/10.3390/atmos12091134.

Domonkos, P., Guijarro, J.A., Venema, V., Brunet, M., Sigró, J. 2021: Efficiency of time series homogenization: method comparison with 12 monthly temperature test datasets. J. Climate, 34, 2877-2891. https://doi.org/10.1175/JCLI-D-20-0611.1.

2. Content of software package ACMANTv5_S5

2.1. Essence of scientific content

This software homogenizes the first moment of climatic variables listed in Sect. 1, provides the probable positions of sudden shifts in the means (hereafter breaks) and short-term large size inhomogeneities (outlier values and outlier periods), assesses the size of breaks and outliers, provides the homogenized time series and fills data gaps with interpolated values (optional). The homogenization is based on the concept that inhomogeneities result in linear biases. Although the linearity is not perfect, the use of linear model gives good results in the homogenization of annual and multiannual means and linear trends.

Seasonally varying biases may be treated a) with the inclusion of a sinusoid model of seasonal changes (S-model), b) with individually estimated monthly correction terms (for irregular seasonality, I-model), or c) with the exclusion of the seasonal variation of correction terms (flat, F-model). Users must choose between these three options according to the advice below:

In the middle and high latitudes not only the mean climate, but also the inhomogeneities often have quasi-harmonic shape with modes around the solstices. In these regions, the selection of S-model is recommended for the homogenization of mean temperature, maximum temperature, relative humidity, sunshine duration, radiation and sea level pressure, while the I-model is recommended for the homogenization of minimum temperature and wind speed. For tropical belt regions and regions of subtropical monsoon climate, the application of I-model is recommended for all climate variables.

When the source area of the input dataset covers regions for which varied inhomogeneity seasonality models are favourable, the division of the initial dataset is recommended, since in ACMANTv5 varied seasonality models cannot be applied within one homogenization procedure.

In the homogenization of sunshine or radiation the F-model cannot be chosen.

Daily values are adjusted relying on the estimated biases of long-term means and seasonality, while the non-linear component of daily biases are not treated.

2.2. Structure of directories and files

The software is put into a directory named “ACMANTv5.0_S5”. In its main directory one can find the main executive file “ACMANT5run.BAT” together with three subdirectories (hereafter they are referred to as directories) and the file of this guide. The BAT file of the main directory opens an executive file of directory “Works”, which manages the homogenization procedure using 20 further BAT files and 16 further executive files.

Directory “Input” is for the deposition of input dataset, while the homogenization products will appear in directory “Output”. Within directory Works one can find several files and one subdirectory “Auxiliary”. Note, however, that during the homogenization procedure the temporary subdirectory “Subnetws” is created when the procedure divides the input dataset to networks. This subdirectory is deleted at the end of a homogenization procedure.

3. Preparation of input data

Before starting a homogenization procedure with ACMANTv5, you are expected to prepare your data in the required format and put them into directory Input. It is strongly recommended to remove any other files from “Input” (e.g. the input data of an earlier homogenization) before this step. In Sects. 3.1 – 3.4, the rules for the preparation of input climatic datasets are presented. Note that ACMANT can be applied if the amount and

temporal compactness of spatially fairly comparable data reaches certain (rather low) thresholds. The thresholds are shown in Sect. 3.5. Sect. 3.6 presents the rules for preparing metadata lists. Users may change default parameterization with showing the requested parameters in request files. The rules for editing such files are shown in Sects. 3.7 – 3.9.

Note again that all kinds of input information must be put into directory Input before starting a homogenization procedure.

3.1. Rules for all kinds of input climate data files

i) Number of time series per dataset is between 4 and 5000. As a general rule, a time series or its section will be homogenized if it has at least 3 neighbour series with temporally coincidental periods of observed data and sufficient spatial correlations (default threshold = 0.4) with the candidate series. A neighbour series or its section is usable only when that can be homogenized based on the same conditions as which are prescribed for the candidate series.

ii) Length of time series (may include data gaps): between 10 years and 200 years.

iii) Time series are separated into distinct files according to observing stations. In other words: 1 file is for 1 station time series only.

iv) Each file must contain a headline with the name of the station or other station identifier. There is no restriction for the length of the headline, but only its first 40 characters will be usable for identifying the station.

v) Default units of climatic elements: °C for temperature, % for relative humidity, hour for sunshine duration, m/s for wind speed and hPa for atmospheric pressure. You are allowed to use other units with the following precautions: a) It is indispensable to use the same unit throughout a given homogenization process. b) The wrong use of units might cause overshooting of physical thresholds inbuilt in programs or overshooting of user defined climatic thresholds even when the input data is correct. c) Climatic values below -998 are considered missing data and those higher than 9999 cannot be treated by the software.

vi) Each line (after the headline) contains a pre-determined number of values (i.e. date identifier(s) and 12 monthly values in monthly datasets or 28-31 daily values in daily datasets). The values in the same line are separated with one or more space characters. TAB also can be applied for separation (optional). The use of other characters (comma, semicolon, etc.) is not allowed. Date identifiers are shown with whole numbers, while the data of climatic variables are usually presented with 1 decimal preciseness, but note that other forms are also accepted. Decimals are separated by dot from the other digits. E.g. value – 4°C can be presented in any form of – 4, – 4.0 or – 4.00.

vii) Missing values are represented by –999.9.

viii) Gap filling with missing value code

Input time series must be temporally continuous from the first calendar day (or month, in case of monthly input) of the year of the first observed data until the last calendar day (month) of the year of the last observed data, although some exceptions are allowed (Sect. 3.4).

Best Site Station

1951 -999.9 -999.9 4.5 7.3 15.6 17.6 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9
1952 -1.5 0.0 7.1 9.8 10.9 15.6 19.8 20.3 13.8 -999.9 8.1 0.4
1953 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9
1954 -999.9 -999.9 -999.9 8.6 14.2 16.8 18.8 18.5 16.4 9.5 3.6 -0.2
…….
…….
…….

In Best Site Station the temperature observation started in March 1951, but there are several gaps in the first few years. Please remember that it is not allowed to jump over the year 1953, in spite of that there was no observation at all in that year. It is also important to note that the time series starts with the January of the first observed data (missing data codes for January and February of 1951).

3.2. Specifics for monthly input datasets

i) Time-resolution is monthly.

ii) Name of files: “ANAMEJJJJ.txt”.

- “ANAME” is the name of the dataset. This substring is always 5-character wide and may contain any alphanumeric characters.
- “JJJJ” is the serial number of time series, it always consists of 4 digits.
- “.txt”: these 4 characters are fixed, they are always “.txt”

Example: If the name of the dataset is “Lucky” and it consists of 275 time series, then the names of the time series must be:

Lucky0001.txt
Lucky0002.txt
Lucky0003.txt
……..
……..
Lucky0274.txt
Lucky0275.txt

iii) Each line (after the headline) is for one calendar year. Thus the number of lines in file is the same as the number of years in the time series, plus the headline.

iv) Each line consists of 13 values: the first is a calendar year, while the other twelve are the monthly climatic values for that year in their natural order (from January to December).

3.3. Specifics for daily input datasets

i) Time-resolution is daily.

ii) Name of files: “ANAMEJJJJd.txt”.

- “ANAME” is the name of the dataset. This substring is always 5-character wide and may contain any alphanumeric characters.
- “JJJJ” is the serial number of time series.
- “d.txt”: these 5 characters are fixed, they are always “d.txt”

Example: If the name of the dataset is “Smart” and it consists of 7 time series, then the names of the time series must be:

Smart0001d.txt
Smart0002d.txt
……..
Smart0007d.txt

iii) Each line (after the headline) is for one calendar month. The first line (after the headline) is for the January of the first year, while the last line is for the December of the last year. Thus the number of lines in file is the same as the number of years in the time series multiplied by 12, plus the headline.

iv) Each line consists of 2 date identifiers and 28-31 values. The date identifiers are the calendar year and the serial number of calendar month. The other values are the observed climatic data in their temporal order (from the first day to the last day of the month). If a line contains less values than the number of days in the month, the program will stop with an error message. By contrast, if lines contain more values than the number of days in the month, the program will not stop, but the data at the end of lines will not be read by the program.

v) Example of daily wind speed series with observed data between 5 May 1954 and 29 November 2006. Observation was suspended between 30-Dec-1954 and 03-Jan-1955.

Nice Park City
1954 1 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9……..-999.9 -999.9 -999.9 -999.9
1954 2 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9……..-999.9
1954 3 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9……..-999.9 -999.9 -999.9 -999.9
1954 4 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9……..-999.9 -999.9 -999.9

1954	5	-999.9	-999.9	-999.9	-999.9	1.4	0.8	……	10.2	4.9	4.3	2.1
1954	6	6.0	4.0	7.3	4.2	2.0	3.4	….....	2.0	0.8	11.3	
1954	7	6.2	3.0	3.2	1.8	8.3	3.1	……	1.7	1.5	5.9	4.3
1954	8	1.4	3.6	5.1	9.0	2.6	4.4	……	7.2	2.2	1.4	4.5
1954	9	3.0	5.3	13.6	8.7	2.8	2.6	……	9.1	3.8	4.8	
1954	10	4.2	4.4	4.0	2.7	6.6	5.7	……	2.2	4.4	2.7	5.2
1954	11	5.2	3.0	2.9	2.6	6.4	4.6	..……	5.0	6.1	3.8	
1954	12	7.4	2.1	1.1	4.0	4.3	3.5	..……	1.0	3.6	-999.9	-999.9
1955	1	-999.9	-999.9	-999.9	2.8	3.1	5.7	……	7.2	7.1	7.4	3.4
1955	2	3.0	3.6	5.8	8.8	8.2	4.1	……	4.3			
1955	3	6.2	6.7	4.6	7.1	6.4	6.0	……	8.4	10.0	3.7	3.9
……..												
……..												
……..												
2006	9	4.0	2.7	2.6	7.0	5.2	4.5	…...	4.7	4.0	3.8	
2006	10	3.3	2.0	1.9	2.1	7.4	6.3	…...	3.6	3.2	14.1	7.3
2006	11	5.0	5.1	4.1	1.8	1.6	4.7	…...	6.1	4.7	-999.9	
2006	12	-999.9	-999.9	-999.9	-999.9	-999.9	-999.9	…….-999.9		-999.9	-999.9	-999.9

3.4. Exceptions when input time series are allowed to be fragmented

Before starting the homogenization, the user must define the target period of the homogenization (Sect. 4.2). The input data out of this target period are allowed to be fragmented if they comply with the following conditions.

i) Before the beginning of the target period: If data for at least one date is included from a given year, then all the dates of that year (i.e. all the months or all the days according to the resolution of the dataset) must be represented either with observed data or with missing data code. However, years without any observed data can be jumped out.

ii) After the target period: If the input time series includes the last year of the target period, then the reading of that file will be terminated with the data of that year, and thus anything might be present in the later part of the file. However, if the series of continuous data terminates before the last year of the target period, the file must be ended together with the end of the series of continuous data.

Note that if you use any of these alleviations, this might cause problems if you would like to repeat the homogenization of the same time series with varied target periods.

3.5. Conditions for treatment and homogenization by ACMANTv5

i) Requirements for the treated period of a time series

Sometimes time series include large missing data fields at the beginning and/or at the end of their period. During the homogenization procedure, the input time series are split into three sections: a treated section (referred to as treated period), and two not treated sections, one in the beginning part and another one in the ending part of the time series. Not treated sections remain out of most steps of the homogenization procedure, yet inhomogeneity adjustments and data completion by spatial interpolation may be applied in them. When the number of missing values is low, the whole period of the time series is treated period. The ratio of missing values within the treated period is limited according to the following rules:

a) In the first *k* year and last *k* year of the treated period the expected ratio of available observed data is higher than 25% for any *k* between 1 and 20, or between 1 and the length of the treated period if the latter is the shorter. For instance, if the first year is 1953 and the last year is 2002, the minimum number of observed monthly data for periods 1953-1953, 1953-1954, 1953-1967 and 2000-2002 are 4, 7, 46 and 10, respectively. Also, at least two neighbour series with sufficient ratio of observed data are needed.

b) The minimum length of a treated period is 10 years.

c) The minimum number of observed monthly data is 9 for each month of the year.

d) The minimum number of the total of observed monthly data is 114.

e) Even when the input data is daily, the adequacy of the amount and temporal coherence of data is checked on monthly scale, according to the rules listed above. For this check, the status of monthly climatic value is “observed” when there are no more than 7 missing daily values in month and the status is “missing” in the reverse case.

In case of too few or too fragmented data, the above conditions might remain unsatisfied throughout the time series. In this case, the time series will fully be excluded from the homogenization procedure: for these series neither inhomogeneity adjustments nor data completion are applied. Note that the homogenization procedure does not stop for the exclusion of one or more time series.

ii) Requirements for homogenizing series or their sections

a) If a station series (or its section) does not have at least 3 neighbour series with sufficient spatial correlation, that series (or its section) will not be homogenized, but otherwise the program will run normally. If none of the series can be homogenized, the program will stop with the adequate control message. Note that when no section of a candidate series can be homogenized, neither inhomogeneity adjustments nor data completion are applied to that series.

b) When there is a section of candidate series without 3 neighbour series, that section will not be homogenized. Sometimes both before and after that section the number of comparable time series is higher, thus two distinct sections could be homogenized. However, in such cases only one section, i.e. the late section will be homogenized, even if a longer section could have been homogenized in the early part of the candidate series. If more than one separate homogenizable sections are found by ACMANT, then this information will appear in the output file "Control message.txt".

The section of time series which can be homogenized according to the presented conditions is referred to as the homogenized period of the time series. Often the whole period of the input time series is homogenized period. Not treated sections cannot be part of the homogenized period. The use of data falling within the treated period but out of the homogenized period is strongly limited in the homogenization procedure.

3.6. Metadata list

Users may provide a list of metadata dates before starting a homogenization procedure. When such a list is provided, the homogenization procedure will use the metadata dates in combination with the break dates detected by statistical methods.

In the evaluation of pairwise detection results, the value of a piece of metadata equals to 1 statistically detected break of high certainty (score = 1.0). The minimum threshold score for validating a break position at that phase of the homogenization procedure is 2.1 (see the explication of pairwise detection result scores in Domonkos, 2021). Later in the homogenization procedure, metadata dates are used also to refine the dates of statistically detected breaks.

All the metadata of a given input dataset must be shown in one common file. Its format is the same for daily and monthly climatic datasets, since metadata are always shown with daily preciseness. Even when you do not know the exact date, the estimated date must be shown with daily preciseness.

The name of metadata list file: "ANAMEmeta.txt", where

- "ANAME" is the dataset name (the same as for the climatic data)
- "meta.txt" is constant

The size of this file depends on the number of known metadata dates, and there is no need to order the pieces of metadata either according to time or station serial numbers.

In each line of the file, 4 whole numbers must be shown, they are from left to right: a) serial number of station, b) day, c) month, d) year. For instance, LUCKYmeta.txt has 20 time series, but only 4 metadata:

14 31 11 1970
14 12 4 1970
5 31 12 1988
9 7 11 1963

The example shows that more than one metadata may occur for the same year and same station, but note that too high metadata frequency is generally neither recommended nor reasoned. The maximum number of pieces of metadata for a station is 40, while for an entire input dataset 40,000. Warning: Any erroneous date will stop the program with the relevant error message. However, when metadata file is not provided, it does not cause problem, and the program will run without metadata use.

3.7. Thresholds for climatic outliers

ACMANTv5 can be run with the inbuilt physical thresholds, or user defined climatic thresholds can be applied.

i) Inbuilt thresholds
temperature: -98 and 60
relative humidity and wind speed: 0 and 100
sunshine and radiation, daily: 0 and 24
sunshine and radiation, monthly: 0 and 744
atmospheric pressure: 0 and 1099

ii) User defined thresholds
User defined thresholds overwrite the inbuilt threshold values. While the inbuilt thresholds are season-independent, user defined thresholds consist of monthly low threshold and high threshold values for each calendar month. These thresholds are written to the request file “Thresholds.txt”. The file must include 12 lines, with the serial number of calendar month and the monthly low threshold and high threshold values in each line. An example of Thresholds.txt for the homogenization of temperature maximums in a Mediterranean region:

```
1  -10   35
2  -10   35
3  -10   40
4   0    50
5   5    50
6   10   55
7   10   55
8   10   55
9   10   50
10   5   45
11  -5   40
12  -10  35
```

Note that lower (higher) threshold values than -997.9 (9999) cannot be defined.

When the use of inbuilt physical thresholds of the software is preferred, the input package should not contain Thresholds.txt.

3.8. Spatial correlation

Two kinds of spatial correlations are used in ACMANT, one is for homogenization, and another is for gap filling. The ways of their calculation differ (Domonkos, 2020). For gap filling, the minimum threshold is 0.4 and users cannot modify it. For homogenization, the spatial correlations are calculated from de deseasonalised values of monthly increment series (see more in Domonkos, 2020). The default minimum threshold for homogenization is 0.4, but users can alter this by the edition of the relevant request file.

The filename is constant: “Rth.txt”. The requested spatial correlation threshold must be written into its first line without adding any other thing. The program accept any value between 0.1 and 0.99. When Rth.txt is not shown in directory Input, or a valid correlation threshold cannot be found in that, the program will use the default correlation threshold.

3.9. Target series for homogenization

For relatively small datasets, ACMANT usually homogenizes all the series together without grouping them into networks.

When networks are constructed by the program, still all the series of a network are homogenized together, but in networks the results are saved only for the candidate series of the network. It is because a principle of the network construction in ACMANT is that a network is optimal for one selected candidate series, and therefore the number of networks equals to the number of the time series of the input dataset (see also Domonkos, 2020).

In homogenizing datasets of larger than 40 time series (multi-network datasets), users may save time with defining which time series are expected to be homogenized when not all the series are needed to be homogenized. The serial numbers of the requested time series must be written into the request file “ANAMEtarget-homg.txt”. In this filename

- “ANAME” is the dataset name (the same as for the climatic data)
- “target-homg.txt” is constant

Each line of the file must contain one serial number. You may write the relevant serial numbers in any order. If the same serial number is shown more than once, this does not cause error. However, error will occur if a) an invalid serial number is shown, then the program stops with error message; b) the file includes more lines with the presentation of valid serial numbers than the number of time series (N) in the dataset (it is possible by repeated presentations of some serial numbers), then the program considers only the first N serial numbers.

Example: dataset WHITE includes 1622 time series, but you need the homogenization results only for series 7, 78, 922, 926, 1473 and 1496. Then WHITEtarget-homg.txt file must be created with the following content:

7
78
922
926
1473
1496

If WHITEtarget-homg.txt is not shown in directory Input, ACMANT will homogenize all the 1622 networks. If you write 1622 into the first line of WHITEtarget-homg.txt and nothing else, then ACMANT will homogenize only the last network.

4. Running ACMANTv5_S5

4.1. General conditions

The software package with its directory structure must be placed in your computer. Saving a copy is advised. The preferred environment is Windows, since the compiled FORTRAN programs of ACMANT are directly executable under Windows. Under Linux, de program can be run by “Wine” application, starting the run with the command “wineconsole ACMANT5run.BAT”.

The computational time demand widely varies: for datasets of up to 40 time series 1-network homogenization is performed, which may last from a few seconds to a few minutes. In multi-network homogenization the typical time demand is from a few hours to a few days, but for very large datasets (>1000 series) the time demand may be several weeks. In multi-network homogenization the time demand increases linearly with the number of time series and exponentially with the length of the target period. The homogenization of daily data needs notably more time than the homogenization of monthly data.

Although this software is automatic, you must introduce some parameters manually at the beginning of the homogenization procedure.

4.2. Initiation I. Starting a new homogenization

Once the input dataset and supplementary input files have been prepared, they are placed to directory “Input”. Please check that directories “Input” and “Output” do not include data of earlier homogenizations.

The homogenization starts with clicking on “ACMANT5run.BAT”. The program will ask for typing some parameters on the screen. When the answer includes (or may include) letters, please introduce the answer from the left end of the line, without space characters.

i) Climatic element: Two characters. It is “TT” for temperature, “HH” for relative humidity, “FF” for wind speed, “SS” for sunshine duration and radiation, and “PP” for atmospheric pressure.

ii) Temporal resolution: One character, “m” for monthly homogenization and “d” for daily homogenization.

iii) Name of the dataset: 5 characters, the same as which is included in the file names of the input time series.

iv) Number of stations: The number of time series in the input dataset must be given here.

v) – vi) First calendar year of time series and number of years in time series: These two parameters determine the target period of homogenization. If all the input time series cover the same time period, it is straightforward to define that period as target period, but the period of observations might vary widely in large datasets. You are allowed to choose either shorter or longer target period than which is covered with input data, there are two requirements only: a) the length of target period must fall between 10 and 200 years; b) All dates of the metadata list (if metadata list is provided) must fall within the target period.

vii) Seasonality of inhomogeneities: It can be sinusoid (“S”), irregular (“I”) or flat (F). See explanation and advice in Sect. 2.1. Option “F” is not offered in sunshine duration / radiation (SS) homogenization.

viii) Would you like outlier filtering?: This question appears only in monthly homogenization, and the answer is 1 character, “Y” if yes and “N” if no. The recommended response here is yes.

ix) Would you prefer default output package?: The answer can be “Y” or “N”, and if “Y” is chosen, the program starts immediately the homogenization procedure. The default output package comprises the

- homogenized time series in the same time resolution as that of the input time series;
- list of neighbour series and their spatial correlations with the candidate series;
- list of detected breaks and outliers.

A more detailed description of the default output package is presented in Sect. 5.1 and 5.2.

If you prefer other output items (see Sect. 5.3) than those of the default package, you may not respond with Yes to this question. When the default output package has not been accepted, ACMANTv5 puts further questions:

x) Time resolution of homogenized time series. This question appears only when the input dataset is of daily resolution. “d”= daily, “m”= monthly, “2”= both kinds of output items are required.

xi) Gap filling. “0”= never, “1”= within the homogenized period, “2”= completion of series for the target period, “3”= all kinds of output items of the previous three options.

xii) Table of confidence indicators: “Y” if yes, “N” if no. --- Indicator values show for individual monthly or daily values if they are homogenized observed data, interpolated data, etc. The indicators are whole numbers between 0 and 9. Lower values (except for 0) mean higher reliability, see detailed description in Sect. 5.4.

xiii) List of detected breaks and outliers. “Y” if yes, “N” if no. This item is a part of the default output package, but if you opt here for “N”, the item will not be output.

xiv) List of neighbour series and their spatial correlations with the candidate series. “Y” if yes, “N” if no. This item is a part of the default output package, but if you opt here for “N”, the item will not be output.

4.3. Initiation II. Continuation of a suspended homogenization

You may suspend the homogenization of multi-network datasets. When you want to continue the homogenization, click again on ACMANT5run.BAT, then the program will ask for the climatic element. To that question, please respond with “CC” instead of the code of the climatic element. From this response ACMANTv5 understands that you want to continue a suspended homogenization, and will ask you how many time series have already been homogenized. You can check that from directory Output and introduce the correct response. Then ACMANTv5 will run without putting further initiation questions.

5. Results of the ACMANTv5_S5 homogenization procedure

The output may include various versions of homogenized series, a summary of the detected breaks and outliers, files with monthly and daily reliability indicators related to the data treatment during the homogenization procedure, the list of neighbour series and their spatial correlations with the candidate series and one or more control messages. All these depend on the user’s choices provided at the beginning of the homogenization procedure (Sect. 4.2). One easy option is to choose the default output package, its content is detailed in Sect. 5.1 and 5.2. However, users may choose other output items than which belong to the default output package. The optional output items are described in Sect. 5.3.

5.1. Default output package of monthly homogenization

i) Homogenized monthly series with gap filling within the homogenized period (for possibly existing missing data of the raw series). Their names have the form of "ANAMEJJJJt.txt", where

"ANAME" – name of the dataset
"JJJJ" – serial number of the time series
"t.txt' – constant

This file contains the homogenized time series, but note that the requested homogenized data tables are always generated, even when no section of a time series has been homogenized. So that the number of files of this kind is the same as the number of time series in the input dataset. The data format is exactly the same as the input data format.

Note that monthly data tables can be requested also when the input dataset is of daily resolution. In this case, output monthly data tables present monthly mean values for the majority of climate variables, but monthly totals for sunshine duration and radiation (SS).

ii) Summary of the homogenization procedure with the list of the detected breaks and outliers and some other characteristics. The name of this file has the form of "ANAME_breaks.txt".

"ANAME" – name of the dataset
"_breaks.txt" – constant

The organisation of the data in this file is as follows: The list comprises *N* sub-lists, where *N* stands for the number of stations. A sub-list generally contains the following 3 parts: a) headline, b) list of breaks, c) list of outliers. The headline contains (from the left to the right) the serial number of the time series, the starting and ending years of the treated period (Sect. 3.5), the starting and ending years of the homogenized period, the number of detected breaks, the number of detected outliers (only in outlier filtering "yes" mode), and the station identifier. In part (b), the properties of the detected breaks are presented. They are the serial number of break, the year and month of break position and the break size for the annual mean of the tested variable. In bivariate homogenization two break sizes are shown, one for each variable. In part (c), the properties of the detected outlier periods are presented, they are the serial number of the outlier period, the year and month of the last month of the outlier period, the duration of outlier period (in months) and the magnitude of deviation. Some notes:
- If no break is detected in the time series, there is no b) part in the sub-list.
- If no outlier is detected in the time series, or outlier filtering "No" mode is applied, there is no c) part in the sub-list.
- The most frequent length of outlier periods is 1 month (i.e. one single outlier value), but lengths may vary between 1 and 4 months. Note that outlier periods of longer than 4 months are treated as a pair of breaks, and their properties are shown in the break list (part b).

- If homogenization has not been performed, character “0” is written into the places of the starting and ending years of the homogenized period, and “–1” into the place of the number of breaks.
- If a time series does not have treated period, “0” are shown in the places of the starting and ending years of the treated period.

iii) List of neighbour series for each candidate series, and the spatial correlations between candidate series and their neighbour series. The name of the file has the form of “ANAME_rlist.txt”.

“ANAME” – name of the dataset
“_rlist.txt” – constant

In multi-network homogenization this file shows the network structures and the correlations between the candidate series and neighbour series. In 1-network homogenization (for small datasets), the candidate series – neighbour series relations are also shown for every time series, since the roles of being candidate series or neighbour series change during the homogenization procedure. Note that when a candidate series does not have homogenized section, no neighbour series or spatial correlation is shown for that. Furthermore, when a neighbour series cannot be homogenized, that series is not shown in any neighbour series list, even if it was selected for some candidate series during the automatic network construction. The form of data is as follows: for each candidate series with non-zero homogenized period has a headline, followed by the list of neighbour series ordered according to decreasing spatial correlations. The headline includes (from the left to the right) the serial number of the candidate series, the starting and ending years of its homogenized period, and its station identifier. Then, in each line after the headline, the serial number of correlation value, the correlation value itself and the station identifier of the neighbour series are shown.

iv) Control message. In running ACMANT, “Control message.txt” file is generated. If no problem has been detected during the run of the program, “….homogenization has been completed” will be the message. However, other messages are possible. In multi-network homogenization, at least one message is provided for each network homogenization.

5.2. Default output package of daily homogenization

v) Homogenized daily series with gap filling within the homogenized period. Their names have the form of “ANAMEJJJJv.txt” where

“ANAME” – name of the dataset
“JJJJ” – serial number of the time series
“v.txt” – constant

Its data format is exactly the same as the input daily data format.

The items ii) - iv) of the standard monthly homogenization output (Sect. 5.1) are included also in the standard daily homogenization outputs, however the content of "ANAME_breaks.txt" (item ii) slightly differs:
- In daily homogenization, break positions are shown with daily preciseness. Note that for low signal-to-noise ratio on the daily scale (which is frequent), the calendar day for the break position is the default value, i.e. the last day of the month.
- For outlier periods, the starting and ending dates are presented with daily preciseness, instead of their length in months.
- Sometimes missing data code (-999.9) stands in the place of the deviation (systematic bias) size of the outlier period. It is either because the detection results do not prove a platform shaped structure (which is necessary to estimate the mean systematic bias), or for the lack of sufficient observed data in neighbour series to provide reliable estimations.

5.3. Optional output items

Optional output items are data tables of homogenized time series or tables of reliability indicators. All of them have the same format as "ANAMEJJJJt.txt" (for monthly data) or "ANAMEJJJJv.txt" (for daily data). The letter before ".txt" shows the kind of the output item, the other characters of the filenames are the same for a given time series.

vi) ANAMEJJJJs.txt – homogenized monthly time series without gap filling for missing data.

vii) ANAMEJJJJh.txt – homogenized monthly time series completed with interpolated data to the target period when data gaps occur in the input data.

viii) ANAMEJJJJu.txt – homogenized daily time series without gap filling for missing data.

ix) ANAMEJJJJg.txt – homogenized daily time series completed with interpolated data to the target period when data gaps occur in the input data.

x) ANAMEJJJJj.txt – Monthly data table of reliability indicators. The data format is the same as for climate data tables, except that whole numbers (0…9) of reliability indications stand on the places of climate data of climate data tables. The meanings of reliability indicators are presented in Sect. 5.4.

xi) ANAMEJJJJi.txt – Daily data table of reliability indicators. The data format is the same as for climate data tables, except that numbers of reliability indications stand on the places of climate data of climate data tables.

5.4. Meaning of reliability indicators

Although the file ANAME_breaks.txt shows the homogenized period for each time series, reliability indicators might show more about the reliability of the homogenized data for two reasons:

- When less than three time series have data for a specific year, that year is excluded from the homogenization for all the stations. Such years might occur within the homogenized period.

- ACMANT automatically performs gap filling with spatial interpolation, but when the number of neighbour series or the spatial correlations are low, the confidence of interpolated values is limited.

Generally the lower values mean higher reliability, except for the 0 code.

0 – sporadic observed value, out of the treated period of the time series, or observed value in a section of the candidate series where the number of neighbour series with observed values is 0 or 1.
1 – observed value within the homogenized period
2 – observed value out of the homogenized period
3 – Interpolated value from at least 4 sufficiently correlating neighbour series.
4 – Interpolated value from 3 highly correlating neighbour series, or from more than 3 but only moderately correlating neighbour series.
5 – Interpolated value from at least 3 fairly correlating neighbour series.
6 – Interpolated value from 2 fairly correlating neighbour series, or from more but poorly correlating neighbour series.
7 – Interpolated value by the use of 1 only neighbour series.
8 – Missing value is substituted with the climatic normal value due to the complete lack of observed values in neighbour series. Frequent occurrence of this code number may be present when the target period is longer than the periods of observed data, and full completion of time series is performed. When no data is available for spatial interpolation, the climatic normal values are included (repeatedly) in the artificially lengthened series. Note, however, that this code sometimes occur within the homogenized period due to synchronous data gaps (e.g. for political events).
9 – missing data for which interpolation has not been performed. This code may occur out of the homogenized period, when data tables of time series completed to the target period are not requested.

Note: The coding between 3 and 7 is based on the number of neighbour series, the spatial correlations between series, the frequency of comparable data pairs around the date of the interpolation, and if the data are within the homogenized period yes or not. For this complexity, not all details of the coding rules are presented.

6. Possible technical problems with using ACMANTv5_S5

The program might fail to accomplish the homogenization for various reasons. The use of a managing program (i.e. the program ACMANT5run manages the running of other "inner" programs of the software) has an inconvenience: when an inner program stops during the homogenization of a network, the running of the managing program still continues, and the procedure turns on the homogenization of the next network. As a consequence, users might fail to realise in time that a problem occurred during the homogenization of a dataset.

The following symptoms indicate that the program failed during the homogenization of a candidate series: (i) In the list of control messages ("Control message.txt") a candidate series is missing; (ii) Some of the output items or all the output items of a given candidate series are missing; (iii) The directory "Output" includes error message from the program (the homogenization may stop with or without giving error message). Note that the homogenization of some candidate series might be denied for the insufficient spatial-temporal coherence of the data, but this is not a software error, and in this case control message and some other output items are still generated.

When a failure of homogenization is realised, first the input data should be checked, as errors in the input data preparation may cause stop of the homogenization. In such cases, likely an error message from the program will be present in directory Works. A relatively frequent data preparation error is that the same time series occurs repeatedly, under different station names. In this case, the homogenization will stop with the error message of "Indefinite equation system".

The programs of ACMANTv5 have been tested, but in spite of this, maybe that some software errors will appear only during its practical use. Thus if you do not find irregularity in your input dataset, maybe that a software error causes the failure of homogenization, and then please contact me and report the problem.