ACMANT homogenization software: Manual

Version 3.1

Content

1. Introduction

ACMANT is a fully automatic homogenization method for homogenizing temperature and precipitation time series. ACMANT was born during the European COST ES0601 project "HOME" (www.homogenisation.org) on the basis of one the best performing non-automatic homogenization methods "PRODIGE" at that time (ACMANT = Applied Caussinus – Mestre Algorithm for homogenizing Networks of Temperature series). The first ACMANT version homogenized monthly temperatures only and was the most successful fully automatic method in the internationally controlled efficiency tests of HOME (Venema et al., 2012). From the release of the second version ACMANT2, the precipitation homogenization is also included and homogenization can be performed either with daily or monthly data (Domonkos, 2014, 2015).

The present third generation of ACMANT includes several technical changes which elevates the efficiency relative to the earlier versions.
(i) An ensemble pre-homogenization is included in a way that one potential reference component is excluded from the reference series creation in each ensemble member then the minimum adjustment terms of all ensembles are applied as preliminary corrections.
(ii) Ordinary kriging is applied for weighting reference components when at least 6 reference components are available.
(iii) Missing data treatment and gap-filling have been modified.
(iv) If the annual cycle of biases is not expected to be sinusoid in temperature homogenization (i.e. for the series of daily minimum temperatures and for temperatures of tropical or monsoonal regions), the shape of annual cycle of biases is estimated from the data (in the earlier version no seasonal change of adjustment terms was included).
(v) In daily homogenization several steps are done on daily scale, the interpolation for filling data gaps, precision of the timings of the breaks for large magnitude breaks, and the ANOVA equation system for the calculation of adjustment terms is also applied on daily scale.

ACMANT3 supports the joint application of the common quality control with ExtraQC (http://www.c3.urv.cat/data.html), homogenization with ACMANT3 and extreme index calculation with RClimDex (http://etccdi.pacificclimate.org/RClimDex/RClimDexUserManual).

The developer hopes that ACMANT3 will contribute to using more accurate temperature and precipitation data in climate variability and climate change studies. In case of any problem or question around the software, please contact with me by e-mail:

Peter Domonkos
dpeterfree@gmail.com

Reference:

Domonkos, P. 2014: The ACMANT2 software package. Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques In Climatology and Meteorology. WMO, WCDMP-84,46-72.

Domonkos, P. 2015: Homogenization of precipitation time series with ACMANT. Theor. Appl. Climatol., 122, 303-314. doi: 10.1007/s00704-014-1298-5.

Venema, V., 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., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Duran, M.P., Likso, T., *Esteban, P. and* Brandsma, T. 2012: Benchmarking monthly homogenization algorithms. Climate of the Past, 8, 89-115, doi:10.5194/cp-8-89-2012.

2. Content of software package ACMANT3

The objective of this software is to provide an effective, high quality tool for the automatic statistical homogenization of temperature and precipitation data. Input and output data are monthly or daily according to the choice of the user.

ACMANT is a relative homogenization procedure, at least 4, spatially correlated climate series are necessary for its application. The software contains auxiliary programs for the calculation of spatial correlations, since their knowledge helps to find realistic networks.

The package consists of 12 files:
i) "ACMANT3monthsinus.exe": the executive file of the ACMANT homogenization procedure for homogenizing monthly means of daily mean or daily maximum temperatures of mid- and high latitudes
ii) "ACMANT3daysinus.exe": the executive file of the ACMANT homogenization procedure for homogenizing daily mean or daily maximum temperatures of mid- and high latitudes
iii) "ACMANT3monthirreg.exe: the executive file of the ACMANT homogenization procedure for homogenizing monthly means of daily temperature minimums of any geographical region and for any kind of monthly temperature data of the tropical belt or monsoon regions.

iv) "ACMANT3dayirreg.exe: the executive file of the ACMANT homogenization procedure for homogenizing daily minimum temperatures of any geographical region or any kind of daily temperature data of the tropical belt or monsoon regions.

v) "ACMANP3month.exe: the executive file of the ACMANT homogenization procedure for homogenizing monthly precipitation totals;

vi) "ACMANP3day.exe: the executive file of the ACMANT homogenization procedure for homogenizing daily precipitation amounts;

vii) "Rtcheck.exe" executive file for calculating spatial correlations of monthly temperatures;

viii) "Rrcheck.exe" executive file for calculating spatial correlations of monthly precipitation totals, it works with transformed precipitation values;

ix) "Transform_dailydata.exe" executive file for the transformation of data format from ExtraQC output to ACMANT input.

x) "attab.dat": a supplementary data-file for the homogenization procedure;

xi) "paramf.txt" a text file including parameters for the transformation of daily data format.

xii) "ACMANT3.0_Manual.pdf": this guide.

The executive file for homogenization and file "attab.dat" must be stored in the same directory. If you use the data transformation program, then its executive file and the "paramf.txt" must be stored in the same directory together with your data.

2.1. Brief characterisation of the function of specific executive files

i) "ACMANT3monthsinus.exe" is for the homogenization of monthly means of daily mean or daily maximum temperatures. The model of the temporal variation of the biases caused by inhomogeneities (hereafter: station effect) is a step-function plus a sinusoid annual cycle. This model is based on the hypothesis that the temporal variation generally can be modelled well with step-functions, but there are substantial seasonal changes due to the seasonally varying radiation intensity in several kinds of station effects. According to simulation experiments this model functions well even when only the 30-50 % of station effects possesses the seasonal cycle. The use of this model is recommended for the higher than 25-30° geographical latitudes, except monsoon regions.

ii) "ACMANT3daysinus.exe": Its model is the same as of ACMANT3monthsinus.exe, but it works with daily data.

iii) "ACMANT3monthirreg.exe" is for the homogenization of monthly means of daily minimum temperatures, as well as for the homogenization of any kind of monthly temperature data from the tropical belt or from monsoon regions. The model of the temporal variation of station effects is a step-function with irregular annual cycle. The model presumes that the seasonal cycle of irradiation is weak, thus the radiation effect on the seasonal cycle of station effect is not sinusoid. The shape of the seasonal cycle of the station effect is estimated by the homogenization procedure.

iv) "ACMANT3dayirreg.exe": Its model is the same as of ACMANT3monthirreg.exe, but it works with daily data.

v) "ACMANP3month.exe": The model of the temporal variation of station effects is a step-function with two seasons: a) rainy season, and b) snowy season. The reasoning of the two-season model is that station effects and their temporal variation often substantially differ for rain and snow. The timing of snowy season must be defined by the user for networks. If snow is rare throughout the year, there is 1 season only, and station effects are assumed to be independent from the part of the year.

vi) "ACMANP3day.exe": Its model is the same as of ACMANP3month.exe, but it works with daily data.

vii) "Rtcheck.exe": With this program the spatial correlations between monthly temperatures can be checked. The program calculates the correlations between the increment series of monthly values, for instance change between the temperatures of April and May of a given year is paired with the same monthly change in the temperature data of other stations. This program is for the selection of climatologically coherent networks. Its use is optional, networks can be created also without the use of this program.

viii) "Rrcheck.exe": Its purpose and operation is similar to "Rtcheck.exe", but it is for monthly precipitation totals. The use of distinct file is recommended, because in ACMANT, precipitation data are converted (quasi-logarithmic transformation). With this transformation the distribution of monthly totals are closer to the normal distribution, hence the empirical correlations for the transformed values characterise better the spatial coherence.

ix) "Transform_dailydata.exe". Before the use of ACMANT, a general quality control of input data is recommended. The present software package supports the joint execution of "ExtraQC" and homogenization with ACMANT3, therefore the package includes the data reformatting program from ExtraQC output to ACMANT daily homogenization input (see details in sect. 6).

3. Preparation of input datasets

3.1. For all kinds of input data files

i) Number of time series per network: between 4 and 99.

ii) Length of time series (in years): between 10 and 200.
iii) In the input dataset time series are separated into different files according to the 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) Values are shown in °C unit for temperature and in mm unit for precipitation.

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 temperature and precipitation data are recommended to be presented with 1 decimal preciseness, although 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) Trace precipitation must be coded with 0.

ix) Justification of the input file lengths
Although the lengths of observations from different stations can be different, each file must contain the same number of lines. It is managed in a way that the files of relatively short time series are filled with -999.9 at the beginning and/or the ending sections of the time series. For instance, if in a temperature network the starting date of the observations is usually in 1951, but for a specific series the starting date is later, in April of 1953 only. Then the first three lines of the monthly input may appear as follows:
1951 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9
1952 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9 -999.9
1953 -999.9 -999.9 -999.9 8.6 14.2 16.8 18.8 18.5 14.4 9.5 3.6 -0.2

Naturally, missing value codes are not only for such a justification, so that they may occur in any part of the dataset.

The violation of the rules of input data preparation usually causes the stop of the homogenization procedure with an error message.

3.2. Specifics for monthly input datasets

i) Time-resolution is monthly.

ii) Name of files: "ANAMEJJ.txt".
- "ANAME" is the name of the network. This substring is always 5-character wide and may contain any alphanumeric characters.
- "JJ" is the serial number of time series.
- ".txt": these 4 characters are fixed, they are always ".txt"
Example: If the name of the network is "Sky1b" and 10 time series belong to the network, then the names of the time series must be:

Sky1b01.txt
Sky1b02.txt
Sky1b03.txt
Sky1b04.txt
Sky1b05.txt
Sky1b06.txt
Sky1b07.txt
Sky1b08.txt
Sky1b09.txt
Sky1b10.txt

iii) Each line (after the headline) is for one calendar year. Thus the number of lines in file is exactly 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 temperature 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: "ANAMEJJd.txt".
- "ANAME" is the name of the network. This substring is always 5-character wide and may contain any alphanumeric characters.
- "JJ" 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 network is "Smart" and 7 time series belong to the network, then the names of the time series must be:

Smart01d.txt
Smart02d.txt
Smart03d.txt
Smart04d.txt
Smart05d.txt
Smart06d.txt
Smart07d.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 exactly 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 daily temperature or precipitation values in their natural 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.

3.4. Specifics for applying Rtcheck.exe or Rpcheck.exe

i) Each input file is for one monthly time series prepared according to the rules of Sect. 3.2.

ii) Each file must hold a 11-character name (together with extension), but otherwise the name and extension can be freely chosen.

iii) The "Names.dat" file must be prepared with the list of the input file-names.

iv) Name of network is not asked, since there are no networks yet.

v) Number of input time series: Between 4 and 800.

3.5. Conditions for treatment and homogenization by ACMANT

i) Requirements for the treated sections of time series
Due to the justification, input files often contain sections with missing data. During the homogenization procedure the input time series are split into three sections: a treated section, and two not treated sections, one in the beginning part and another one in the ending part of the time series. When the number of missing values is low, the not treated

sections may have zero extension. On the other hand, the ratio of missing values within the treated section is limited, the requirements are as follows:

a) In the first k year and last k year of the treated section 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 section if the latter is the shorter. For instance, if the first year is 1953 and the last year is 2002, the number of available data for periods 1953-1953, 1953-1954, 1953-1967 and 2000-2002 are at least 4, 7, 46 and 10, respectively.

b) The minimum length of treated sections is 10 years.

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

d) The minimum number 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 temperature is "observed" when there are no more than 7 missing daily values in month and the status is "missing" in the reverse case. The status of monthly precipitation total is "observed" only if all the daily data are observed.

If the treated period of a time series is too short, that time series will be excluded from the homogenization procedure, but the procedure does not stop for this reason.

ii) Requirements for homogenizing series or their sections

a) If a station series (or its section) does not have at least 3 partner series with sufficient spatial correlation (R) 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. The minimum condition for performing homogenization with ACMANT is at least 3 partner series with $R \geq 0.50$ for each reference composites, or at least 4 partner series with $R \geq 0.40$.

b) Reference series or their sections are expected to be homogenizable in the same network as in which the candidate series is subjected to the homogenization. It is because ACMANT uses pre-homogenized reference series in the procedure. Consequently, if the number of homogenizable series is lower than 4, none of the time series will be homogenized.

c) When there is a section of candidate series without 3 partner 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 longer section could be homogenized in the early part of the candidate series. If more than one separate homogenizable sections are registered by ACMANT then this information will appear in the output file "Control message.txt".

4. Running the executive files

4.1. Parameters of the input dataset

Before running, all the input files must be put into the directory of the executive file. The running starts after a simple click on the .exe file. In the beginning of the running, some parameters must be input from the console. The following four parameters are asked in each version: a) name of the network (exception: it is not asked for Rtcehck and Rrcheck), b) number of stations, c) starting year of the time series (after the justification of the length of time series), d) length of time series (number of years after justification). In certain version further questions are asked. The values of the parameters are always asked one-by-one. The parameters input to the program have to comply with the rules described in Sect. 2, thus, for instance, the input will not be accepted if shorter than 10 years length, more than 200 stations, etc. are tried to be input to the program. For starting year, values between 1100 and 2020 can be introduced. Network-names of less than 5 characters are not allowed. In case of more than 5 characters are typed as network-name, the input will be accepted, but only the first five characters of the name will be retained by the program. After a failed inputting of a parameter-value, the request for inputting a value of that parameter appears again. If there is no formal problem with the parameter-values input, but they do not agree with the content of the input files, the program will stop with some error-message.

In homogenizing precipitation series, the first and last snowy months are asked. A month is snowy, if more than 50% of the precipitation for that is solid. The snowy season is the same for the whole network by definition, so in case of substantial spatial variation within network, the most typical (most frequent) season must be input.

4.2. Other parameters requested

a) At the beginning of monthly data homogenization or daily precipitation homogenization, ACMANT asks if the user requires outlier filtering or not. This outlier filtering works on monthly scale and spatial outlier values are searched. Generally the "Yes" is recommended to this question, however, the confidence of filtering relative outliers in an automatic process is not absolute. If the user knows well that his data is quit of outliers, automatic outlier filtering can be switched off. An alternative suggestion is that after running ACMANT with outlier filtering, the detected outliers can be checked manually for final decision: The true outliers should be corrected or substituted with missing data code, and the running of ACMANT can be repeated without outlier filtering. Caution! If you run ACMANT more than one times with the same data, always the raw data must be input, corrections are allowed only for solely outliers or short (less than 5 months) outlier periods. In ACMANT, biases of shorter than 5 month duration are treated as outliers.

In daily temperature homogenization the check of monthly outliers is connected to the check of daily data, therefore in daily data homogenization the outlier filtering of monthly data cannot be switched off (and such question does not appear among the input parameter requests).

b) The program asks the user if he accepts the default output package or not. If the answer is no, a series of questions follows, offering different output options. The output items are presented in Sect. 5.

5. Results of the ACMANT homogenization procedure

The output consists of various versions of homogenized series, a summery of the detected breaks and outliers, a file with monthly indicators about the treatment during the homogenization procedure, the list of spatial correlations and one or more control messages. At the beginning of the program the question "Do you prefer default outpackage?" question appears and if the answer is yes (Y) the output will include the items of Sect. 5.1 (Sect. 5.2) in monthly homogenization (daily homogenization). If the answer is N, the user will specify one by one, which output items of Sect. 5.1, 5.2, 5.3 he requires.

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 ANAMEJJt.txt, where 'ANAME' is the identifier of the network, JJ is the serial number of the time series and the ending part 't.txt' is a constant indicator of the file type. It contains the homogenized time series, but note that the file (and other versions of homogenized series) is always generated even when no section of the series has been homogenized. So that the number of files of this kind is the same as the number of time series in network. The data format is exactly the same as the input data format.

ii) Indicator of reliability for individual monthly values. The filename format is ANAMEJJj.txt, where the meaning of ANAMEJJ is the same as in i) and 'j.txt' is the indicator of the file type. The number of files of this kind is the same as the number of time series in network. The data format is similar to that of the homogenized temperature or precipitation, but instead of climatic values, indicator values, whole numbers between 0 and 9 are shown. Lower values (except 0) mean higher reliability, see detailed description in Sect. 5.4.

iii) Summary of the homogenization process with the list of the detected change-points and outliers and some other characteristics. The name of this file has the form of ANAMEbrk.txt. ANAME is the name of the network, while 'brk.txt' is a constant part of the name for this kind of files. The organisation of the data 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 change-points, 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, the starting and ending years of the homogenized period, the number of detected change-points and the number of detected outliers (in outlier filtering switched off mode the latter character does not appear). In lines of part (b), the properties of the detected change-points are presented. They are the serial number, the year and month of the timing, the size of change in the mean. In lines of part (c), the properties of the detected outliers are presented, i.e. the serial number and the timing with the data of year and month, and the duration of outlier period (in months) as well.

– Treated period always includes the entire homogenized period. Homogenized period may have the same extension as the treated period.
– If no change-point is detected in the time series, there is no b) part in the sub-list.
– If no outlier is detected in the time series, there is no c) part in the sub-list. The most frequent value for the duration of outlier period is 1, i.e. one month, but any value between 1 and 5 may occur.
– If homogenization has not been performed, character "0" is written into the place of the starting and ending years of homogenized period, and "-1" into the place of the number of change-points.
– It may occur that a time series has not treated period. In this case "0" is shown in the place of the starting and ending years of treated period.
– With precipitation homogenization, two shift sizes are shown for each breakpoint, when rainy and snowy seasons are defined. In these cases the first shift size is for the rainy season and the second is for the snowy season. Two shift sizes are shown also with ACMANT3monthsinus.exe and ACMANT3daysinus.exe, in these cases the first value shows the shift in the annual mean and the second value is for the shift in the summer – winter difference. By contrast, in case of irregular shaped seasonality of biases (with ACMANT3monthirreg.exe and ACMANT3dayirreg.exe), no information of the seasonality is shown.
– In daily temperature homogenization, in part (b) timing is always shown with daily preciseness, although ACMANT works on daily scale only when the pre-estimated shift magnitude is large. Therefore the calendar day is often the last day of the month. Also in daily temperature homogenization, more pieces of data are shown when outliers had been detected on monthly scale. In these cases short-term platform shaped inhomogeneities are searched on daily scale, and the results are shown with daily preciseness of the estimated period of the bias, and the size of the bias as well. However, sometimes missing data code (-999.9) stands in the place of the bias size. It is either because the detected monthly outlier does not show platform shaped structure on daily scale, or for the lack of sufficient observed data in partner series to produce a reliable estimation of the bias size.

iv) Spatial correlations between time series. The name of the file has the form of ANAMEr1.dat. ANAME is the name of the network, while "r1.dat" is a constant part of the name for this kind of files. The file comprises *N* lists of the spatial correlations where *N* stands for the number of stations. A particular list shows the spatial correlations of one candidate time series with all the other series. The spatial correlations are calculated from the increment series of monthly values in the same way as with Rtcheck.exe and Rpcheck.exe. However, a difference is that this side product of the homogenization shows the correlations for outlier-filtered and pre-homogenized data, while Rtcheck.exe and Rpcheck.exe use raw data. The form of data is as follows: Each particular list of the file has a headline. It includes (from the left to the right) the serial number of the candidate series, the starting and ending years of its treated period, and its station identifier. Then, in the following lines, the spatial correlations with each time series of the network are shown. In each line of the list after the headline, the serial number of the partner series, the correlation with the partner series and the name of the partner series are shown. If a calculated correlation is lower than 0.4, zero is shown, since in this case ACMANT considers the spatial relationship to be zero. If less than 50 value-pairs were available for the calculation, 0 is shown. Where the partner series is identical with the candidate series, 0 is shown.

Note that the results of Rtcheck and Rrcheck have the same format as ANAMEr1.dat except that in the former ones instead of 0 code, 9.999 stands.

v) Control message. In running ACMANT, "Control message.txt" file is generated. If there is no problem during the run of the program, "….homogenization has been completed" will be the only message. However, several other messages are possible.

Note that if you would like to have homogenized data in the format of the input of extreme index calculations (i.e. each month is in separate line in monthly data and each day is in separate line in daily data) you must answer "No" to default output package, and answer yes when the relevant question appears among the offered optional output items.

5.2. Default output package of daily homogenization

i-v) The same items in the same form as in monthly homogenization, see Sect. 5.1.

vi) Homogenized daily series with gap filling within the homogenized period. Their names have the form of ANAMEJJv.txt for temperature. 'ANAME' is the identifier of the network, JJ is the serial number of the time series and the ending part 'v.txt' is the indicator of the file type. The format of data within files is the same as in the input files.

5.3. Optional output items

The following items are output only on request.

vii) Homogenized monthly series without gap filling. The file name has the form of
ANAMEJJs.txt, the only formal difference relative to ANAMEJJt.txt is that the eights
letter is 's'. This output type can be chosen either with daily or monthly homogenization.

viii) Homogenized monthly series with completing the time series with interpolated data
(in case it is in need) for the entire period involved in the homogenization procedure. The
file names has the form of ANAMEJJh.txt, the only formal difference relative to
ANAMEJJt.txt is that the eights letter is 'h'. This file can be chosen either with daily or
monthly homogenization.

ix) Homogenized daily temperature series without gap filling. The file name has the form
of ANAMEJJu.txt. It is similar to ANAMEJJv.txt , the only difference is that gap filling
is not shown for this file.

x) Homogenized daily series with completing the time series with interpolated data (in
case it is in need) for the entire period involved in the homogenization procedure. The file
name has the form of ANAMEJJg.txt, the only formal difference relative to
ANAMEJJv.txt is that the eights letter is 'g'.

xi) Series of reliability indicators for daily values. The file name has the form of
ANAMEJJi.txt, the difference relative to ANAMEJJv.txt is that the eights character of
the name is "i". The data format is similar to that of the homogenized temperature or
precipitation, but instead of climatic values, indicator values, whole numbers between 0
and 9 are shown. Lower values (except 0) mean higher reliability, see detailed description
in Sect. 5.4.

xii) Homogenized monthly series in one column form. It can be chosen either with daily
or monthly homogenization. The file name has the form of ANAMEJJm.txt. In each line
of the file the year, month and the monthly value are shown. Only the homogenized
period is included. Data-gaps are always filled with interpolated values.
In case of precipitation homogenization, the values are multiplied with 100. This item can
directly be applied as input for SPI (Standardised Precipitation Index) calculation
(http://digital.csic.es/handle/10261/10006).

xiii) Homogenized daily series in one column form. The file name has the form of
ANAMEJJn.txt. In each line of the file the calendar year, month, day and the daily value
are shown. Only the homogenized period is included. Missing data may occur and they
are coded with -99.9. This form is harmonized with the input data requirements of the
extreme index calculations with RClimdex and quality control with ExtraQC
(http://www.c3.urv.cat/data.html).
        In RClimdex and ExtraQC precipitation, Tmax and Tmin values are required to
be in the same line. The easiest way to produce it from ANAMEJJn.txt files is that the

precipitation, Tmax and Tmin files are opened one-by-one by Excel, and the 3 relevant columns of the 3 files can be copied next to each-other in one only sheet. Be cautious with the dates, since the homogenized periods may differ for the different variables.

xiv) Raw and homogenized series of monthly and annual temperature anomalies of the homogenized period. Raw series: <u>ANAMEJJb.txt,</u> homogenized series: <u>ANAMEJJf.txt</u>. These two output items can be chosen together or either of them, in homogenizing monthly temperature or daily temperature data. In the files each line contains the values for one calendar year, but apart from several other output items,14 values are shown in a line, since just after the number of year the annual mean is shown. In the files ANAMEJJb.txt raw values are shown, without gap filling and homogenization. Annual values are shown when at least 10 months of the year have observed values. ANAMEJJf,txt files shows completed and homogenized anomalies.

5.4. Meaning of reliability indicators

Although the file ANAMEbrk.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 for the entire treated period, but when the number and/or correlation with series is low, the confidence of interpolated values is limited.

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

In all programs except for daily temperature homogenization:

0 – sporadic observed value, out of the treated period of the time series.
1 – homogenized value
2 – observed value, not homogenized
3 – interpolated value from sufficient number of other values, within the homogenized period
4 - interpolated value from sufficient number of other values, out of the homogenized period
5 – interpolated value from limited number of other values, within the homogenized period
6 – interpolated value from limited number of other values, out of the homogenized period
7 – interpolated value, with the use of 1 only observed value
8 – missing value is substituted with the climatic normal value
9 – missing data

15

In daily temperature homogenization (ACMANT3daysinus.exe and ACMANT3dayirreg.exe)

0 – sporadic observed value, out of the treated period of the time series.
1 – homogenized value
2 – not applied
3 – daily values are treated on daily scale, after monthly outlier had been detected.
4 – observed value, not homogenized
5 – interpolated value from sufficient number of other values, within the homogenized period
6 – interpolated value from limited number of other values in the homogenized period, or interpolated value out of the homogenized period
7 – interpolated value, with the use of only 1-2 observed values
8 – missing value is substituted with the climatic normal value
9 – missing data

In daily homogenization, both daily and monthly indicator values are generated. The monthly code always equals to the eighth highest daily code in the month.

Notes
a) "Observed value not homogenized" means that break search and outlier check were not performed at the timing of the observed value, but such values might hold adjustment terms due to detected breaks in some later section of the time series;
b) Codes of interpolated values are shown even when output with interpolated values is not required;
c) Code 9 is shown for the missing data out of the homogenized period as a default. However, when the output package includes completed data to the entire period, code 9 is not applied.

6. Joint use of ExtraQC and ACMANT3

i) Input data must be prepared in the format of ExtraQC input. This format is the same as which is applied in RClimDex use and the description can be found at http://etccdi.pacificclimate.org/RClimDex/ .

ii) A particularity of the joint application of ExtraQC and ACMANT3 is that the files of your input data for ExtraQC must hold file names of the form "ABCJJ.txt" where "ABC" is the network name and JJ is the serial number of the station. "ABC" may consist of any

alphanumeric characters, but the number of characters is fixed by 3. A network may contain 4 to 99 stations of the same climatic region. For instance, if you have 7 stations in a network named "Cat", the file names must be

Cat01.txt
Cat02.txt
Cat03.txt
Cat04.txt
Cat05.txt
Cat06.txt
Cat07.txt

It is recommended to create a file in which the matching between original station names or codes and the new input file names is kept.

iii) Run ExtraQC (see its manual at http://www.c3.urv.cat/data.html).

iv) Correct the found errors or put missing data code where reliable correction is not possible.

v) Prepare the parameter file "Paramf.txt". Its default content is:

Name of the network (First 3 characters of the next line)
ABC
Number of stations
22
First and last years for precipitation homogenization
1924   2020
First and last years for temperature homogenization
1772 1988

a) Please change the first three characters of the second line "ABC" to the actual name of the network (in the example: Cat)
b) Please write the actual number of stations, and actual periods covered with precipitation data and temperature data.
Notes: The time period indicated can be longer than the period really covered with data. Remember, that ACMANT needs 4 comparable time series, therefore periods with data of less than 3 stations are uninteresting. If you do not have data of each kind of variables included in ExtraQC, (e.g. you have only precipitation data), you may write anything to the period of the missing variable, but if you prefer not to keep unnecessary data, put higher number to the starting year than to the ending year for the missing variable, and the result will be empty files for that variable.

vi) Click on "Transform_dailydata.exe"

vii) Continue with running the daily homogenization programs of ACMANT3 (Sect. 4 and Sect. 5 of this Manual).