

Magnus Joelsson, Nils Slättberg, Alicia Carnebring, Christophe Sturm, and Erik Engström

Automation of the interactive mode of the homogenisation software HOMER for climatological applications

Motivation

- SMHI requires homogenisation of observations from the network of weather stations for climatological studies of temperature, precipitation, wind, and other measurables
- A homogenisation protocol needs to be established including the selection of a homogenisation tool

Evaluation of homogenisation tools

- Four methods are evaluated
 - HOMER-automatic
 - HOMER-interactive (with criteria previously used at SMHI)
 - CLIMATOL (Guijarro, 2014)
 - ACMANT (Domonkos, 2014)
- Three variables are considered
 - Maximum daily temperature, monthly mean
 - Minimum daily temperature, monthly mean
 - Daily mean temperature according to Ekholm-Modén's method (SMHI, 2014), monthly mean (observations data set only)
- Two sets of data is used:
 - Indecis project homogenisation benchmark dataset (Indecis project, 2017), "Angry Baboon" (with missing values and homogeneity break points inserted). Two networks with data in 1950–2005:
 - Swedish network, 100 time series
 - Slovenian network, 30 time series
 - Observations for daily maximum and minimum (daily mean) temperature from 77 (81) coupled time series from SMHI's network of weather stations in 1880–2019 (1860–2019)

HOMER

- HOMER (Mestre et al. 2013) is a well-established homogenisation software (e.g. Kivinen et al. 2017, Morán-Tejeda et al. 2016, Vicente-Serrano et al. 2017)
- HOMER combines the results of essentially four methods for detection of homogeneity breaks:
 - Pairwise comparison (with individual, well-correlated, reference time series) on annual data
 - Pairwise comparison on winter and summer data
 - Joint comparison with larger parts of the network
 - ACMANT-style, which includes detection of seasonal cycle range inhomogeneities
- HOMER has two modes "automatic" and "interactive":
 - Automatic homogenisation is dissuaded (Gubler et al. 2017) and still requires an operator
 - Interactive homogenisation is:
 - Time-consuming
 - Toilsome
 - Subjective
 - Error-prone
 - May not be reproducible
 - Difficult to evaluate

Problem

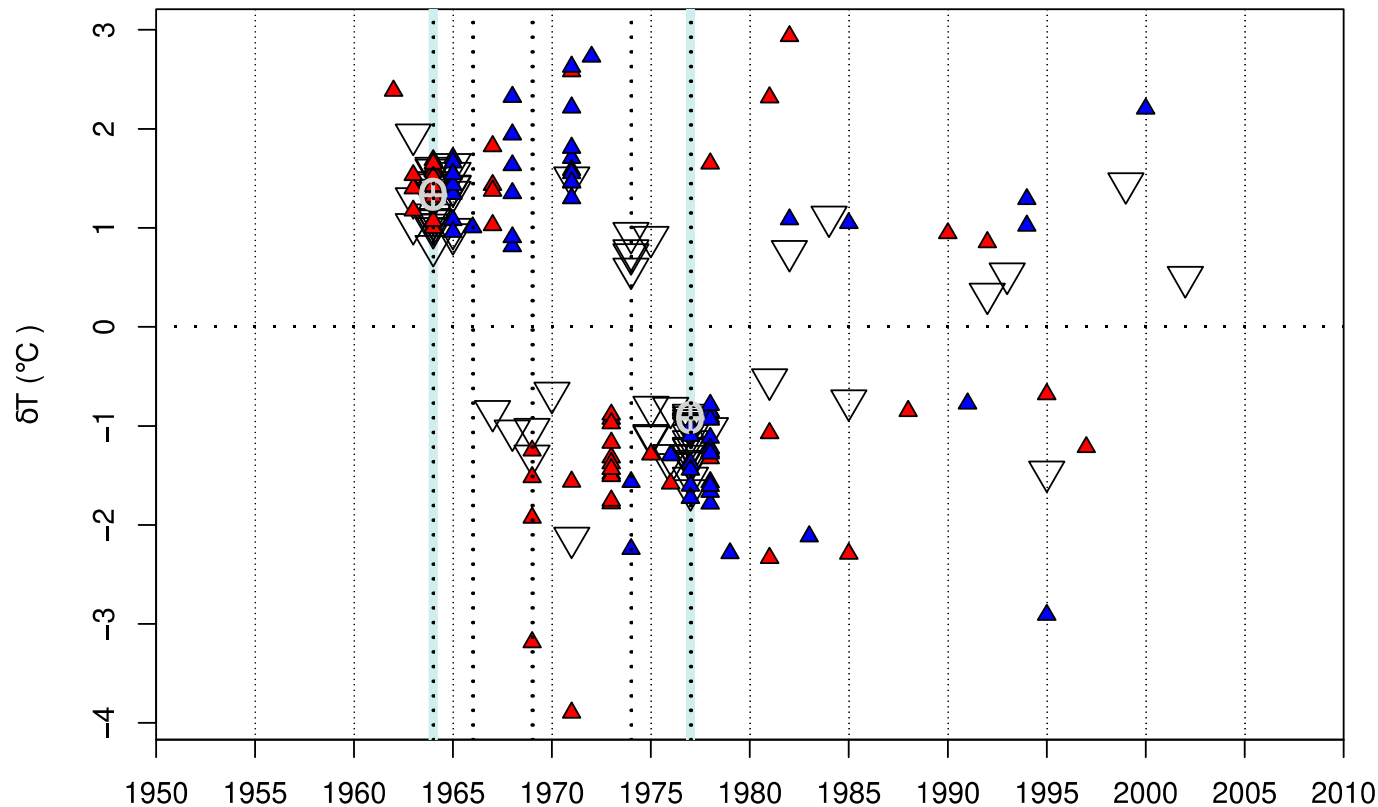
- In order to evaluate HOMER and possibly use it operational the software must be automatised
- (At least) two different sets of criteria for the acceptance/rejection of homogeneity breaks must be possible to apply:
 - The criteria used by the HOMER-automatic method (accept all and only breaks reported by either the joint-detection or the ACMANT-style methods)
 - The criteria previously used at SMHI
- Furthermore, the homogenisation procedure should be possible to customise to meet certain needs or preferences, e.g.:
 - Only correct large and obvious breaks
 - Minimize the error at each data point
 - Reproduce the network wide linear trend

The SMHI acceptance criteria

- For each candidate break (a year in a time series with a possible homogeneity break), all the reported breaks from each method at the same year of a time series are counted. Note that the pairwise-detection methods can report as many breaks as there are reference series (i.e. well-correlated neighbours).
- The breaks reported from the seasonal method counts only as 1/3 break.
- If the reported breaks sum up to four or more (or for example three breaks reported by annual methods and three breaks reported by the seasonal method) the candidate break is accepted. If the breaks sum up to three, the candidate break is noted.
- If meta data or a following homogenisation round supports a noted break (i.e. if the break is already in the note book), the candidate break is accepted
- Reported breaks in the year previous or following the year of a candidate break adds to the candidate break, but with a penalty of one break
- Breaks can not occur in consecutive years in the same time series

The SMHI acceptance criteria, example

Daily maximum temperature at station "34" of the Indecis benchmark dataset Swedish network



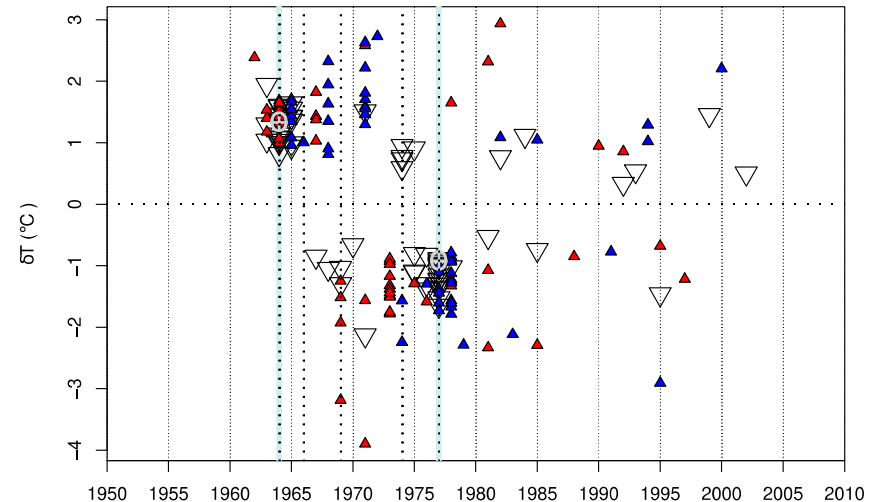
- Large triangles: Pairwise-detection, annual
- Small red triangles: Pairwise-detection, summer
- Small blue triangles: Pairwise-detection, winter
- Cross-hairs: Joint-detection
- Dotted vertical lines: ACMANT-style detection
- Cyan vertical line: Accepted break (note that some extra criteria for acceptance are applied in the figure, hence some breaks are not marked)

The SMHI acceptance criteria, example

Year: 1964

- The year (1964)
Joint: 1, acmant 1, pair (annual): 15, pair (winter): 0, pair (summer): 11
- The previous year (1963)
Joint: 0, acmant 0, pair (annual): 3, pair (winter): 0, pair (summer): 3
- The following year (1965)
Joint: 0, acmant 0, pair (annual): 7, pair (winter): 0, pair (summer): 9
- Three-year breaks = $[3 + 3/3; 17 + 11/3; 7 + 9/3]$
breaks = $(4-1) + 20.7 + (11-1) = 33.7 \geq 4$

Note that e.g. 1974 should be accepted according to the SMHI acceptance criteria, however these criteria are developed for a more sparse network than what is presented here



- Large triangles: Pairwise-detection, annual
- Small red triangles: Pairwise-detection, summer
- Small blue triangles: Pairwise-detection, winter
- Cross-hairs: Joint-detection
- Dotted vertical lines: ACMANT-style detection
- Cyan vertical line: Accepted break (note that some extra criteria for acceptance are applied in the figure, hence some breaks are not marked)

Development of “Bart”, child of HOMER

- Declaration of intent:
 - Bart is an attempt to automatise the HOMER software
 - No detection or correction methods are altered (however some minor bugs are fixed)
 - The order of the HOMER tasks are predetermined
 - The acceptance or rejection of a homogeneity break point follows predetermined criteria
 - The removal of “empty breaks” are automatised
 - “Remote” operation is enabled
 - Meta data information can be included
 - Bart is possible to run in batch mode
- Bart is guided through the use of inputs, either given as arguments in the command line, or stored in a separate text-file
- The acceptance criteria is translated into a point system, in which points are given to a candidate break. If the points exceed an acceptance threshold, the candidate break is accepted.
- The gapfilling part of HOMER’s correction function is made into its own function to enable gapfilling of time series prior to, or without homogenisation

The point system

Basics, with default values

- Pairwise-detection, annual: 3 points
- Pairwise-detection, season: 1 point
- Joint-detection: 3 point
- ACMANT-style detection: 3 points
- Meta data break: 3 points
- Accepted break threshold: 12 point
- Noted break threshold: 9–12 points
- Adjacent penalty: 3 points
- “Safe space”: 1 year (thus far only 1 and 0 is tested)

Fancy extra parameters

- A amplitude threshold can be applied such that reported breaks smaller than a certain amplitude are disregarded, default threshold is 0
- Large amplitude breaks can be given more weight by a amplitude power:

points = (amplitude of break/normal amplitude of break)^{amplitude power}

The normal break amplitude can be predetermined or set to the median of the break amplitudes of the annual pairwise detection of raw data. The default value of the amplitude power is 0.

- Breaks in well-correlated series can be given more weight by a correlation power:

points = correlation coefficient of the reference vs. the candidate^{correlation power}

This can be used together with the amplitude power. The default value is 0.

Special rules

- Adjacent break points can only be added to one candidate year, the candidate break with highest points. The method checks if the year previous or following the adjacent year has more points than the candidate break year, before adding the points to the candidate. The candidate break year must also have higher points than the adjacent year itself.
- Reported breaks will only be added if they have the same sign. Groups of winter, summer, and annual breaks can however have different signs between the groups but not within the groups. The highest points count in the point sum. This is also applied to reported breaks in adjacent years.
- The pairwise-detection methods are normalised with the number of references and multiplied with the “normal” number of references as such:

points = number of breaks/number of references × minimum(“normal” number of references, number of references) × weight of method

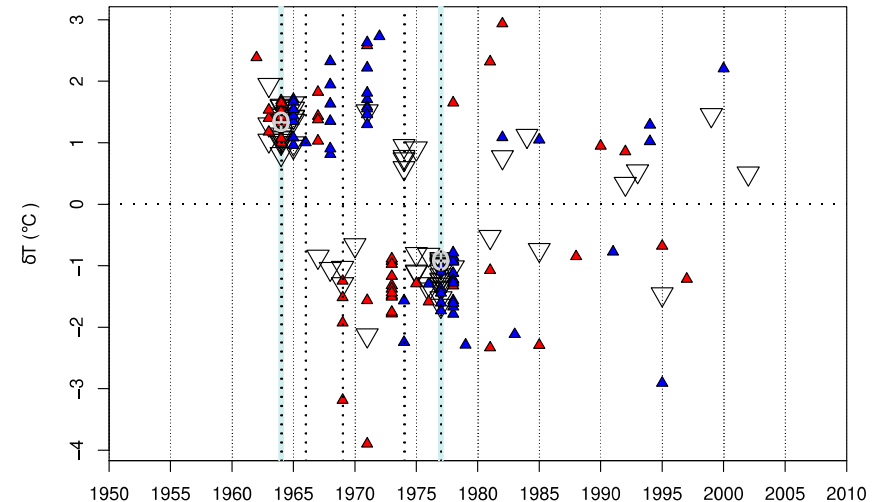
This is to give the pairwise method similar weight regardless of the size and density of the network or the region of a network. To not amplify break signals in sparse (and thus unreliable) regions the “normal” number of references-factor cannot be larger than the actual number of references. The default “normal” number of references is 10.

- New breaks will not be added if there are breaks accepted from an earlier test within the “safe space”-time, even if the signal is stronger for the new break. The reach of the “assess month”-method which usually is applied in the end of the process, is two years before and two years after the break. Breaks within two years of each other can therefore almost be considered identical

Point system, example with default values

Year: 1964 with the score 35.00

- The year (1964)
 Joint: 1, acmant 1, pair (annual, positive): 15, pair (annual, negative): 0, pair (winter, positive): 0, pair (winter, negative): 0, pair (summer, positive): 11, pair (summer, negative): 0, number of neighbours: 28
 $\text{points} = 3 + 3 + 15/28 \times 10 \times 3 + 11/28 \times 10 \times 1 = 26.00$
- The previous year (1963)
 Joint: 0, acmant 0, pair (annual, positive): 3, pair (annual, negative): 0, pair (winter, positive): 0, pair (winter, negative): 0, pair (summer, positive): 3, pair (summer, negative): 0
 $\text{points} = 3/28 \times 10 \times 3 + 3/28 \times 10 \times 1 = 4.29$
 Will not be counted to next previous year (with 0.36 points)
- The following year (1965)
 Joint: 0, acmant 0, pair (annual, positive): 7, pair (annual, negative): 0, pair (winter, positive): 0, pair (winter, negative): 0, pair (summer, positive): 9, pair (summer, negative): 0
 $\text{points} = 7/28 \times 10 \times 3 + 9/28 \times 10 \times 1 = 10.71$
 Will not be counted to next following year (with 3.36 points)
- Five-year points = [0.36; 4.29; 26.00; 10.71; 3.36]
 $\text{Points} = (4.29 - 3) + 26.00 + (10.71 - 3) = 35.00 \geq 12$



Point system:

- Pairwise-detection, annual (large triangles): 3 points
- Pairwise-detection, season (small triangles): 1 point
- Joint-detection (cross-hairs): 3 point
- ACMANT-style detection (dotted vertical lines): 3 points
- Accepted break (cyan vertical line): ≥ 12 points
- No meta data, amplitude threshold, amplitude power, or correlation power

Inputs to Bart

- Basics inputs and inputs inherited from HOMER:
 - Working directory, enables “remote” operation
 - Additive or multiplicative correction (only additive has been tested so far)
 - Correlation threshold
 - Minimum number of neighbours
 - Order of tasks, repeated either a predetermined number of times or until no further breaks are detected
- Parameters that control the acceptance criteria
 - Weight of breaks reported by the pairwise-annual method
 - Weight of breaks reported by the pairwise-seasonal method
 - Weight of breaks reported by the joint method
 - Weight of breaks reported by ACMANT-style method
 - Weight of breaks in the meta data
 - Adjacent penalty
 - “Normal” number of reference series
 - Break acceptance threshold
 - Note threshold
 - Amplitude threshold
 - Power of the correlation
 - Power of the amplitude
 - Break amplitude normal
 - “Safe space”: Time around an accepted break where additional breaks in the time series are disallowed (thus far 1 or 0 are implemented)
- The numerical inputs can also be given as lists such that each step in the homogenisation process can have a dedicated set of parameters

Sanity check: Comparison of detection lists

Daily minimum temperature of the Indecis benchmark dataset
Slovenian network

Bart with HOMER-automatic criteria

00000009	BREAK	1959	12	n	509
00000009	BREAK	1960	12	n	509
00000009	BREAK	1970	22	n	509
00000000	BREAK	1980	22	n	609
00000000	BREAK	1990	10	n	609
00000000	BREAK	1998	10	n	609
00000000	BREAK	1985	8	n	610
00000000	BREAK	2000	20	n	610
00000008	BREAK	1954	3	n	815
00000008	BREAK	1959	12	n	815
00000000	BREAK	1958	92	n	916
00000018	BREAK	1963	30	n	186
00000018	BREAK	1980	12	n	186
00000028	BREAK	1984	12	n	183
00000028	BREAK	1980	12	n	187
00000029	BREAK	1969	82	n	187
00000019	BREAK	1973	12	n	19
00000019	BREAK	1983	12	n	19
00000019	BREAK	1986	4	n	19
00000022	BREAK	1970	4	n	22
00000022	BREAK	1990	7	n	22
00000022	BREAK	1992	3	n	22
00000023	BREAK	1967	3	n	23
00000026	BREAK	1961	10	n	26
00000026	BREAK	1980	12	n	26
00000032	BREAK	1972	2	n	32
00000032	BREAK	2000	8	n	32
00000040	BREAK	1970	11	n	40
00000040	BREAK	1983	3	n	40
00000048	BREAK	1970	9	n	48
00000048	BREAK	1980	12	n	48
00000048	BREAK	1985	9	n	48
00000048	BREAK	1993	11	n	48
00000049	BREAK	1953	4	n	49
00000049	BREAK	1958	2	n	49
00000049	BREAK	1980	12	n	49
00000058	BREAK	1962	4	n	58
00000058	BREAK	1975	9	n	58
00000058	BREAK	1984	12	n	58
00000060	BREAK	1966	12	n	60
00000060	BREAK	1976	12	n	60
00000060	BREAK	1984	12	n	60
00000060	BREAK	1997	12	n	60
00000067	BREAK	1964	4	n	67
00000067	BREAK	1972	9	n	67
00000067	BREAK	1976	12	n	67
00000067	BREAK	1980	12	n	67
00000069	BREAK	1959	12	n	69
00000069	BREAK	1961	12	n	69
00000069	BREAK	1975	12	n	69
00000069	BREAK	1979	11	n	69
00000082	BREAK	1956	12	n	82
00000082	BREAK	1958	12	n	82
00000082	BREAK	1959	3	n	82
00000084	BREAK	1964	12	n	84
00000084	BREAK	1980	2	n	84
00000094	BREAK	1959	12	n	94
00000094	BREAK	1960	9	n	94
00000094	BREAK	1972	10	n	94
00000094	BREAK	1980	12	n	94
00000101	BREAK	1950	12	n	101
00000101	BREAK	1951	2	n	101
00000101	BREAK	1968	12	n	101
00000101	BREAK	1994	12	n	101
00000107	BREAK	1905	12	n	107
00000109	BREAK	1960	3	n	109

HOMER-automatic

00000009	BREAK	1959	12	n	509
00000009	BREAK	1960	12	n	509
00000009	BREAK	1970	22	n	509
00000000	BREAK	1980	22	n	609
00000000	BREAK	1990	10	n	609
00000000	BREAK	1998	10	n	609
00000000	BREAK	1985	8	n	610
00000000	BREAK	2000	20	n	610
00000008	BREAK	1954	3	n	815
00000008	BREAK	1959	12	n	815
00000000	BREAK	1958	92	n	916
00000018	BREAK	1963	30	n	186
00000018	BREAK	1980	12	n	186
00000028	BREAK	1984	12	n	183
00000028	BREAK	1980	12	n	187
00000029	BREAK	1969	82	n	187
00000019	BREAK	1973	12	n	19
00000019	BREAK	1983	12	n	19
00000019	BREAK	1986	4	n	19
00000022	BREAK	1970	4	n	22
00000022	BREAK	1990	7	n	22
00000022	BREAK	1992	3	n	22
00000023	BREAK	1967	3	n	23
00000026	BREAK	1961	10	n	26
00000026	BREAK	1980	12	n	26
00000032	BREAK	1972	2	n	32
00000032	BREAK	2000	8	n	32
00000040	BREAK	1970	11	n	40
00000040	BREAK	1983	3	n	40
00000048	BREAK	1970	9	n	48
00000048	BREAK	1980	12	n	48
00000048	BREAK	1985	9	n	48
00000048	BREAK	1993	11	n	48
00000049	BREAK	1953	4	n	49
00000049	BREAK	1958	2	n	49
00000049	BREAK	1980	12	n	49
00000058	BREAK	1962	4	n	58
00000058	BREAK	1975	9	n	58
00000058	BREAK	1984	12	n	58
00000060	BREAK	1966	12	n	60
00000060	BREAK	1976	12	n	60
00000060	BREAK	1984	12	n	60
00000060	BREAK	1997	12	n	60
00000067	BREAK	1964	4	n	67
00000067	BREAK	1972	9	n	67
00000067	BREAK	1976	12	n	67
00000067	BREAK	1980	12	n	67
00000069	BREAK	1959	12	n	69
00000069	BREAK	1961	12	n	69
00000069	BREAK	1975	12	n	69
00000069	BREAK	1979	11	n	69
00000082	BREAK	1956	12	n	82
00000082	BREAK	1958	12	n	82
00000082	BREAK	1959	3	n	82
00000084	BREAK	1964	12	n	84
00000084	BREAK	1980	2	n	84
00000094	BREAK	1959	12	n	94
00000094	BREAK	1960	9	n	94
00000094	BREAK	1972	10	n	94
00000094	BREAK	1980	12	n	94
00000101	BREAK	1950	12	n	101
00000101	BREAK	1951	2	n	101
00000101	BREAK	1968	12	n	101
00000101	BREAK	1994	12	n	101
00000107	BREAK	1905	12	n	107
00000109	BREAK	1960	3	n	109

Results, Indecis benchmark dataset, numbers

Table: Heidke Skill Score (HSS), Peirce Skill Score (PSS), hits per reported break (Hit rate), Probability Of Detection (POD), number of breaks (Num. brk.), and Centred Root Mean Square Error (CMSE) for homogenisation over four cases (Slovenia- T_{\min} , Slovenia- T_{\max} , Sweden- T_{\min} , Sweden- T_{\max})

Method	HSS	PSS	Hit rate /%	POD /%	Num. brk.	CRMSE* /°C	Trend dev. /°C (10 a)
Bart-SMHI	0.62	0.58	70	59	102	0.70	0.009
HOMER- auto	0.55	0.60	54	63	164	0.70	0.026
CLIMATOL	0.68	0.61	79	61	97	0.73	0.012
ACMANT	0.69	0.75	66	77	182	1.14	0.015

*Only data points with non-missing raw data

Maximum daily temperature, Slovenia

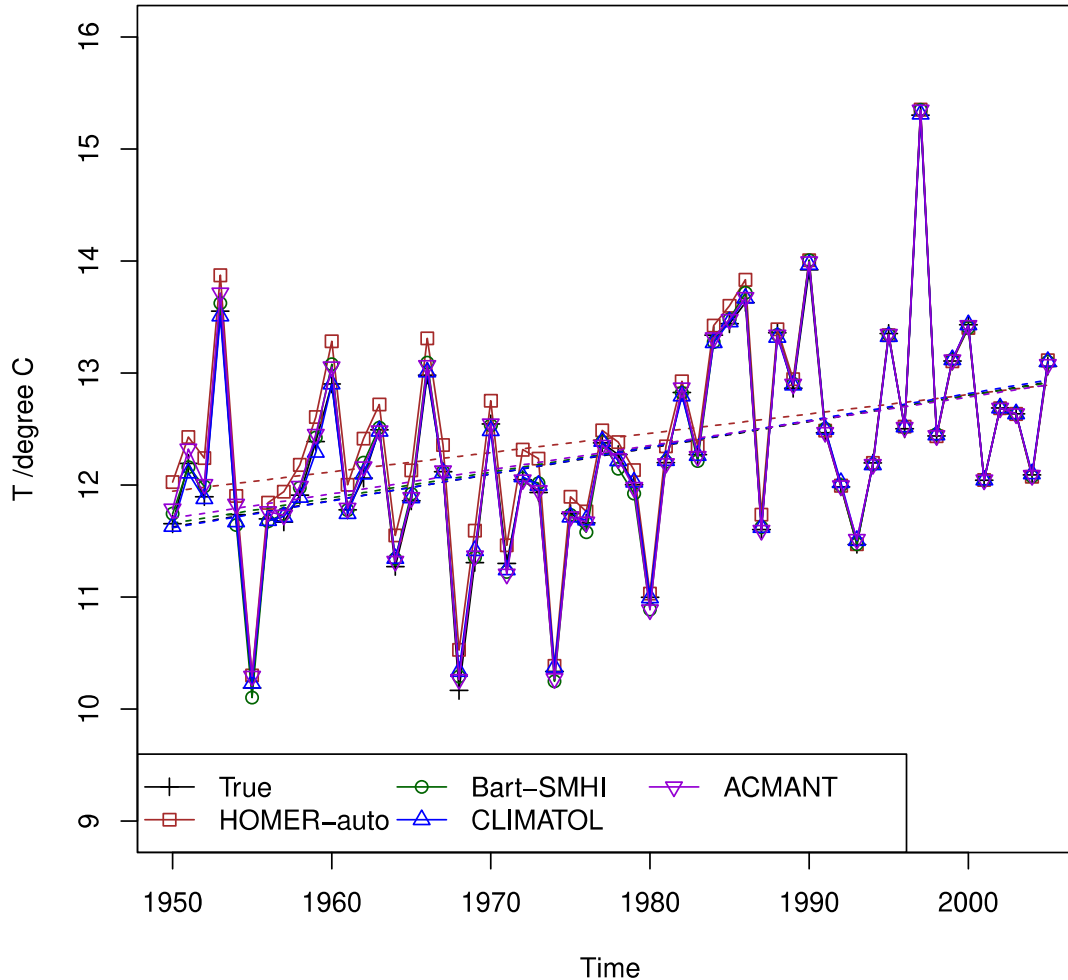


Table: Network average trend

Method	Trend / °C (10 a) ⁻¹
True	0.243
HOMER-auto	0.178
Bart-SMHI	0.235
CLIMATOL	0.249
ACMANT, post-gapfilled	0.223

Figure: Network average of homogenised data

Maximum daily temperature, Sweden

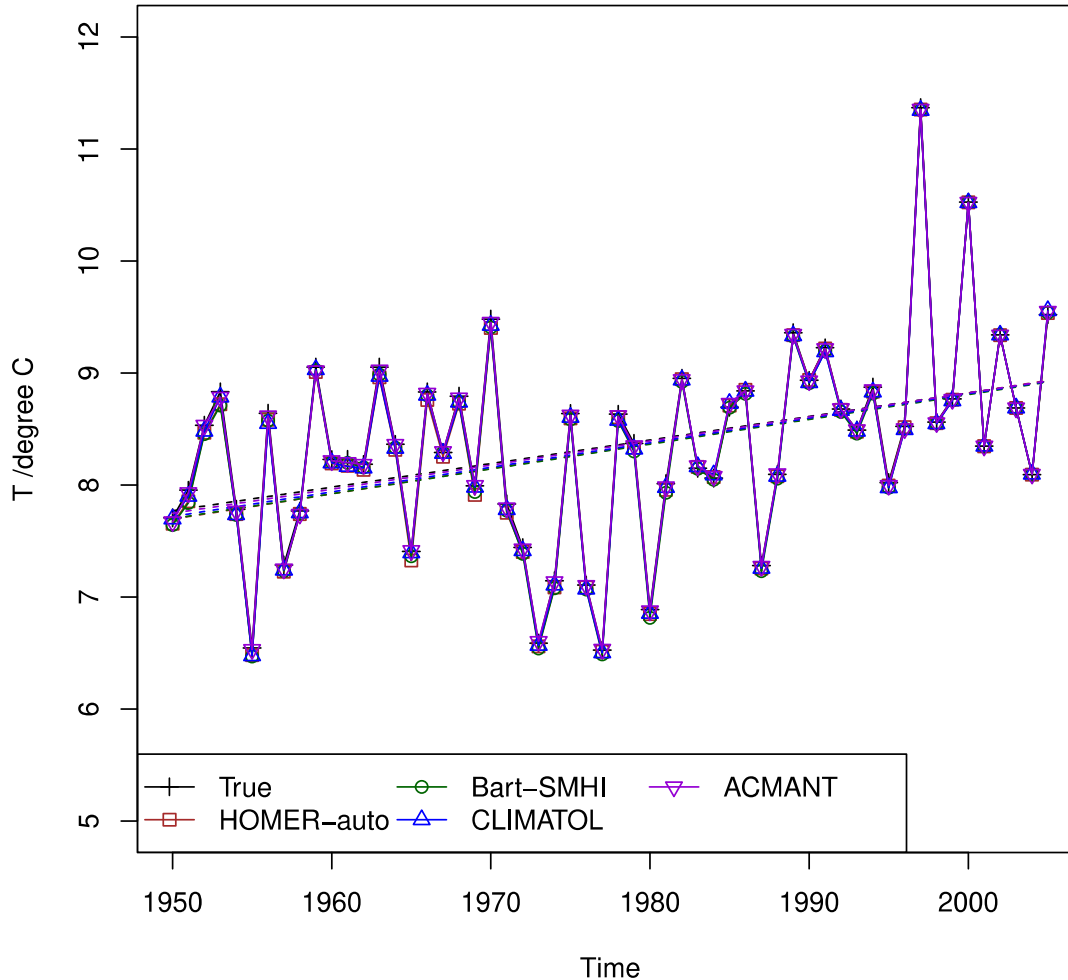


Table: Network average trend

Method	Trend / °C (10 a) ⁻¹
True	0.213
HOMER-auto	0.227
Bart-SMHI	0.223
CLIMATOL	0.221
ACMANT, post-gapfilled	0.217

Figure: Network average of homogenised data

Minimum daily temperature, Slovenia

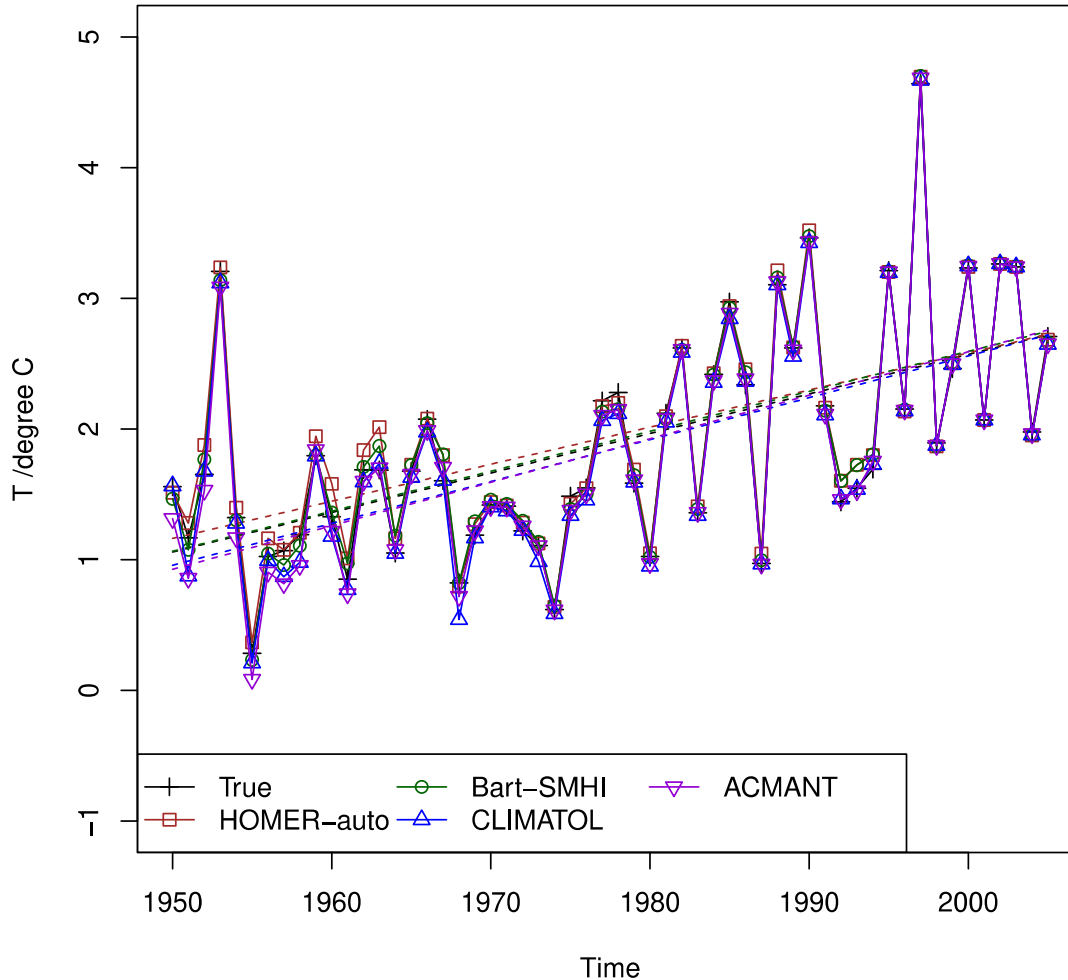


Table: Network average trend

Method	Trend / °C (10 a) ⁻¹
True	0.313
HOMER-auto	0.294
Bart-SMHI	0.316
CLIMATOL	0.332
ACMANT, post-gapfilled	0.344

Figure: Network average of homogenised data

Minimum daily temperature, Sweden

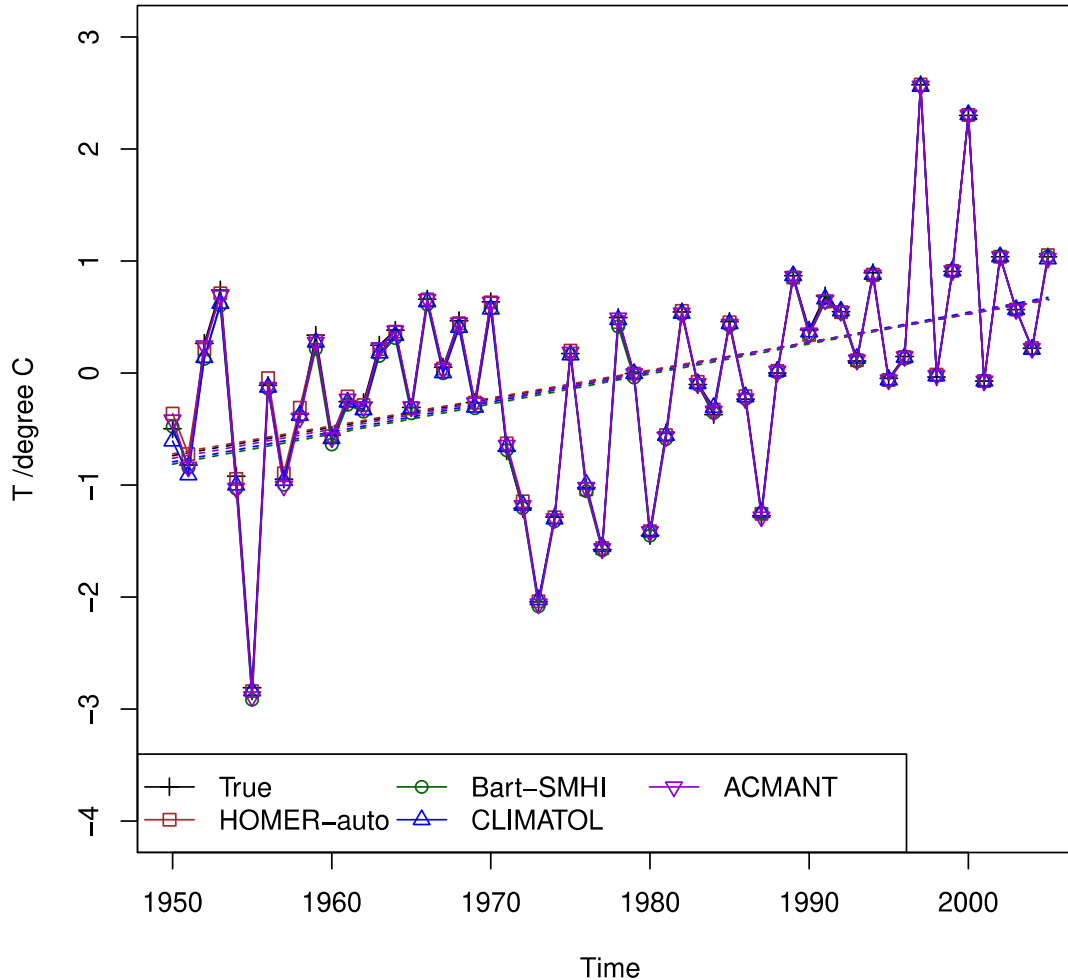


Table: Network average trend

Method	Trend / °C (10 a) ⁻¹
True	0.256
HOMER-auto	0.252
Bart-SMHI	0.271
CLIMATOL	0.270
ACMANT, post-gapfilled	0.261

Figure: Network average of homogenised data

Results, observational dataset

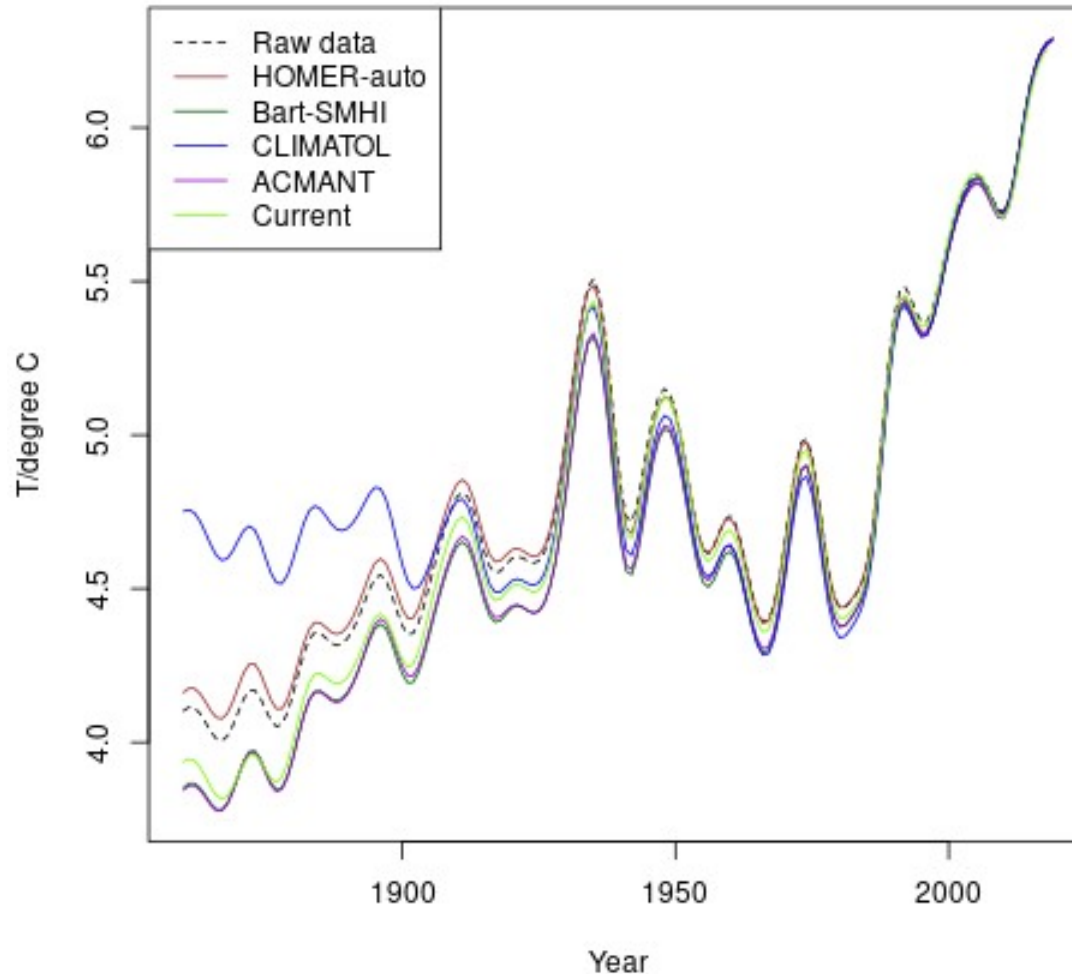


Figure: 35 selected station average of homogenised data, mean daily temperature (Ekholm-Modén), Sweden, along with raw data (black dashed line), “current” is data from the current official SMHI homogenised data set (SNHT method)

Summary

- The standard method for homogenisation HOMER, have been fully automatised in the novel “Bart” script
 - Bart enables
 - Easy evaluation of (especially the interactive mode of) HOMER
 - Tailoring of detected homogeneity break acceptance criteria
 - Batch mode run of HOMER in both automatic and interactive mode
 - Bart has successfully replicated
 - Homogenisation by HOMER automatic mode
 - The network wide trends of the current SMHI homogenised data set used for climate analysis
- Bart’s performances with the current HOMER criteria used at SMHI (Bart-SMHI) and the HOMER-auto criteria (HOMER-auto) have been evaluated along with the fellow standard homogenisation tools CLIMATOL and ACMANT using the Indecis benchmark data set
 - The results of the evaluation test shows that Bart-SMHI/HOMER-auto overall performance compare well with the other methods
 - CLIMATOL has the most hits per accepted break of the methods
 - ACMANT has the most hits per real breaks of the methods
 - Bart-SMHI and HOMER-auto has the lowest Centered Root Mean Square Error of the methods
 - Bart-SMHI has the lowest deviation of the network wide trend

Optimisation of Bart-parameters, motivation

- The development of the Bart script has enabled the application of tailored acceptance criteria to maximize/minimize target features of the homogenised data, e.g.:
 - Maximize HSS or PSS
 - Minimize the number of “false positives”
 - Maximize the detection of large breaks
 - Minimize the error at each data point
 - Minimize the error of the network wide trend
- The criteria can be developed by the use of optimisation of the Bart-parameters

Optimisation of Bart-parameters, method

- The workaround method
 - All reported breaks from one detection round for each method of raw data (except the ACMANT-style method, which need pre-homogenised data) are listed
 - Acceptance checks of all candidate breaks in all time series is performed outside Bart, with any value of the parameters
 - The accepted breaks are compared to the real breaks list and a skill score is computed (e.g. HSS, Heidke skill score)
 - Several cases can be examined with the same parameter set (e.g. Slovenia- T_{\min} , Slovenia- T_{\max} , Sweden- T_{\min} , Sweden- T_{\max}). A representative score of the cases (e.g. the minimum score) is assigned the parameter set.
 - The parameters are optimised according to some protocol
 - Drawbacks
 - Only break-related scores can be applied, not e.g. root-mean-square-error
 - Changes in scores due to the ACMANT-style method's dependence on pre-homogenisation are not captured
 - Changes in scores due to subsequent homogenisation on pre-homogenised data are not captured
- The full method
 - A run of Bart (where detection of breaks in the raw data are pre-run) with a certain values of the parameters
 - A skill score is computed
 - Drawback
 - Slow and expensive

Optimisation of Bart-parameters, results

- We hope to soon publish results of optimised detected homogeneity break acceptance criteria for a set of targets

References

- Mestre, Olivier & Domonkos, Peter & Picard, Franck & Auer, Ingeborg & Robin, Stéphane & Lebarbier, Emilie & Boehm, Reinhard & Aguilar, Enric & Guijarro, Jose & Vertachnik, Gregor & Klančar, Matija & Dubuisson, Brigitte & Stepanek, Petr. (2013). HOMER: A homogenization software - methods and applications. *Idojaras*. 117.
- Kivinen, S.; Rasmus, S.; Jylhä, K.; Laapas, M. Long-Term Climate Trends and Extreme Events in Northern Fennoscandia (1914–2013). *Climate* 2017, 5, 16.
- Morán-Tejeda, E., Bazo, J., López-Moreno, J.I., Aguilar, E., Azorín-Molina, C., Sanchez-Lorenzo, A., Martínez, R., Nieto, J.J., Mejía, R., Martín-Hernández, N. and Vicente-Serrano, S.M. (2016), Climate trends and variability in Ecuador (1966–2011). *Int. J. Climatol.*, 36: 3839-3855. doi:10.1002/joc.4597
- Vicente-Serrano, S.M., Aguilar, E., Martínez, R. et al. The complex influence of ENSO on droughts in Ecuador. *Clim Dyn* 48, 405–427 (2017). <https://doi.org/10.1007/s00382-016-3082-y>
- Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. and Rosas, G. (2017), The influence of station density on climate data homogenization. *Int. J. Climatol*, 37: 4670-4683. doi:10.1002/joc.5114

References, cont.

- SMHI, Hur beräknas medeltemperatur?, 2014, <https://www.smhi.se/kunskapsbanken/meteorologi/hur-beraknas-medeltemperatur-1.3923>, viewed 8 april 2020
- Guijarro, José A. "User's guide to climatol." An R contributed package for homogenization of climatological series. State Meteorological Agency (AEMET), Balearic Islands Office, Palma de Mallorca, Spain (2011).
- 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.
- Indecis Project, Homogenisation Benchmark, 2017, <http://www.indecis.eu/benchmarking.php>, viewed 8 april 2020