

Magnus Joelsson

The Birth of Bart: Automation and Evaluation of the Homogenisation tool HOMER

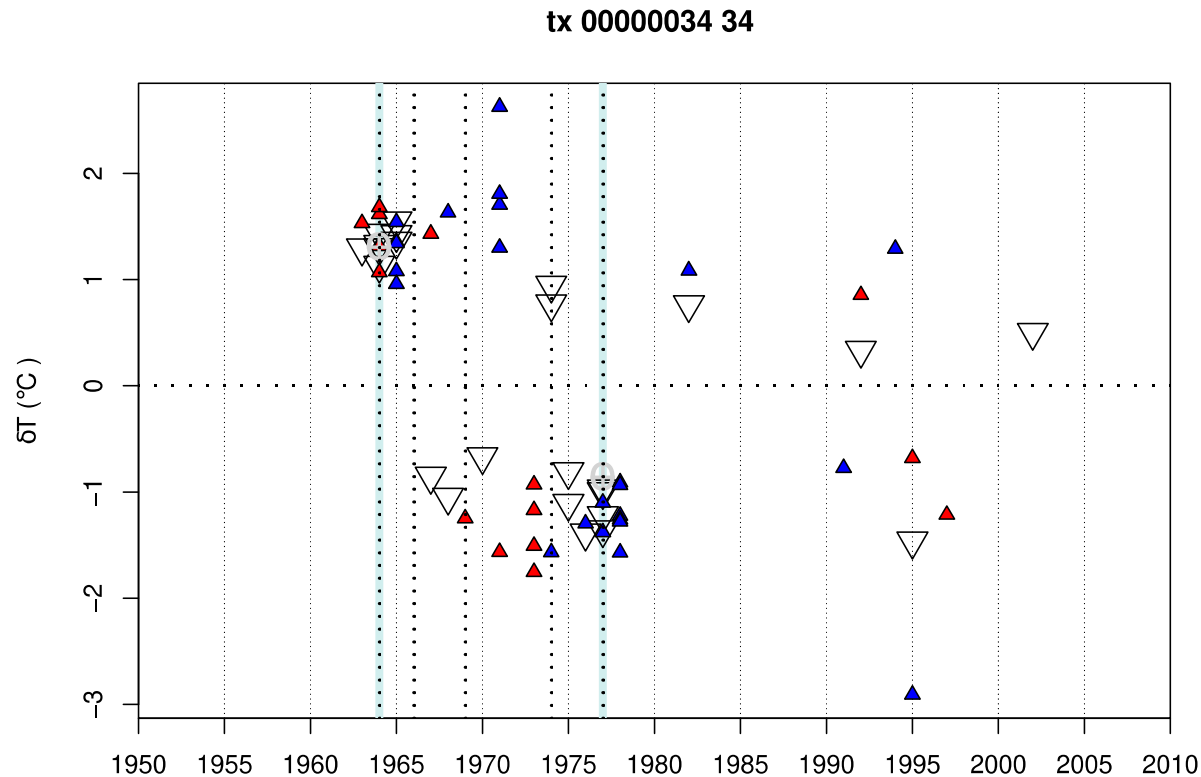
Motivation

- Homogenisation of monthly temperature series
- New method:
 - Climatol
 - HOMER
- Automation of HOMER
 - Objective testing
 - Operational use

HOMER

- Three tasks:
 - Detect homogeneity breaks, hybrid method
 - Pairwise (annual and seasonal) adapted from PRODIGE (d)
 - "Joint", method adapted from genome research (j)
 - Seasonal cycle, adapted from ACMANT (a)
 - Attribution of specific month, adapted from ACMANT (m)
 - Interpolate data gaps (g)
 - Correct homogeneity breaks (c)
- Procedure: d-j-(g)-c-a-c-d-j-c-a-m-c
- Two modes:
 - Automatic (well, semi-automatic)
 - Interactive

Example: HOMER detection of homogeneity breaks



- 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

The “traditional” acceptance criteria

- Reported breaks from each method are counted
- Seasonal method counts only as 1/3 break.
- Sum of reported breaks \geq four \rightarrow break is accepted
- Sum of reported breaks \geq three \rightarrow break is noted.
- If meta data or a previous homogenisation round supports a noted break \rightarrow 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 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
- Restricted time: 1 year (thus far only 1 and 0 is tested)

Optional parameters

- Normalisation of the number of references for the pairwise-detection
- Amplitude threshold, default = 0
- Amplitude power weight:

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

Default = 0

- Correlation power weight:

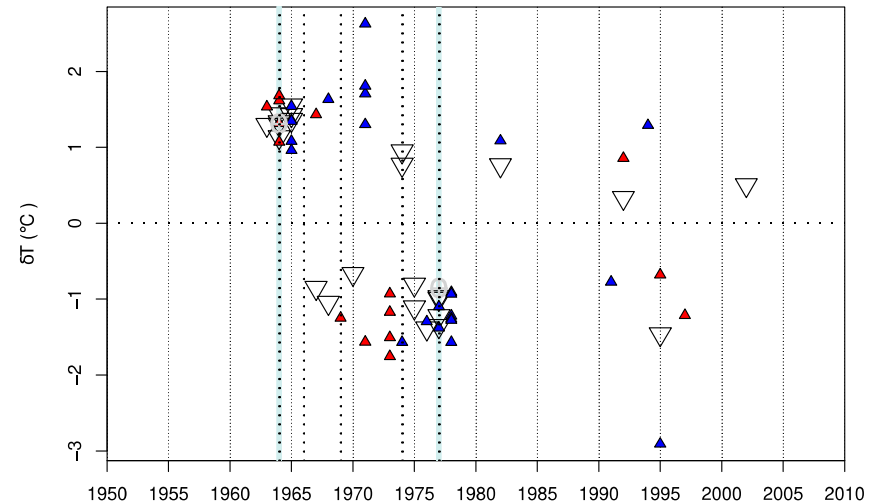
points = correlation reference vs. candidate^{correlation power}

Default = 0

Point system, example with default values

Year: 1964 with the score 34.00

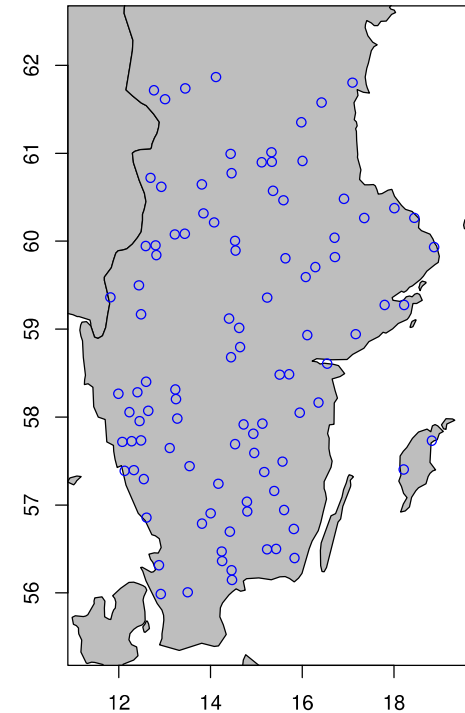
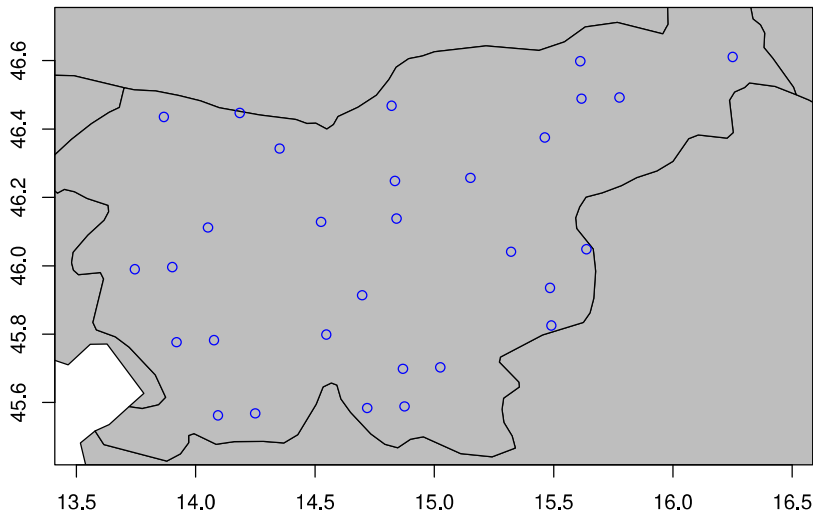
- The year (1964)
 Joint: 1, Acmant: 1, pair (annual, positive): 4, pair (annual, negative): 0, pair (winter, positive): 0, pair (winter, negative): 0, pair (summer, positive): 4, pair (summer, negative): 0
 points = $3 + 3 + 4 \times 3 + 11 = 22.00$
- The previous year (1963)
 Joint: 0, Acmant: 0, pair (annual, positive): 1, pair (annual, negative): 0, pair (winter, positive): 0, pair (winter, negative): 0, pair (summer, positive): 1, pair (summer, negative): 0
 points = $3 + 1 = 4$
 Will not be counted to next previous year (with 0 points)
- The following year (1965)
 Joint: 0, Acmant: 0, pair (annual, positive): 3, pair (annual, negative): 0, pair (winter, positive): 5, pair (winter, negative): 0, pair (summer, positive): 0, pair (summer, negative): 0
 points = $3 \times 3 + 5 = 14$
 Will not be counted to next following year (with 3 points)
- Five-year points = [0; 4; 22; 14; 3]
 Points = $(4-3) + 22 + (14-3) = 34 \geq 12$



- Point system:**

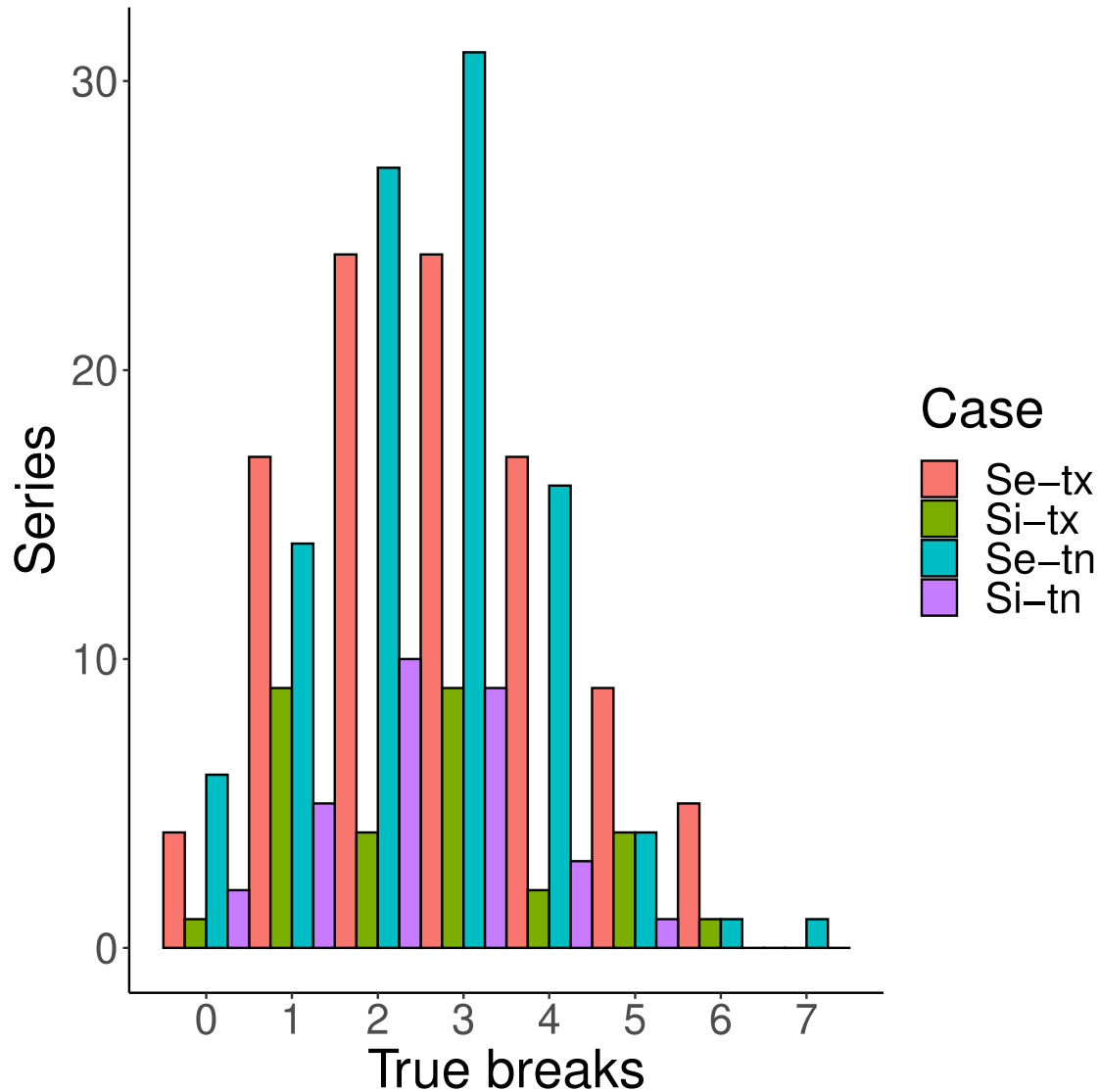
 - Pairwise-detection, annual (large triangles): 3 points
 - Pairwise-detection, season (small triangles): 1 point
 - Joint-detection (cross-hairs): 3 point
 - ACMANT-style detection (dotted vertical lines): 3 points
 - Accepted break (cyan vertical line): ≥ 12 points
 - No meta data, normalisation of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

Indecis benchmark data sets



- Sweden (100 stations), Slovenian networks (30 stations)
- Daily minimum and maximum temperatures, averaged to monthly values
- 1950-2005
- Data gaps and homogeneity breaks (between 69 and 280) inserted

Distribution of inserted ("true") homogeneity breaks



Metrics

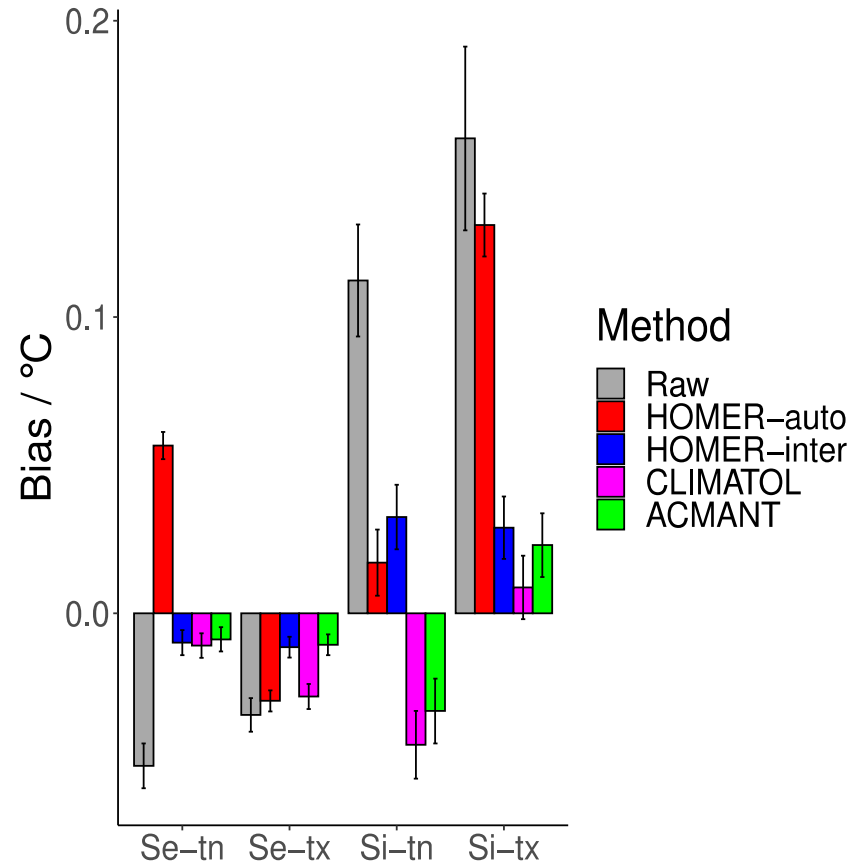
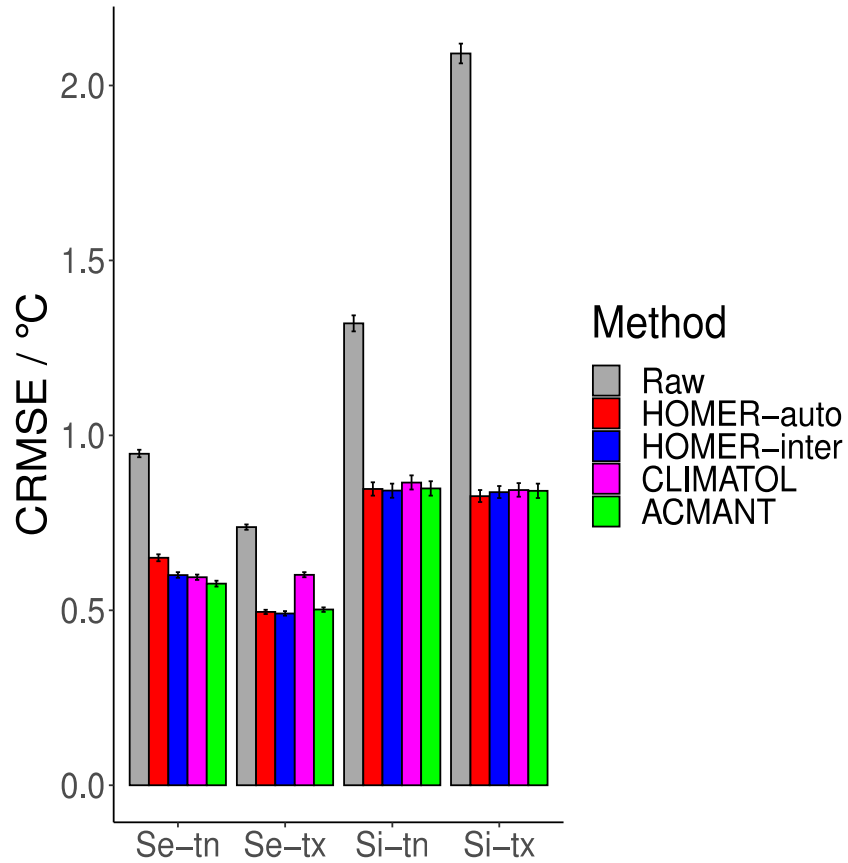
- Reproduction of time series:
 - Centered Root Mean Square Error (CRMSE)
$$\text{CRMSE} = \sqrt{1/n \sum_{i=1:n} ((x_i - \bar{x}) - (y_i - \bar{y}))^2}$$
 - Bias
$$B = 1/n \sum_{i=1:n} (x_i - y_i)$$
 - Error reduction = $\text{RMSE}_{\text{raw}} - \text{RMSE}_{\text{homogen}}$
 - Efficiency
$$E = \text{Error reduction} / \text{RMSE}_{\text{raw}} \times 100$$
- Detection of homogeneity break points:
 - Probability of Detection (POD)
POD = number of years with detected breaks / number of years with breaks
 - Probability of Hit (POH)
POH = number of years with detected breaks / number of years with reported breaks

Part 1: Comparison of Methods

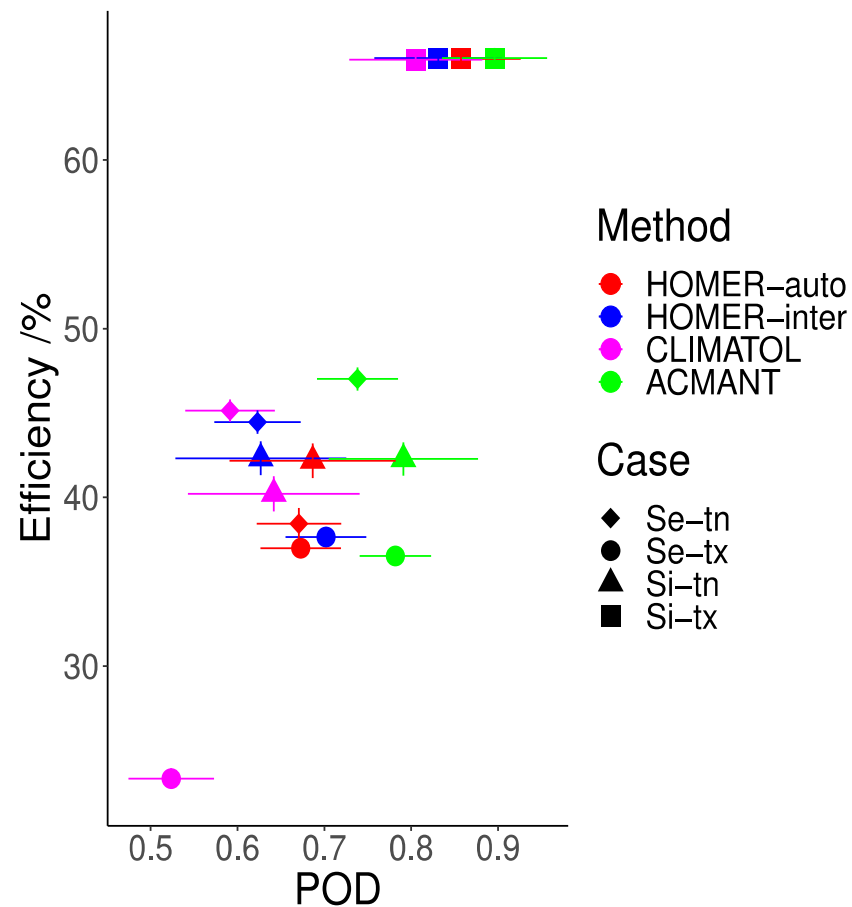
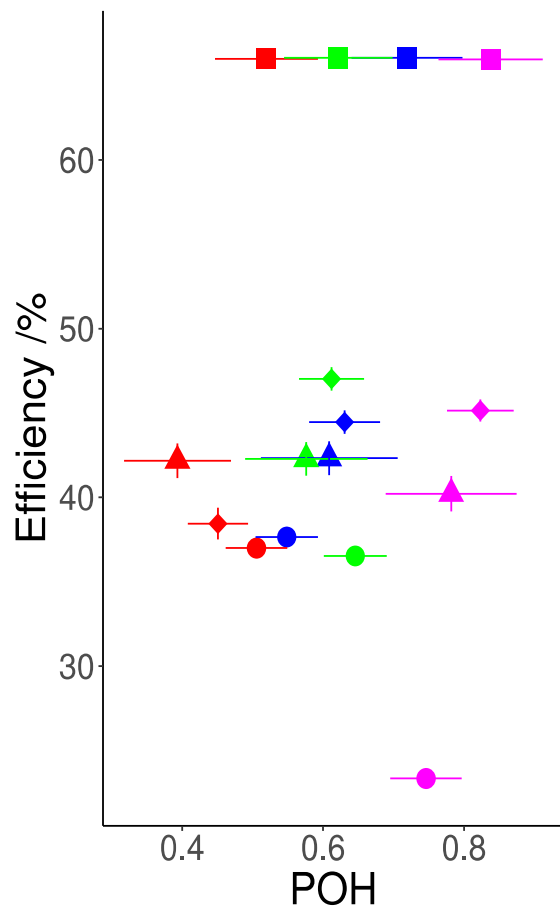
Four homogenisation methods are compared:

- Bart/HOMER with two setups of parameters:
 - The established acceptance criteria used with HOMER's interactive mode (HOMER-inter)
 - The acceptance criteria of HOMER's automatic mode (HOMER-auto)
- Climatol (default settings)
- ACMANT

Results: Reproduction of time series



Results: Detection of homogeneity break points



Discussion

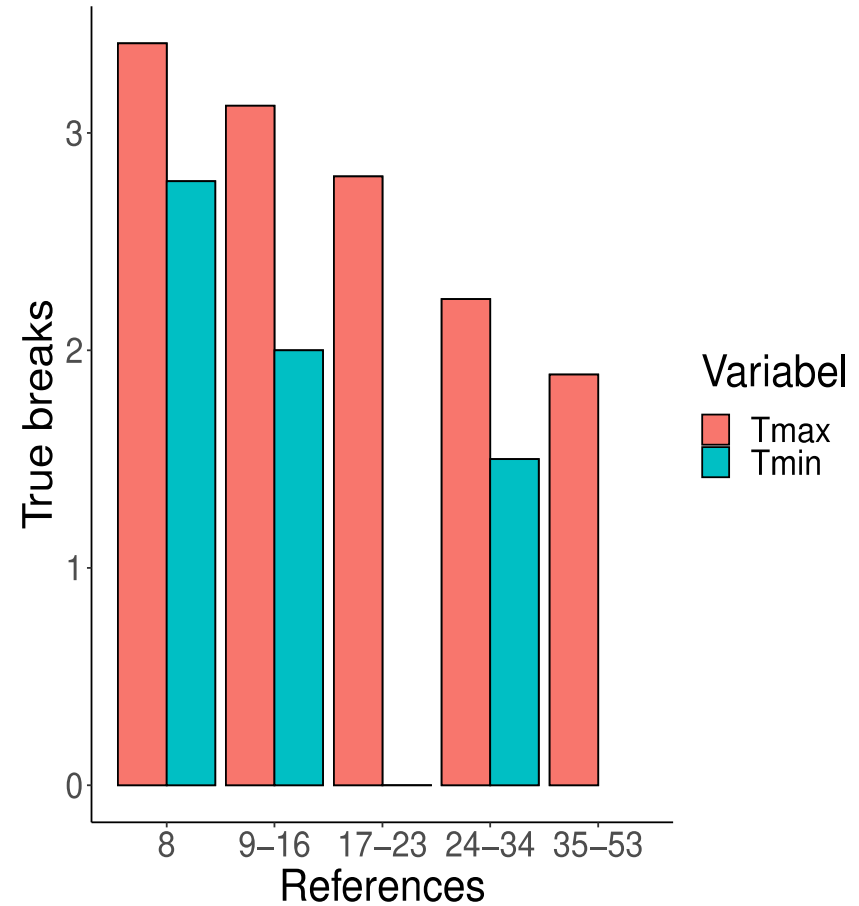
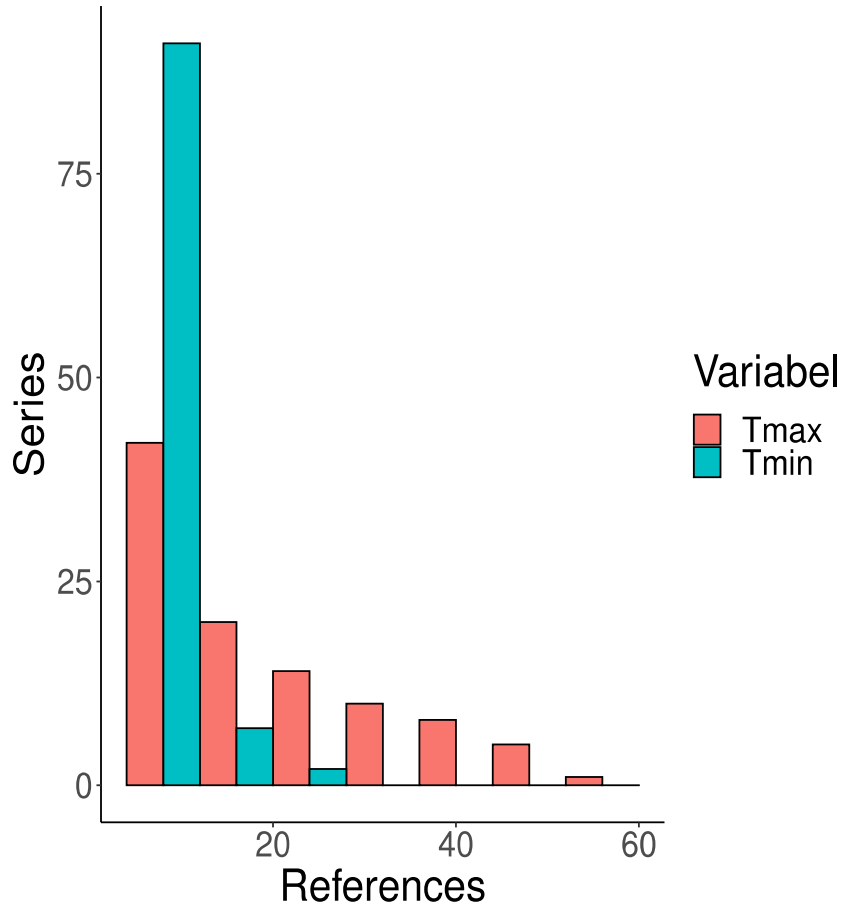
- The usefulness of synthetic benchmark data sets for the evaluation of homogenisation methods is an open question
- The two benchmark methods (ACMANT and Climatol) represents two approaches to homogenisation:
 - ACMANT reports more breaks which improves the data quality but introduces more false positives
 - Climatol reports fewer breaks which leaves a larger residual error but introduces fewer false positives
 - HOMER-inter places itself between ACMANT and Climatol in both POH and POD
 - HOMER-auto reports approximately as many breaks as ACMANT, but improved the data quality less. Additionally, it occasionally has a large bias.
- When dealing with observational data without any possible validation, it can be argued that the correction of uncertain break points should be avoided

Part 2: Selection of reference series

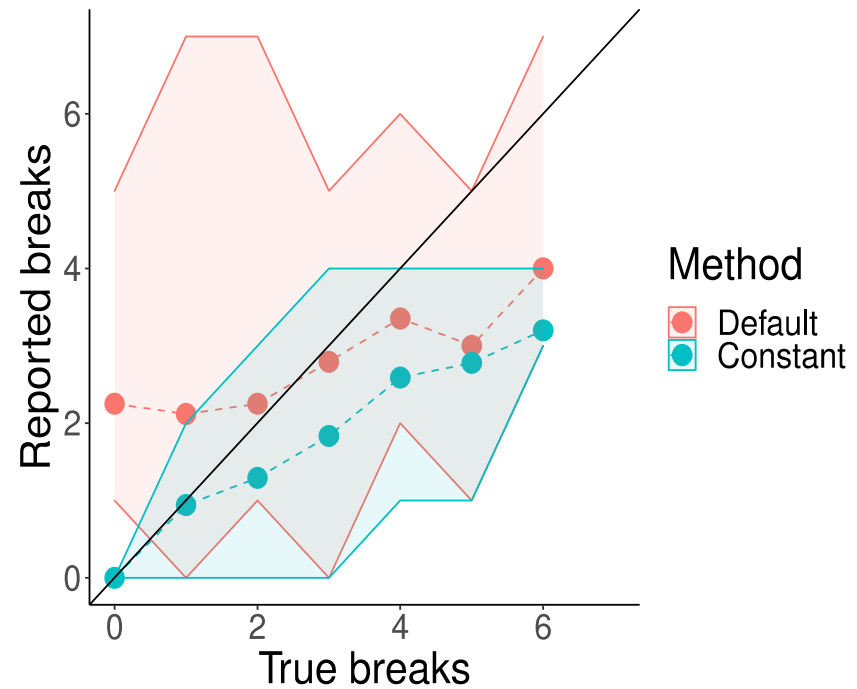
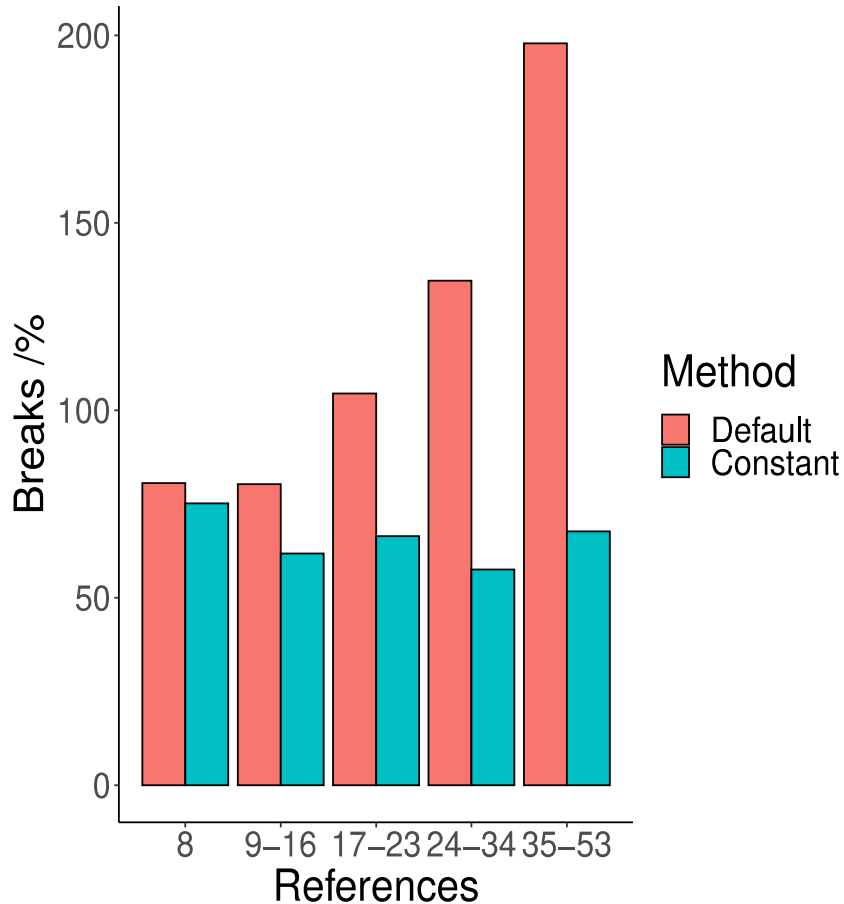
Two selection criteria are compared for homogenisation with HOMER-inter, Swedish network:

- Defaults: Correlation threshold 0.95, minimum 8 references
- Constant: 8 most highly correlated time series

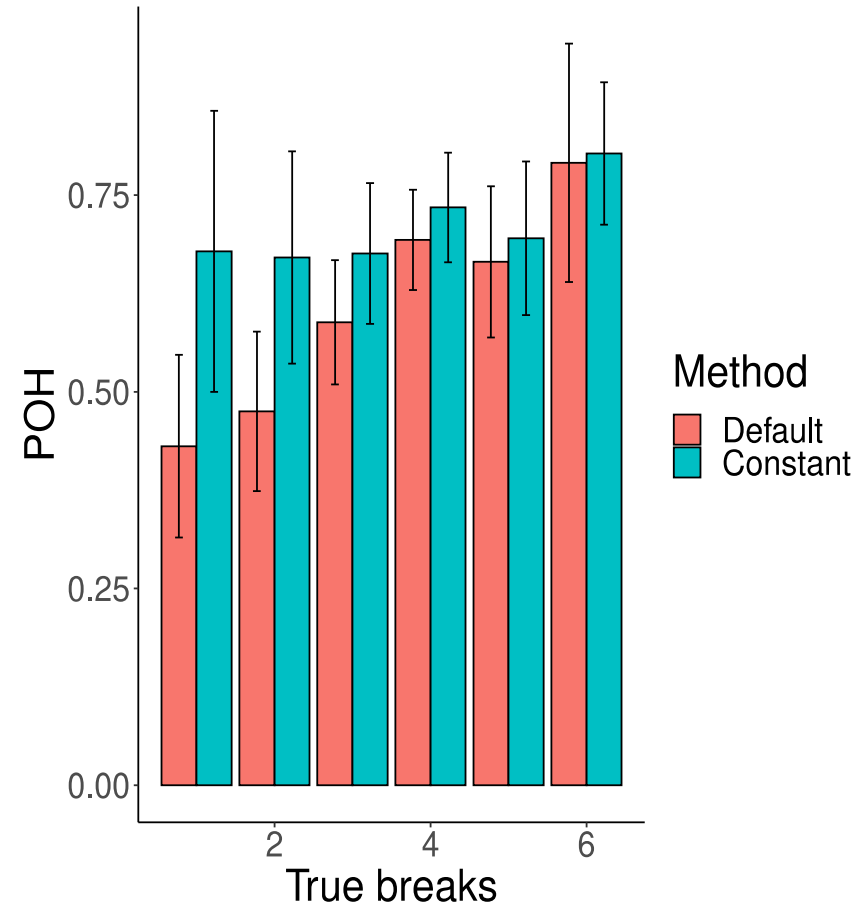
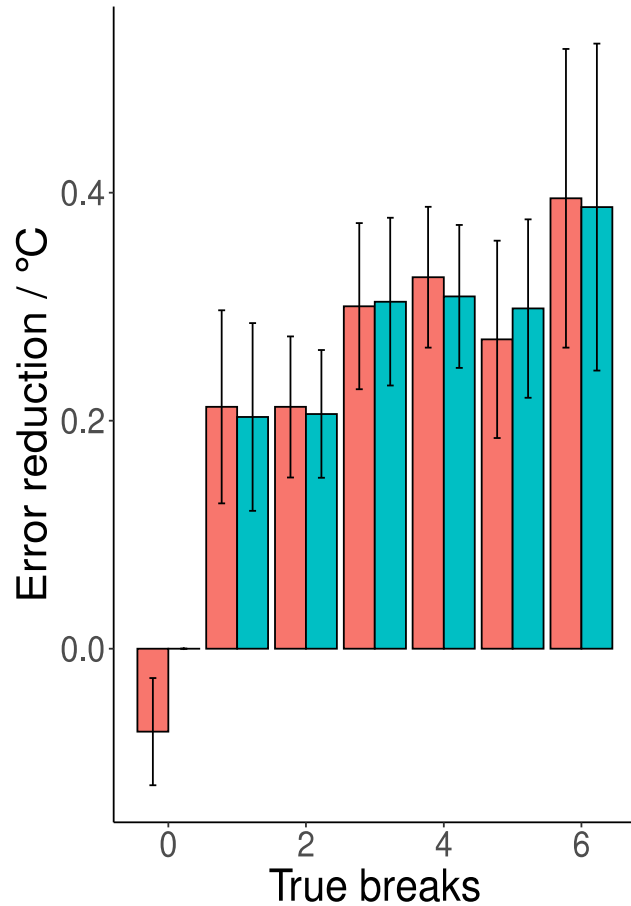
Results: Number of references – true breaks



Results: True breaks - reported breaks (daily maximum temperature)



Results: True breaks - performance



Discussion

- In the default mode the number of references can vary substantially
- There seems to be a connection between the number of references and the number of true breaks. Probably relatively clean series correlate better with each other than more perturbed series.
- For varying number of references there is a risk of overhomogenise relatively clean series
- For constant number of references the POH is less sensitive to the number of true breaks in the data
- Constant number of references does not seem to discard vital information, but might be a problem if the coverage varies in time
- Normalisation of the number of references in the pairwise detection could be used to circumvent this problem and still use all well correlated references. However, there is instead a risk of underhomogenise series with large number of references

Conclusions

- HOMER is automated by the use of parameters (Bart)
- All the methods reduces the average error
- The performance of HOMER-inter compare well with Climatol and ACMANT
 - The average error is within the uncertainty range of the average error of ACMANT and slightly better than Climatol
 - The distribution of error is similiar to those of Climatol and ACMANT
 - 70 % of the homogeneity breaks are found, within the uncertainty range to Climatol, less than ACMANT
 - 70 % of the reported breaks are true breaks, more than ACMANT, less than Climatol
- HOMER-auto can not be recommended
- Relatively clean series seem to correlate strongly with each other which has implications for reference selection