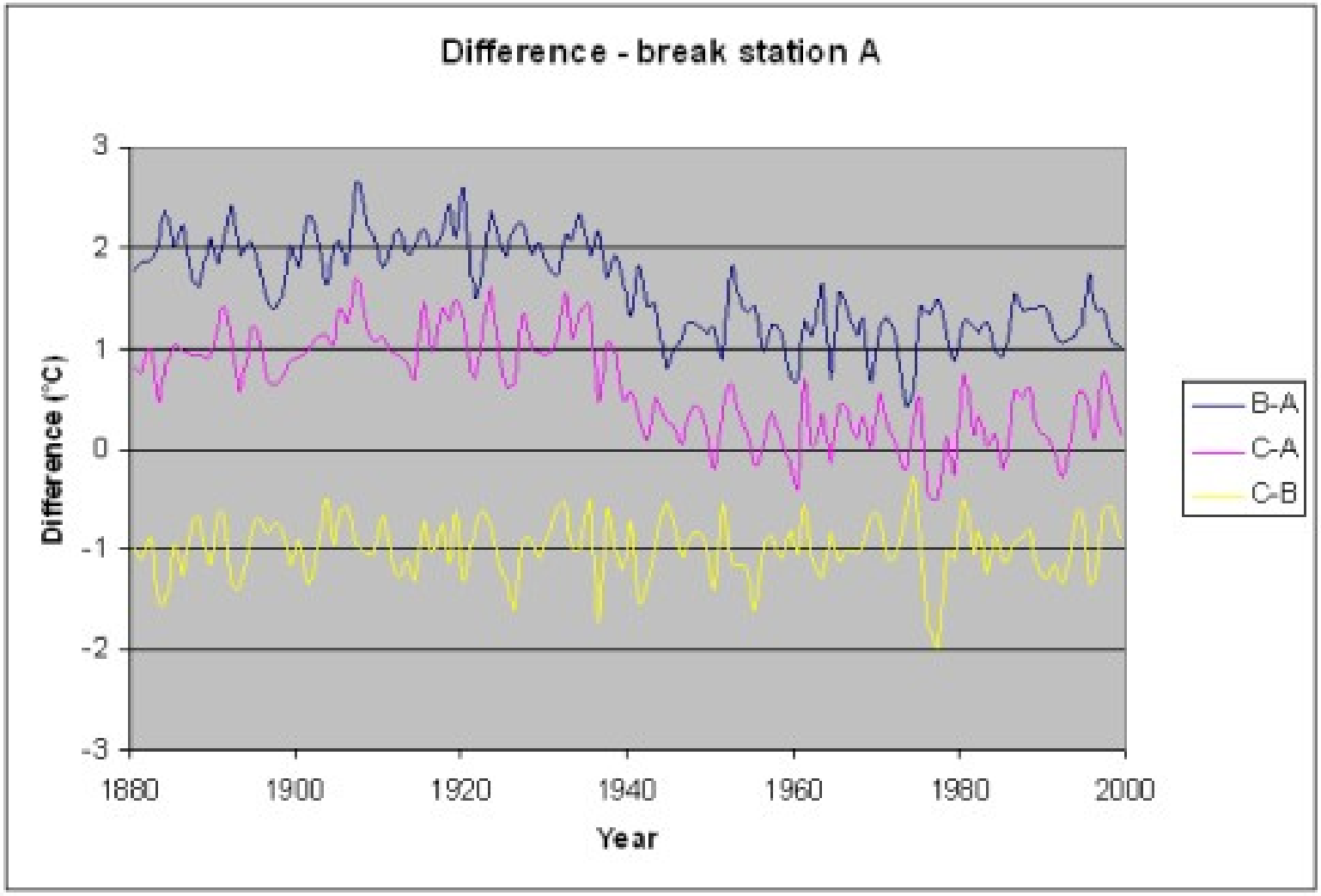
Relative statistical homogenization of observational networks with a low signal to noise ratio

Victor Venema & Ralf Lindau

Relative homogenisation - reference series

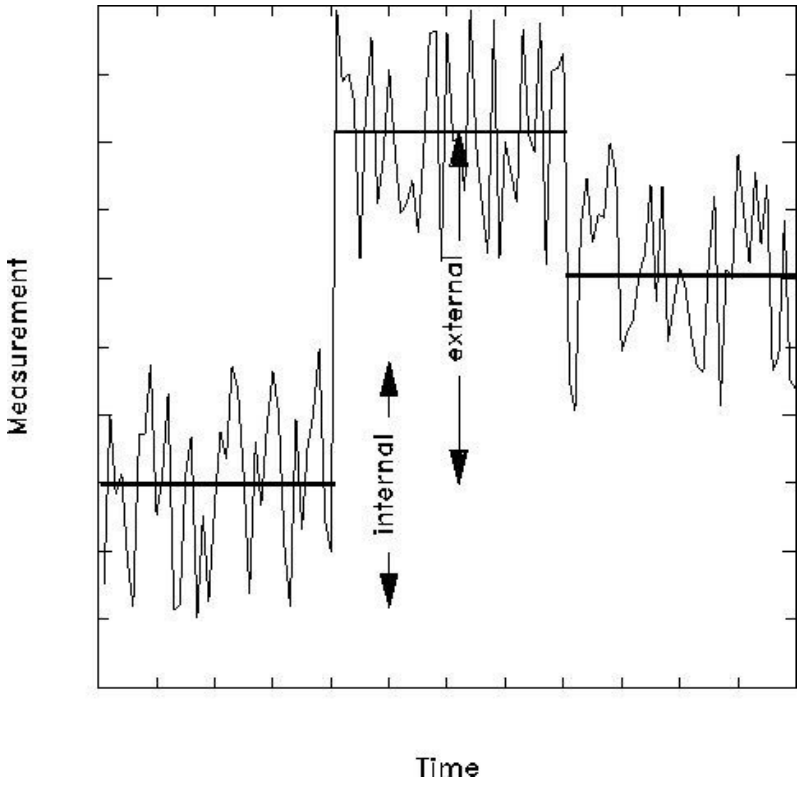
Pairwise homogenization



http://variable-variability.blogspot.de/2012/08/statistical-homogenisation-for-dummies.html

Signal to Noise Ratio (SNR)

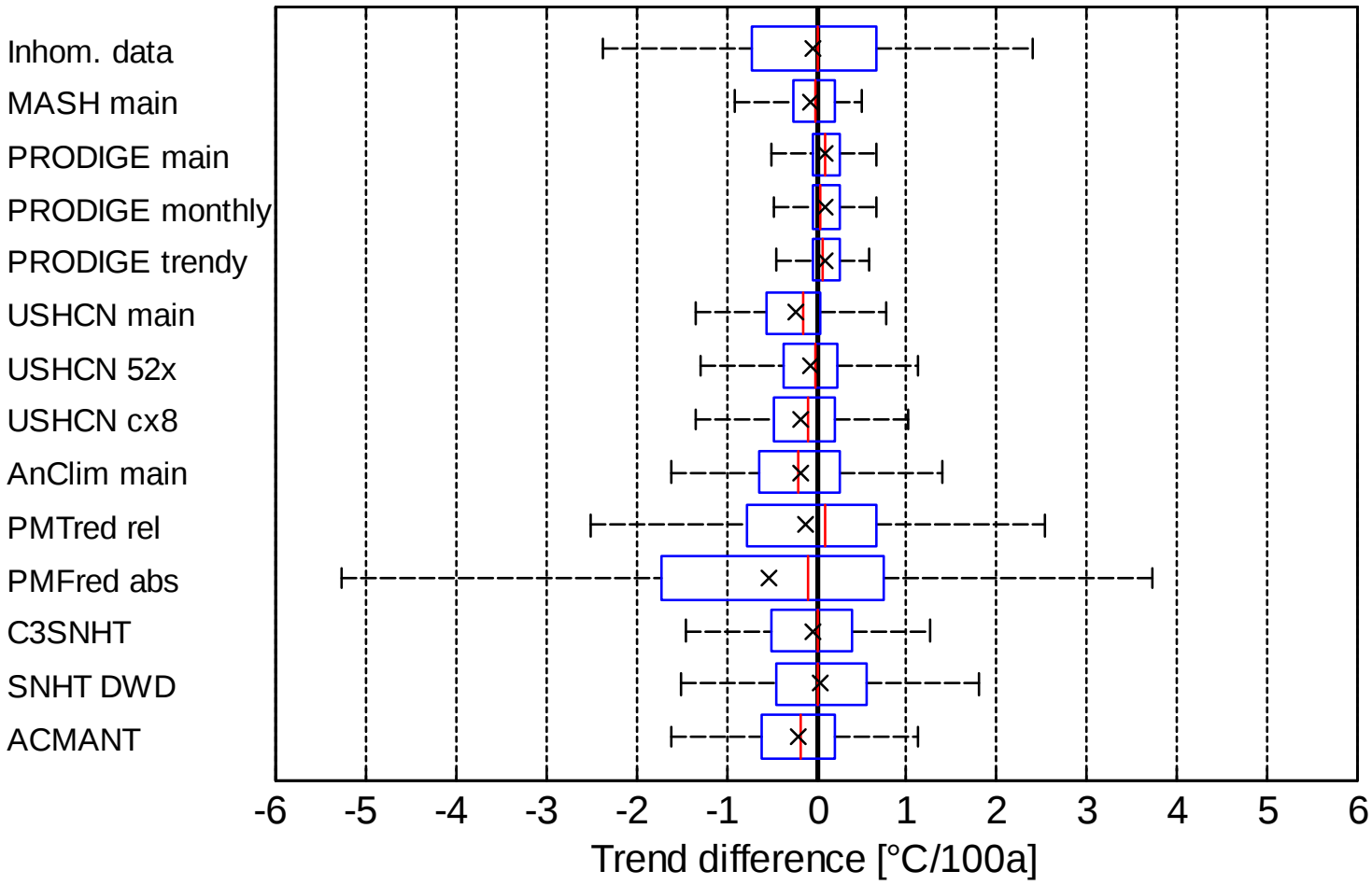
- Noise: variability within the subperiods
  - Depends on station density, climate & measurement quality
- Break variance: between the means of different subperiods



All is fine in 2012 - Benchmarking

Venema, V., O. Mestre, E. Aguilar, I. Auer, J.A. Guijarro, P. Domonkos, G. Vertacnik, T. Szentimrey, et al. Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, **8**, pp. 89-115, doi: 10.5194/cp-8-89-2012, 2012.

Errors in station trends



Main caveats HOME benchmarking

1. Size breaks (SNR)
   - Lindau, R. and Venema, V. K. C.: The joint influence of break and noise variance on the break detection capability in time series homogenization, Adv. Stat. Clim. Meteorol. Oceanogr., 4, 1–18, https://doi.org/10.5194/ascmo-4-1-2018, 2018.
   - HOME: breaks 1.4 times too large
2. Station density (SNR) varies regionally
   - HOME: Europe station density (recent)
3. Network-wide non-climatic trend bias
4. Length of the series
   - Ralf Lindau and Victor Venema. On the multiple breakpoint problem and the number of significant breaks in homogenisation of climate records. *Idojaras*, 117, no. 1, pp. 1-34, 2013.

- Details: http://variable-variability.blogspot.com/2014/06/problems-with-home-benchmark.html

Which SNR is sufficient?

RMS skill for:

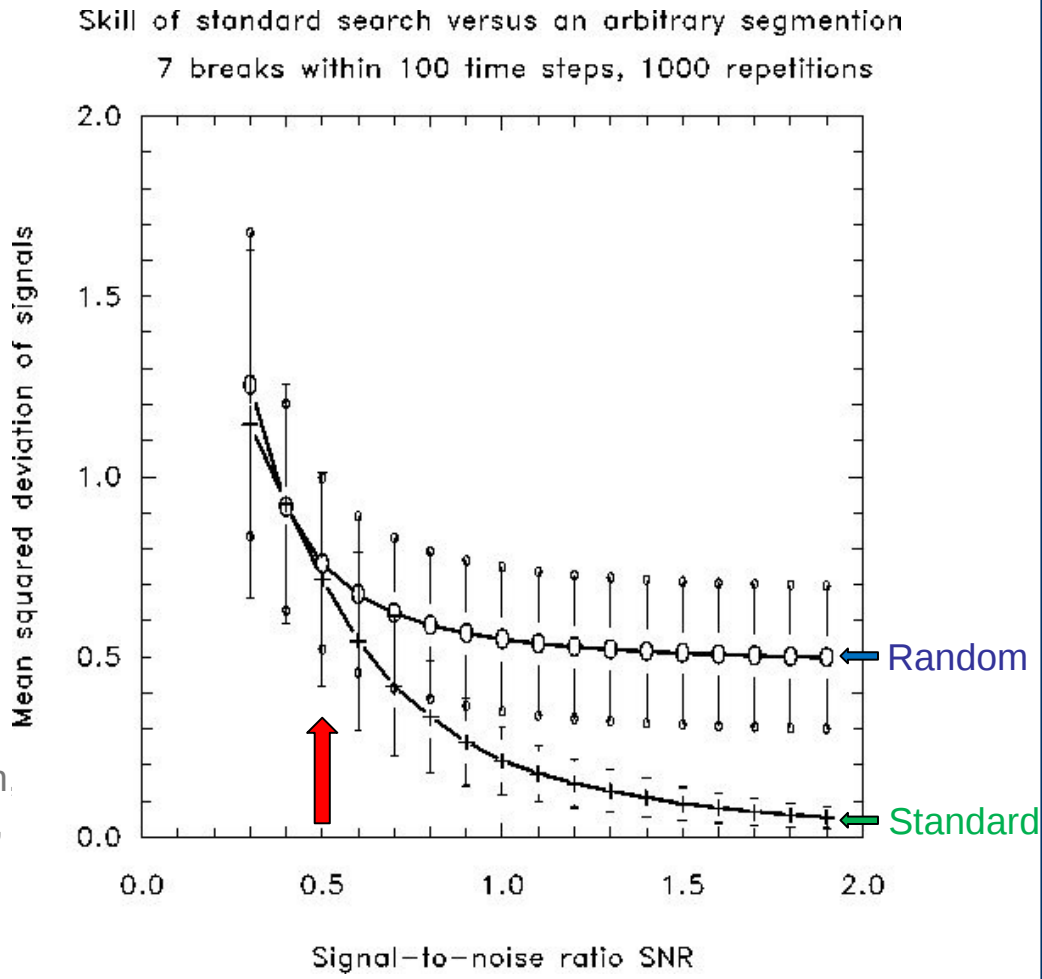
- 0 Random segmentation
- \+ Standard search

for different SNRs.

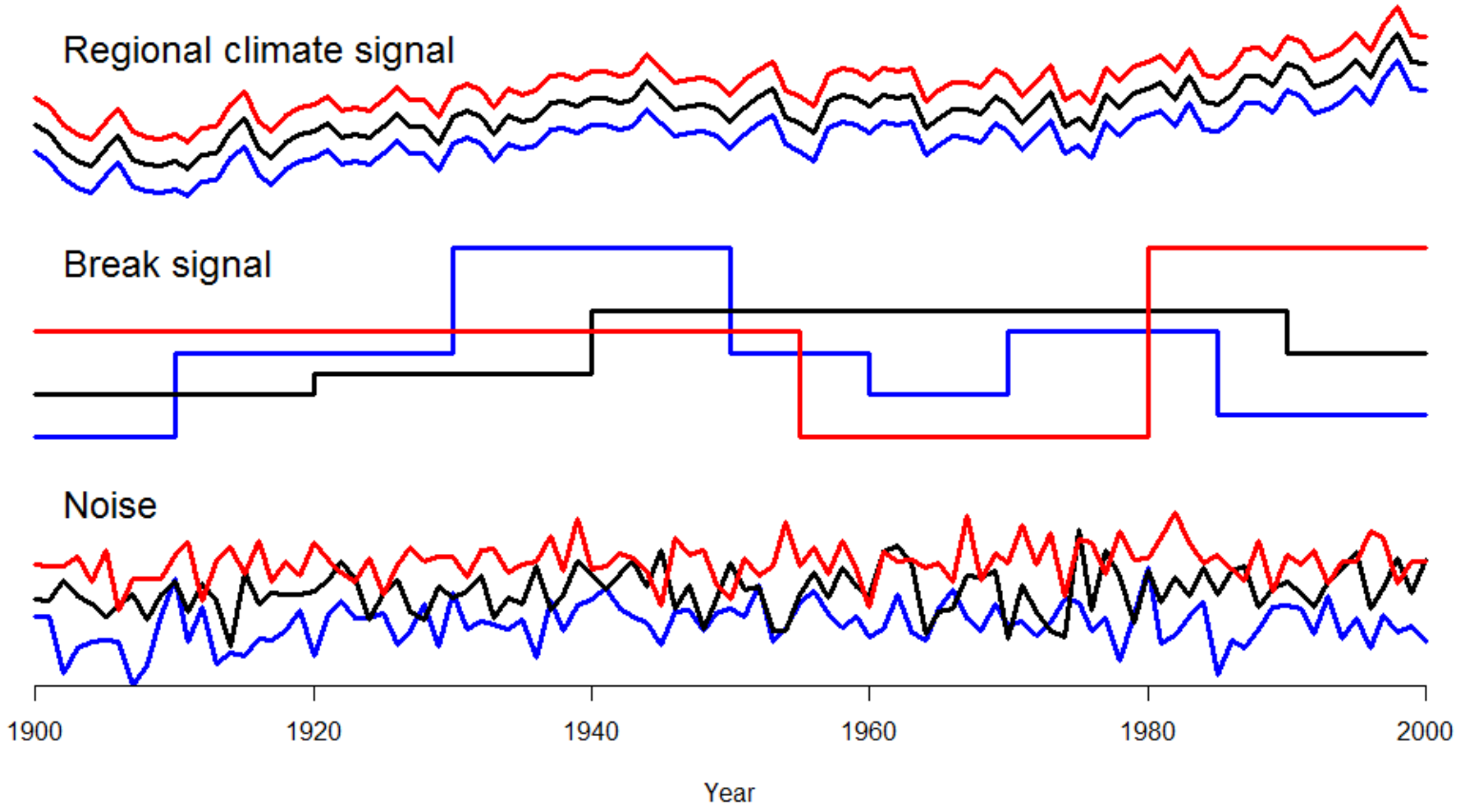
SNR = ½: Random segmentation and standard search have comparable skills.

Only for SNR > 1, the standard search is significantly better.

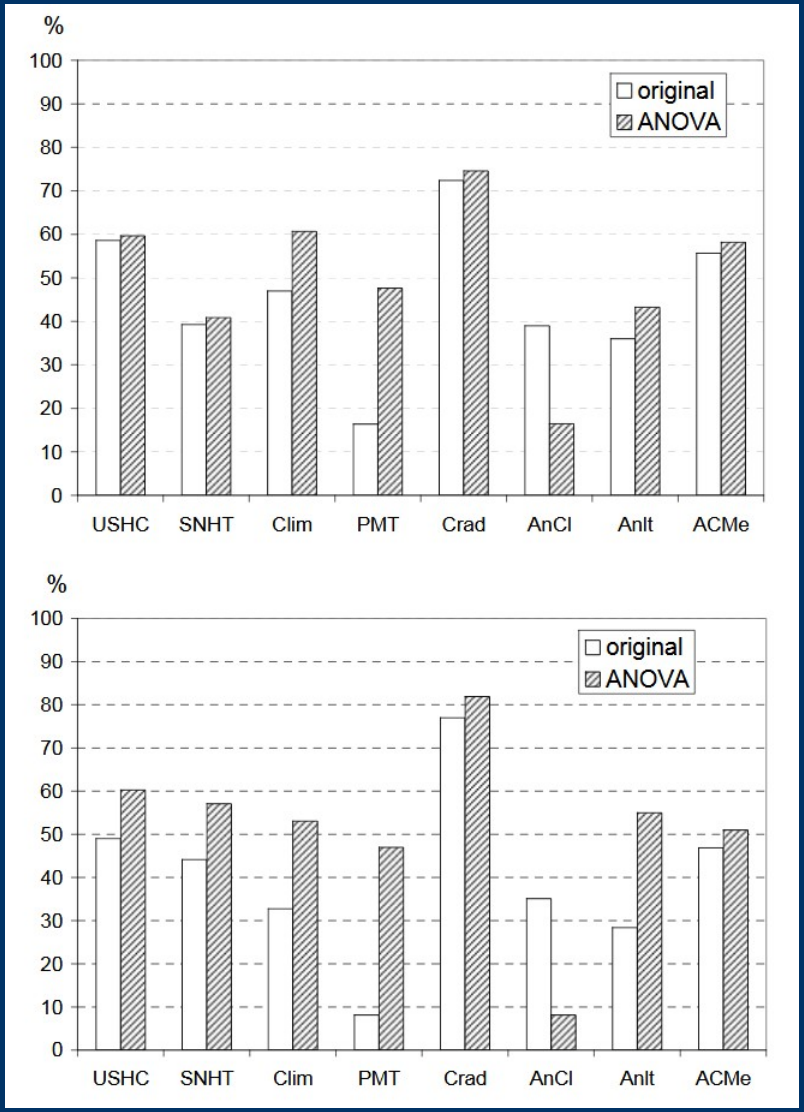
Lindau, R. and Venema, V. K. C.: The joint influence of break and noise variance on the break detection capability in time series homogenization, Adv. Stat. Clim. Meteorol. Oceanogr., 4, 1–18, https://doi.org/10.5194/ascmo-4-1-2018, 2018.



Correction by decomposition (“ANOVA”)



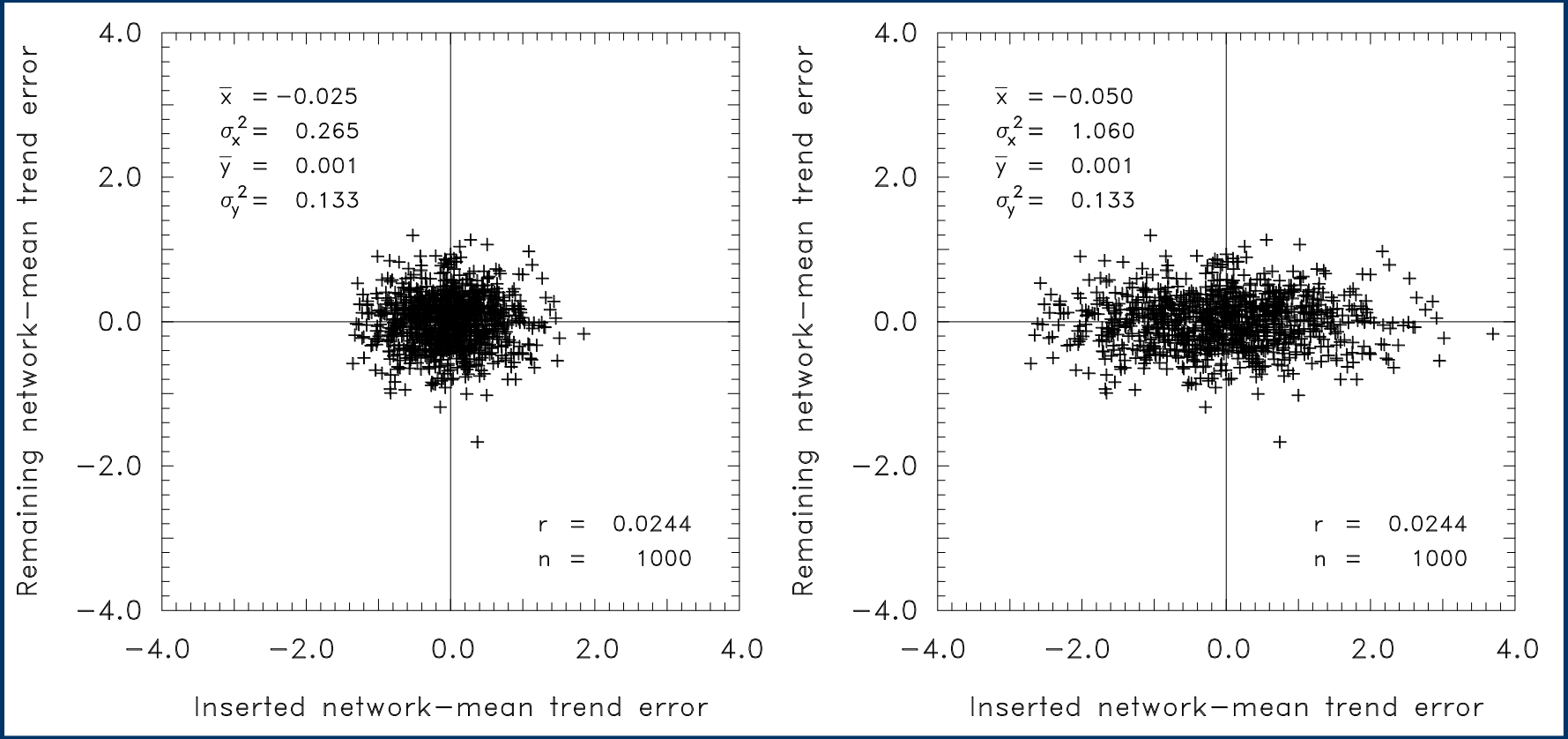
ANOVA corrections



- Top: CRMSE (annual)
- Bottom: Station trend errors
- All contributions, but AnClim, improved by applying ANOVA
  - Break file AnClim may have had errors

- *Domonkos, P., V. Venema, O. Mestre. Efficiencies of homogenisation methods: our present knowledge and its limitation. Proceedings of the Seventh seminar for homogenization and quality control in climatological databases, Budapest, Hungary, 24 – 28 October 2011, WMO report, Climate data and monitoring, WCDMP-No. 78, pp. 11-24, 2013.*

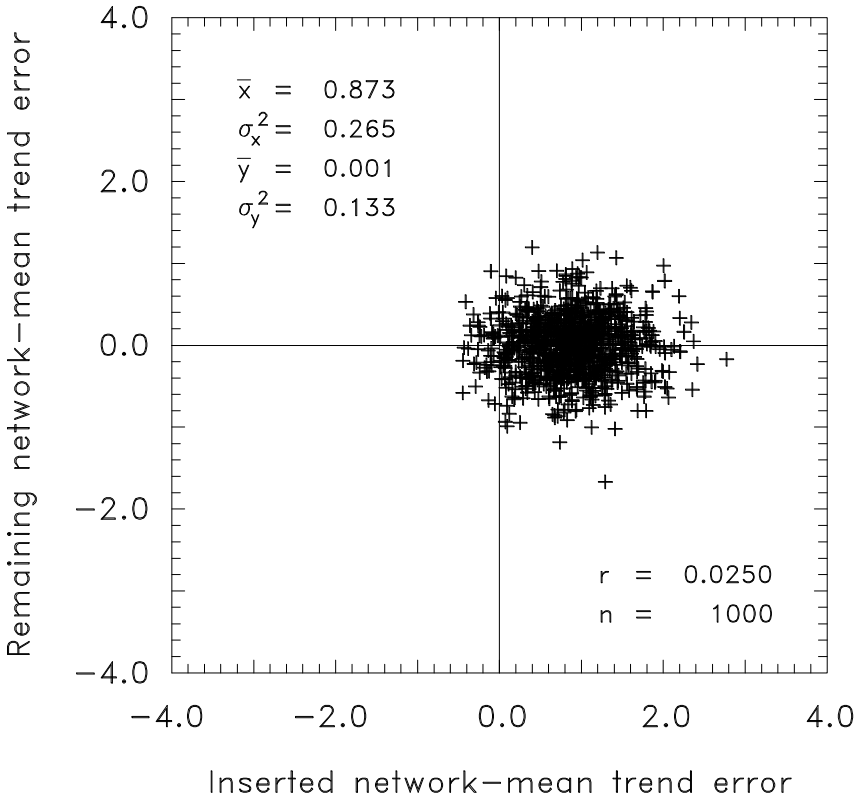
ANOVA – network trend – perfect breaks



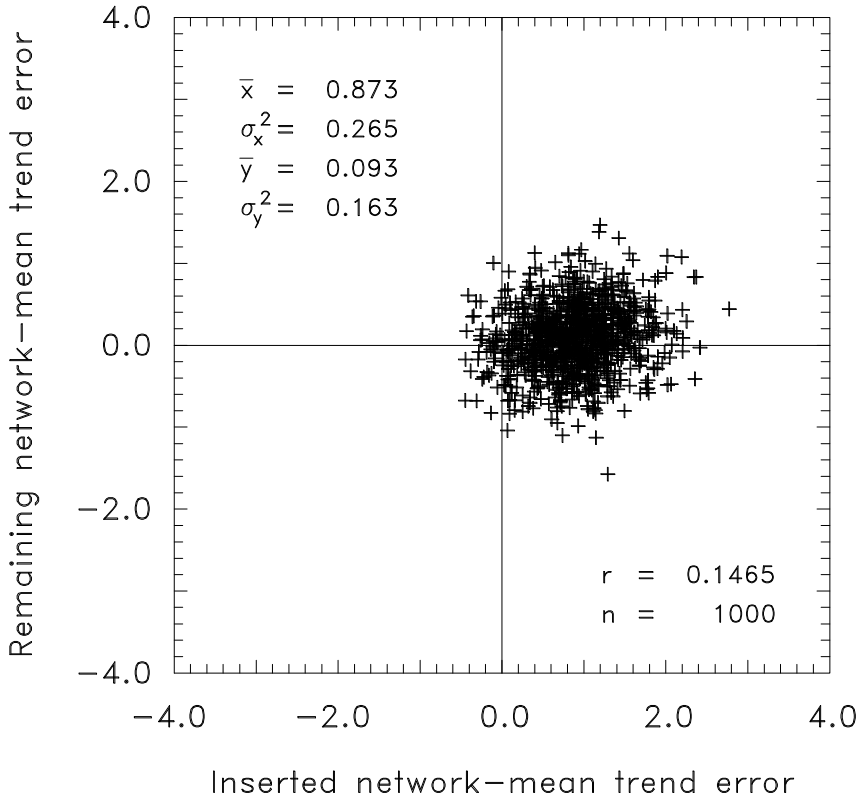
Lindau, R, Venema, V. On the reduction of trend errors by the ANOVA joint correction scheme used in homogenization of climate station records. Int J Climatol. 2018; 38: 5255– 5271. https://doi.org/10.1002/joc.5728

Network-wide trend bias

Perfect break positions



Break position error =1a

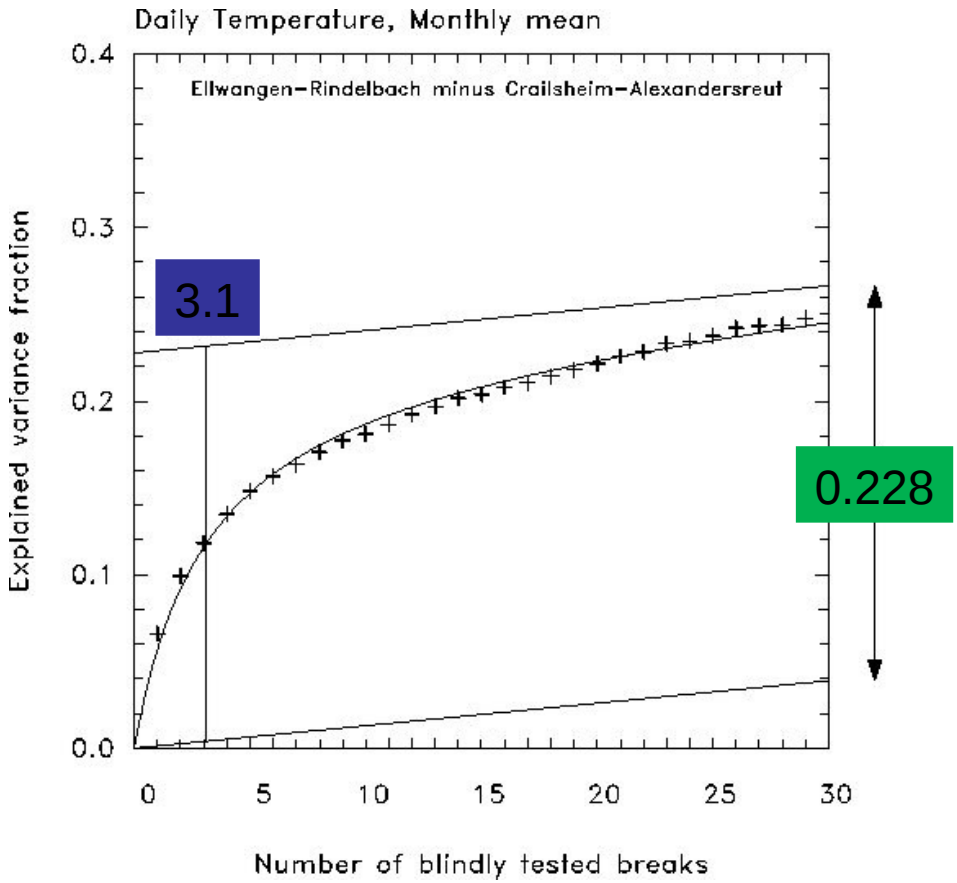


Swiss benchmark

- HOMER homogenised dense network is truth
- Sparse network (similar to Peru)
  - HOMER could reduce the errors in the station data, and the RMSE network-wide.
  - It could not reduce the network average trend errors.
  - A difficult task was a transition to AWS in 1980s.

- 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.
- https://homogenisation.grassroots.is/assessments/the-influence-of-station-density-on-climate-data-homogenization/

A priori formula

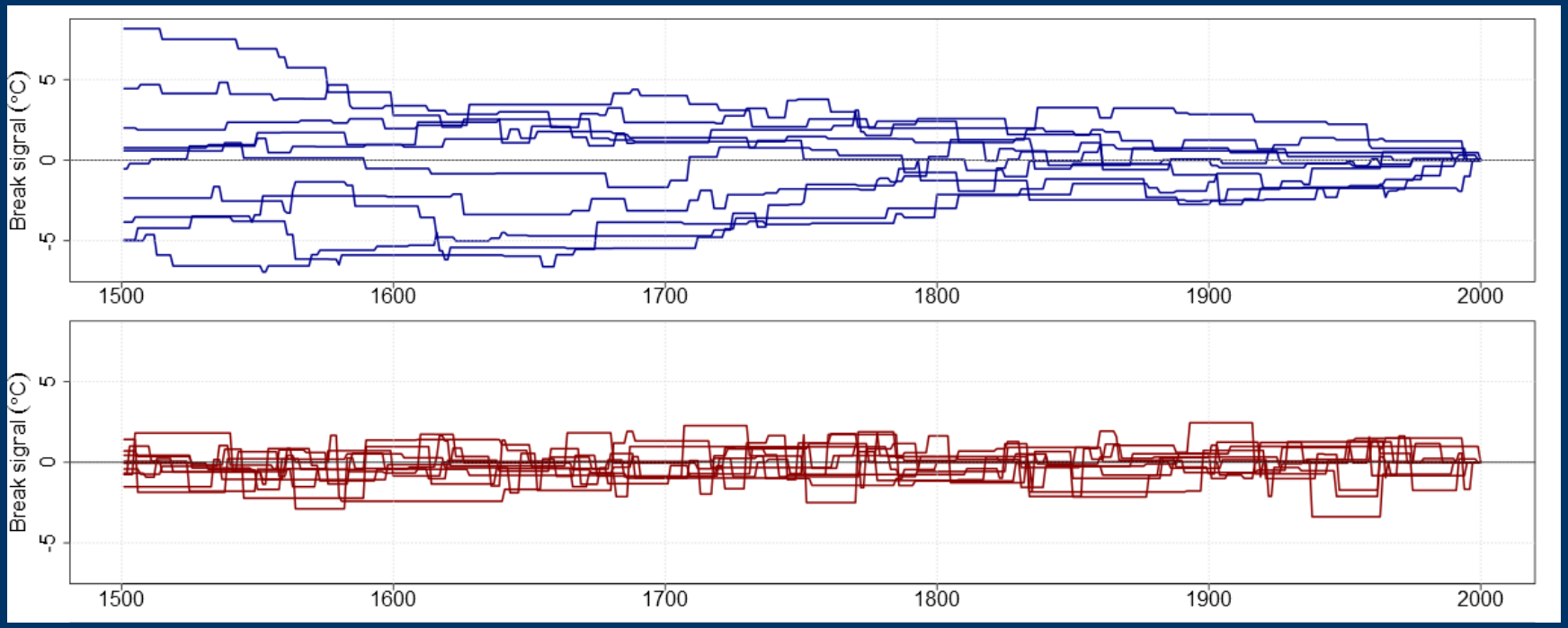


The different reaction of breaks and noise on randomly inserted breaks makes it possible to estimate break variance and break number *a priori*.

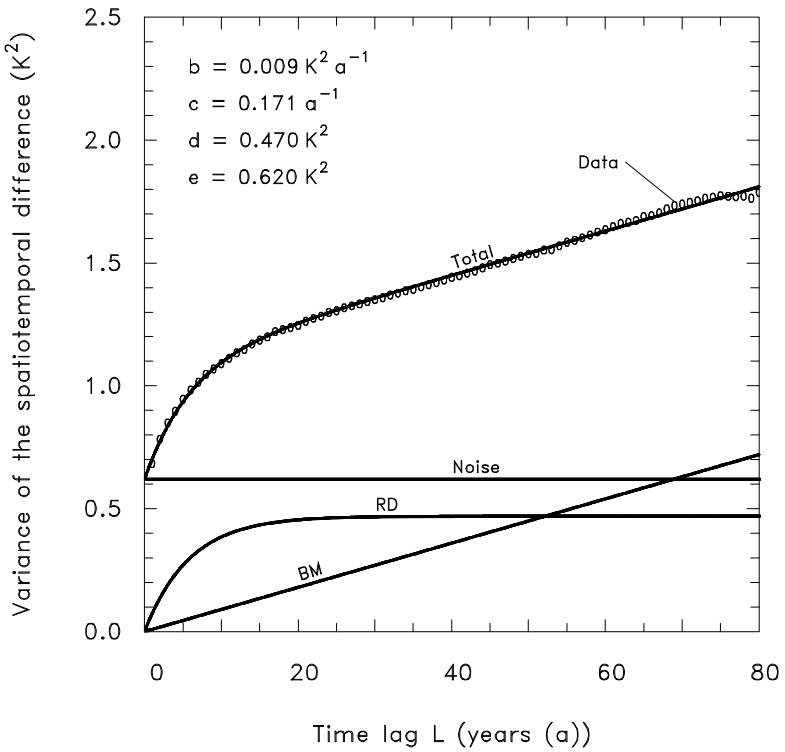
If we insert many breaks, almost the entire break variance is explained plus a known fraction of noise.

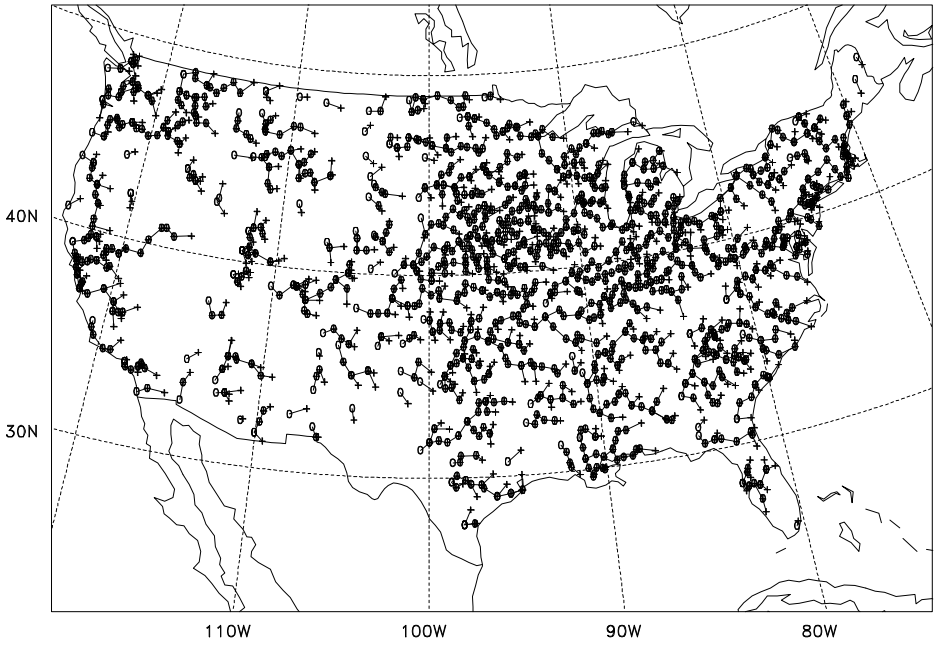
Lindau, R. and Venema, V. K. C.: The joint influence of break and noise variance on the break detection capability in time series homogenization, Adv. Stat. Clim. Meteorol. Oceanogr., 4, 1–18, https://doi.org/10.5194/ascmo-4-1-2018, 2018.

Noise, random walk or trendy noise?



Properties of inhomogeneities - USA

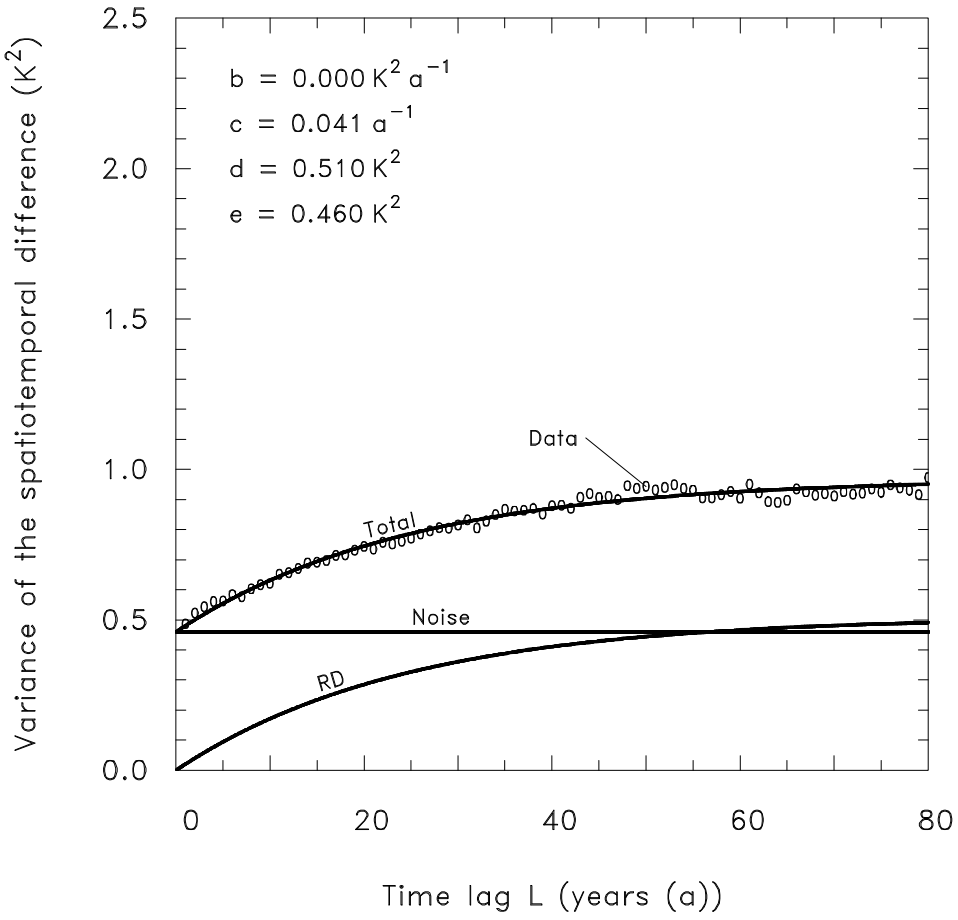


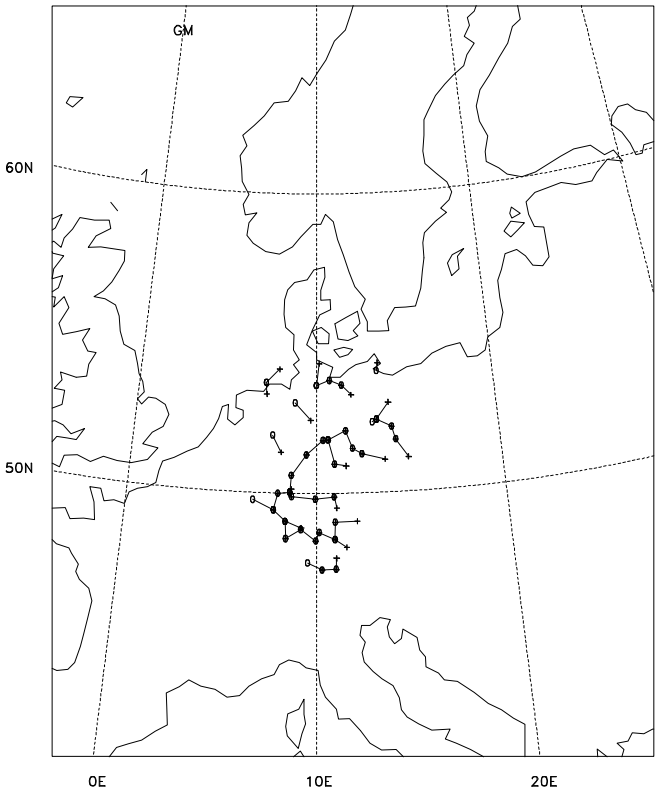


Station trend uncertainty: 0.82K/100a; empirical: 0.71K/100a

Lindau and Venema, Random trend errors in climate station data due to inhomogeneities (submitted to IJC)

Properties of inhomogeneities - Germany





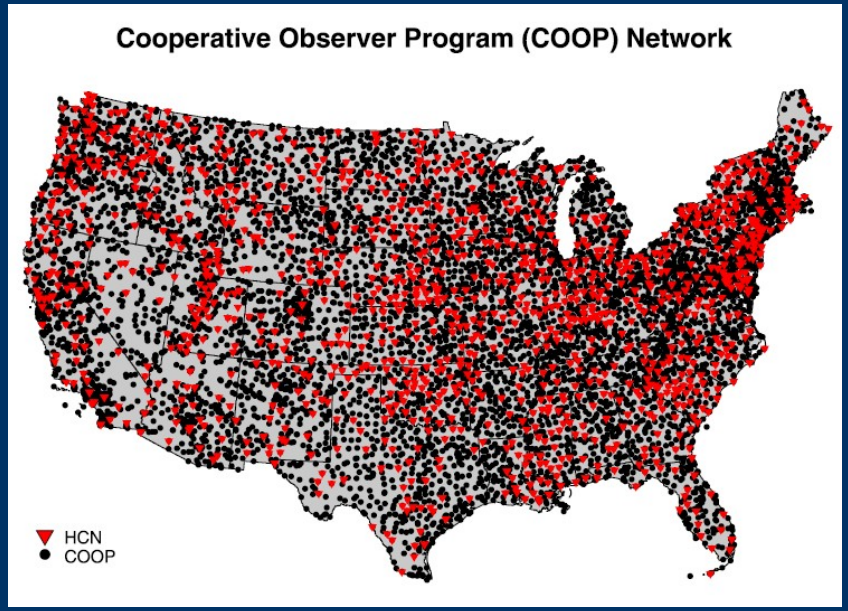
Random deviations (σ=0.36K) breaks every 24.4 years
Trend uncertainty: 0.58K/100a (empirical: 0.5K/100a)

Take home messages

- Statistical homogenisation works for recent WEIRD temperature data
  - White Educated Industrialized Rich Developed countries
  - Most of the world is not WEIRD & early data is sparse
  - Other climate elements mostly harder
- Key parameter is the Signal to Noise Ratio
  - As well as number of breaks
  - Need estimates for the entire world

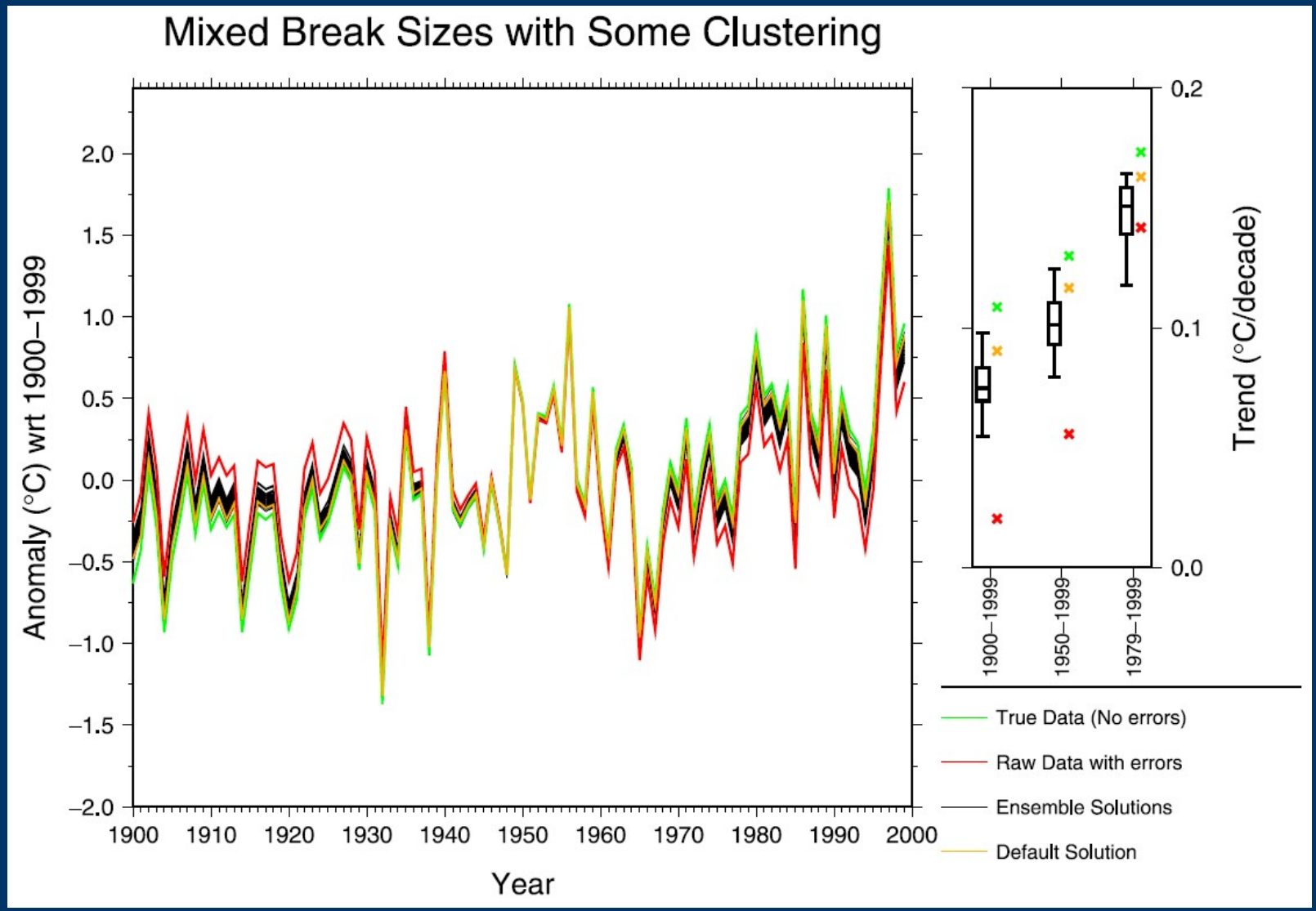
USHCN benchmark

- U.S. COOP network (7200 stations) reference for homogenising the USHCN subset (1218 stations)
- 1 COOP per 1122 km^2
- 8 scenarios for homogeneous data & inhomogeneities
  - Large-scale trend bias



Williams, C.N., M.J. Menne, and P.W. Thorne, 2012: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. J. Geophys. Res., 117, D05116, doi: 10.1029/2011JD016761.

USHCN benchmark – Mixed break sizes, some clustering



USHCN benchmark – many small breaks