The deleted chapter on future research needs

(Of the WMO Guidance on Homogenization)

Victor Venema
Victor.Venema@grassroots.is
@variabilityBlog
https://fediscience.org/@VictorVenema

Outline

- Introduction
- Physical understanding
- Break detection
- Computing uncertainties
- Deterministic or stochastic corrections?
- Correction as model selection problem

Victor.Venema@grassroots.is

https://grassroots.is

Low SNR

- Theory
  - Lindau, R. and Venema, V.K.C., 2018: The joint influence of break and noise variance on the break detection capability in time series homogenization. Advances in Statistical Climatology, Meteorology and Oceanography, 4, p. 1–18. https://doi.org/10.5194/ascmo-4-1-2018
- Statistical properties of inhomogeneities
  - Lindau, R, Venema, V., 2019: A new method to study inhomogeneities in climate records: Brownian motion or random deviations? International Journal Climatology, 39: p. 4769– 4783. Manuscript. https://eartharxiv.org/vjnbd/ https://doi.org/10.1002/joc.6105
  - Lindau, R, Venema, V., 2020: Random trend errors in climate station data due to inhomogeneities. International Journal Climatology, 40, pp. 2393-2402. Open Access. https://doi.org/10.1002/joc.6340
- Numerical confirmation
  - 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.
- Statistical homogenization under-corrects any station network-wide trend biases
  - https://variable-variability.blogspot.com/2020/05/statistical-homogenization-under-correction-trends.html

Victor.Venema@grassroots.is

https://grassroots.is

Climate elements

- We work too much on temperature
  - Other variables often have a lower SNR
- Combined dataset with all elements
  - Copernicus project
  - Help in understanding reasons for inhomogeneities
    - Help for detection?

Victor.Venema@grassroots.is

https://grassroots.is

Scales & goals

- Station, regional, network, continental, global
- Different scales need different methods
- Example:
  - Imagine a network-wide inhomogeneity
  - Important for the network average, but small compared to station noise
  - Most stations it will not be statistically significant, correction likely have a larger error than the break size
  - Correcting it would increase error on a station scale
  - But would decrease the error on the network scale
- Estimate statistical properties of inhomogeneities from raw data

Victor.Venema@grassroots.is

Independent lines of evidence

- Climatology is an observational science
- Independent lines of evidence
- Warming
  - Station record
  - SST observations (lake & river temperatures)
  - Radiosonde observations
  - Satellite retrievals (many different kinds)
  - Paleo reconstructions (many different kinds)
  - Changes in nature and agriculture
  - Sea level rise
  - Melting glaciers
  - Theoretical expectation

Victor.Venema@grassroots.is

https://grassroots.is

Independent lines of evidence

- Station warming estimates
  - Different station classes
  - Comparison with neighbouring networks
  - Comparison of the corrections with parallel measurements

Victor.Venema@grassroots.is

https://grassroots.is

Physical understanding

- Parallel measurements
  - KNMI (Brandsma) wrote a WMO report
- Physical modelling (RT, flow)
- Historical information on measurement methods
  - Marine data community has this
  - Room observation
  - Window observations
  - Sheds, stands and huts
  - Stevenson screens
  - AWS
  - Surrounding (urbanization, airports, irrigation)

Victor.Venema@grassroots.is

https://grassroots.is

Break detection

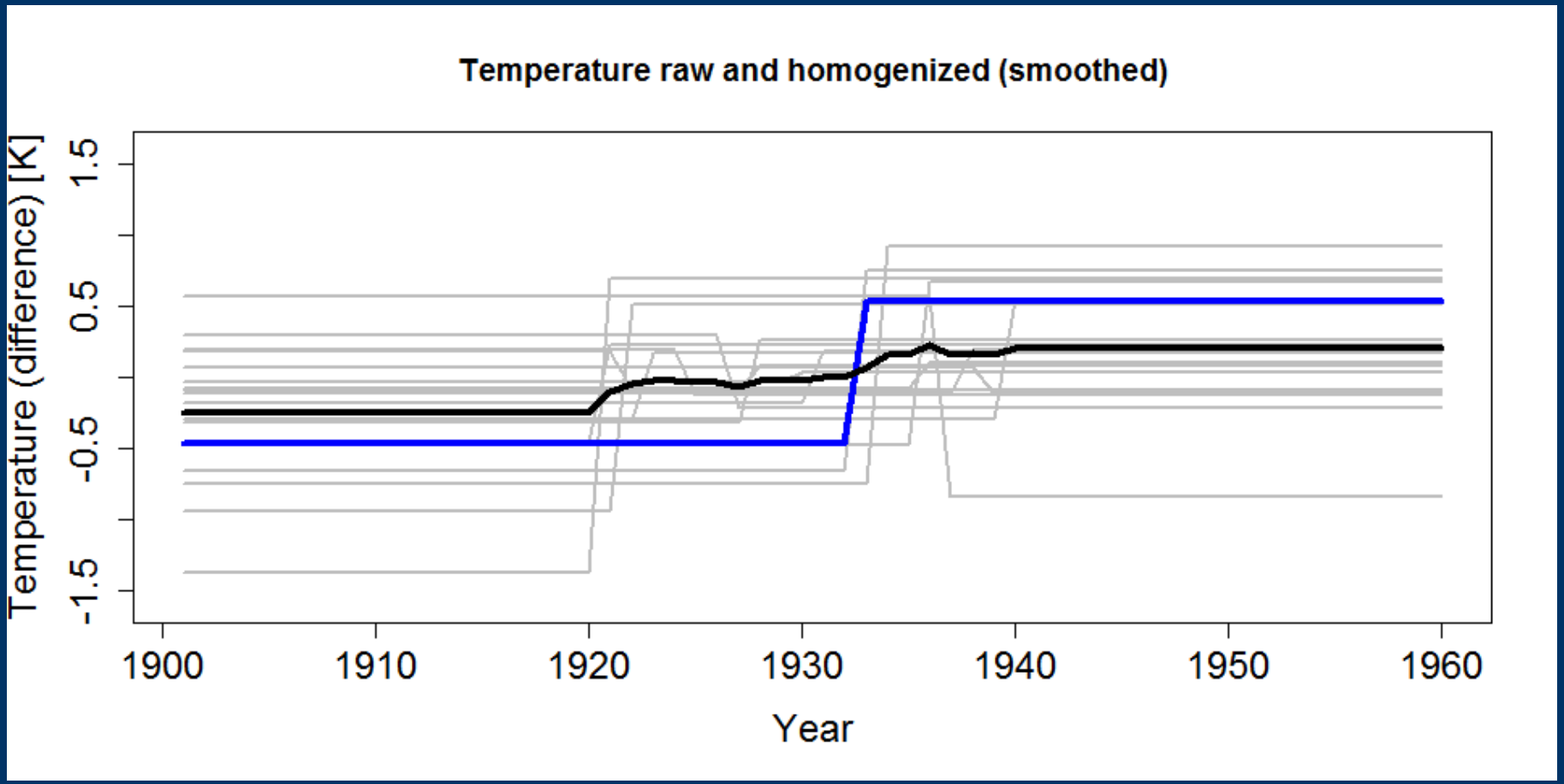
- Multiple breakpoint problem
  - Multi-test
- Inhomogeneous reference problem
- Compute the SNR before you homogenize
  - Another significance level does not solve the SNR problem
  - Compute statistical properties of the inhomogeneities from raw data
    - Temporal correlations (in the difference series)
- Joint detection

Victor.Venema@grassroots.is

https://grassroots.is

Break detection

- Network-wide breaks



Victor.Venema@grassroots.is

https://grassroots.is

Computing uncertainties

- Source:
  - Also after homogenisation uncertainties remain in the data due to various problems: Not all breaks in the candidate station have been and can be detected.
  - False alarms are an unavoidable trade-off for detecting many real breaks.
  - Uncertainty in the estimation of correction parameters due to limited data.
  - Uncertainties in the corrections due to limited information on the break positions.
- We have estimate from benchmarking high SNR networks
  - Only few on: difficult situations (islands, El Nino, mountains, ...)
- Brohan et al. (2006): Errors from a few known inhomogeneities assumed to be similar world-wide

Victor.Venema@grassroots.is

https://grassroots.is

Deterministic or stochastic corrections?

- Authors of daily correction methods find their method improve estimates
- Independent validation is less optimistic

- Purely deterministic
  - For example, a calibration problem:
    - 0°C before and after the break the same
    - At 30°C after the break 2°C warmer
- Purely stochastic
  - For example, a new response time
- How to estimate this from data?
  - Noise in the difference series for identical instruments

Victor.Venema@grassroots.is

https://grassroots.is

Correction as model selection problem

- The number of degrees of freedom (DOF) of the various correction methods varies widely.
  - From just one DOF for annual corrections,
  - 12 DOF for monthly correction
  - 24 DOF for monthly mean and standard deviation
  - 40 DOF for decile corrections applied to every season
  - Large number of DOF for quantile/percentile matching
- HOME, with PRODIGE & for Europe
  - Temperature: monthly
  - Precipitation: annual
  - Will depend on the SNR, on the break
- Model selection problem, with a penalty for complexity

Victor.Venema@grassroots.is

Q&A

- Physical understanding
- Break detection
- Computing uncertainties
- Deterministic or stochastic corrections?
- Correction as model selection problem

- https://variable-variability.blogspot.com/2020/10/deleted-chapter.html
- https://fediscience.org

Victor.Venema@grassroots.is

https://grassroots.is