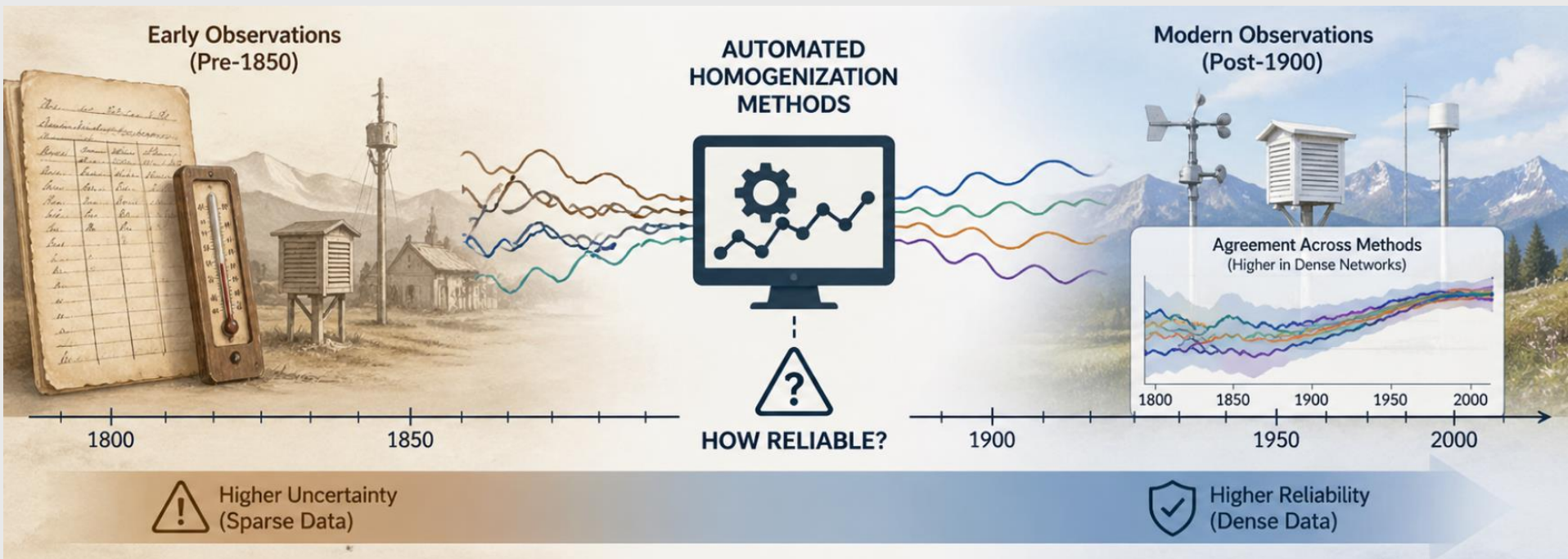


# How Reliable Are Automated Homogenization Methods for Early Climate Records?



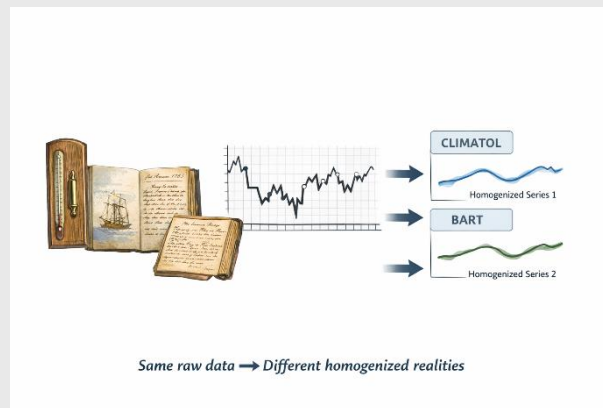
**Elin Lundstad**

Norwegian Meteorological Institute

12th Seminar for Homogenization and Quality Control in Climatological Databases

# Motivation

- Early instrumental data = backbone of historical climate analysis
- Records before ~1850 are:
  - sparse
  - heterogeneous
  - poorly documented
- Automated homogenization increasingly used  
→ **But how reliable are these methods for early instrumental data?**



# Key challenges

- No true reference climate ("ground truth")
- Sparse station networks
- Changing instruments and practices
- Metadata often missing or incomplete
- Strong climatic variability (volcanic eruptions, etc.)



Norwegian Meteorological Institute

# Research Questions

- How consistent are different automated methods?
- How sensitive are results to:
  - network density
  - data availability
- Do methods behave differently in:
  - dense networks (France)
  - sparse networks (South-East Asia)?
- How do results compare with 20CRv3?



# Data



## HCLIM database

- Global early instrumental monthly dataset
- Variables: temperature, precipitation, pressure, rainy days
- Coverage: 1658–present
- Focus:
  - France (dense network)
  - South Asia (sparse network)

# Data



## HCLIM database

- Global early instrumental monthly dataset
- Variables: temperature, precipitation, pressure, rainy days
- Coverage: 1658–present
- Focus:
  - France (dense network)
  - South East-Asia (sparse network)

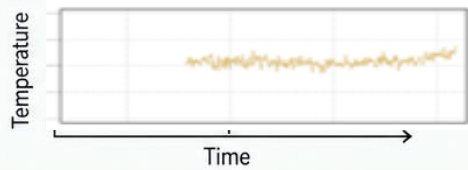
# Methods

## Automated homogenization:

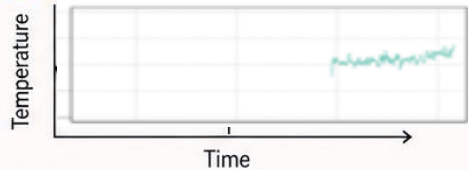
### CLIMATOL



### BART



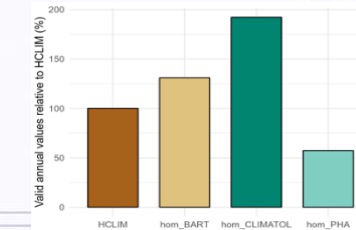
### PHA



## Evaluation metrics:



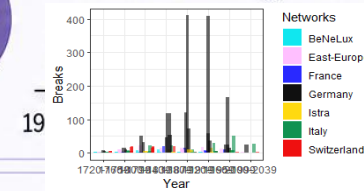
### Data retention after preprocessing



% of data retained



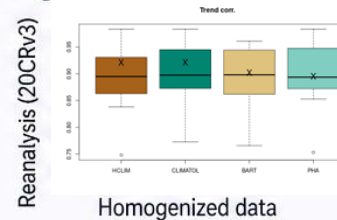
### Number & timing of breakpoints



Count and timing of breakpoints



### Agreement with reanalysis (20CRv3)



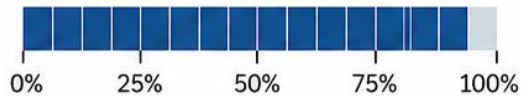
Correlation, RMSE, bias

# Data Retention

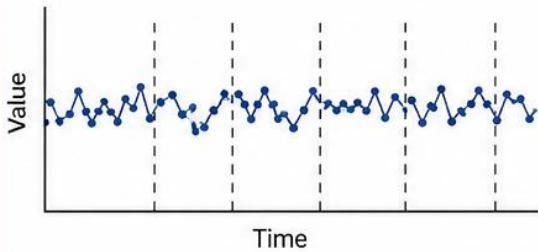
## CLIMATOL



~96% retained



Data Retention

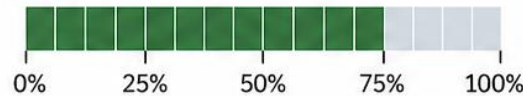


More data retained → Better coverage,  
but includes more shorter segments

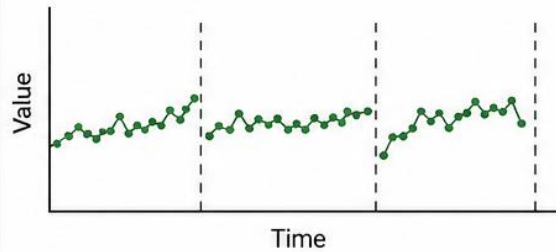
## BART



~80% retained



Data Retention

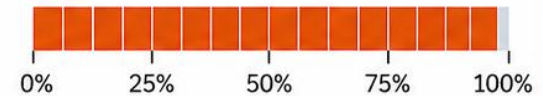


Fewer but longer records →  
More consistent series

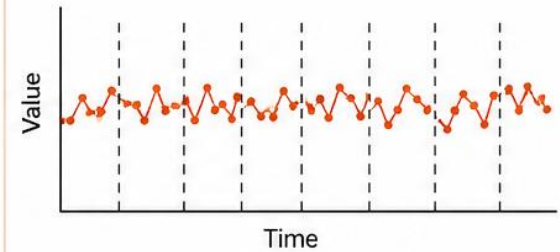
## PHA



~99% retained



Data Retention



Maximum data retained →  
Highest coverage, includes  
many short segments



## Different philosophies of data handling

CLIMATOL prioritizes maximum data retention,  
BART prioritizes consistency and longer record lengths,  
PHA aims for near-complete retention with maximum coverage.

# Breakpoint Detection

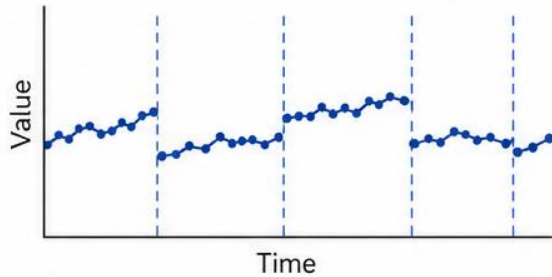
1

## CLIMATOL



**fewer breakpoints**

more conservative



Detects fewer breakpoints, focusing on larger, more robust shifts.

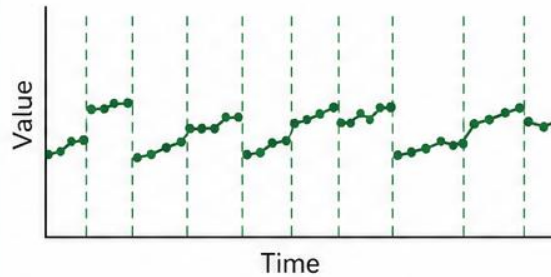
2

## BART



**~8× more breakpoints**

sensitive to small shifts



Detects many breakpoints, including small magnitude shifts.

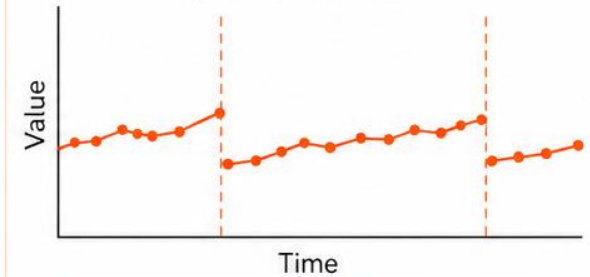
3

## PHA



**fewest breakpoints  
(most conservative)**

focus on robust,  
persistent shifts



Detects the fewest breakpoints, focusing on statistically robust, persistent shifts. May miss smaller or short-lived inhomogeneities.



### Structural differences between methods

**CLIMATOL** applies a balanced approach.

**BART** favors sensitivity and detects more (including smaller) shifts.

**PHA** is the most conservative, focusing on robust, persistent breakpoints.

# Dense vs. Sparse Networks

## France (dense):

- Methods show reasonable agreement
- Network structure stabilizes results

## South-East Asia (sparse):

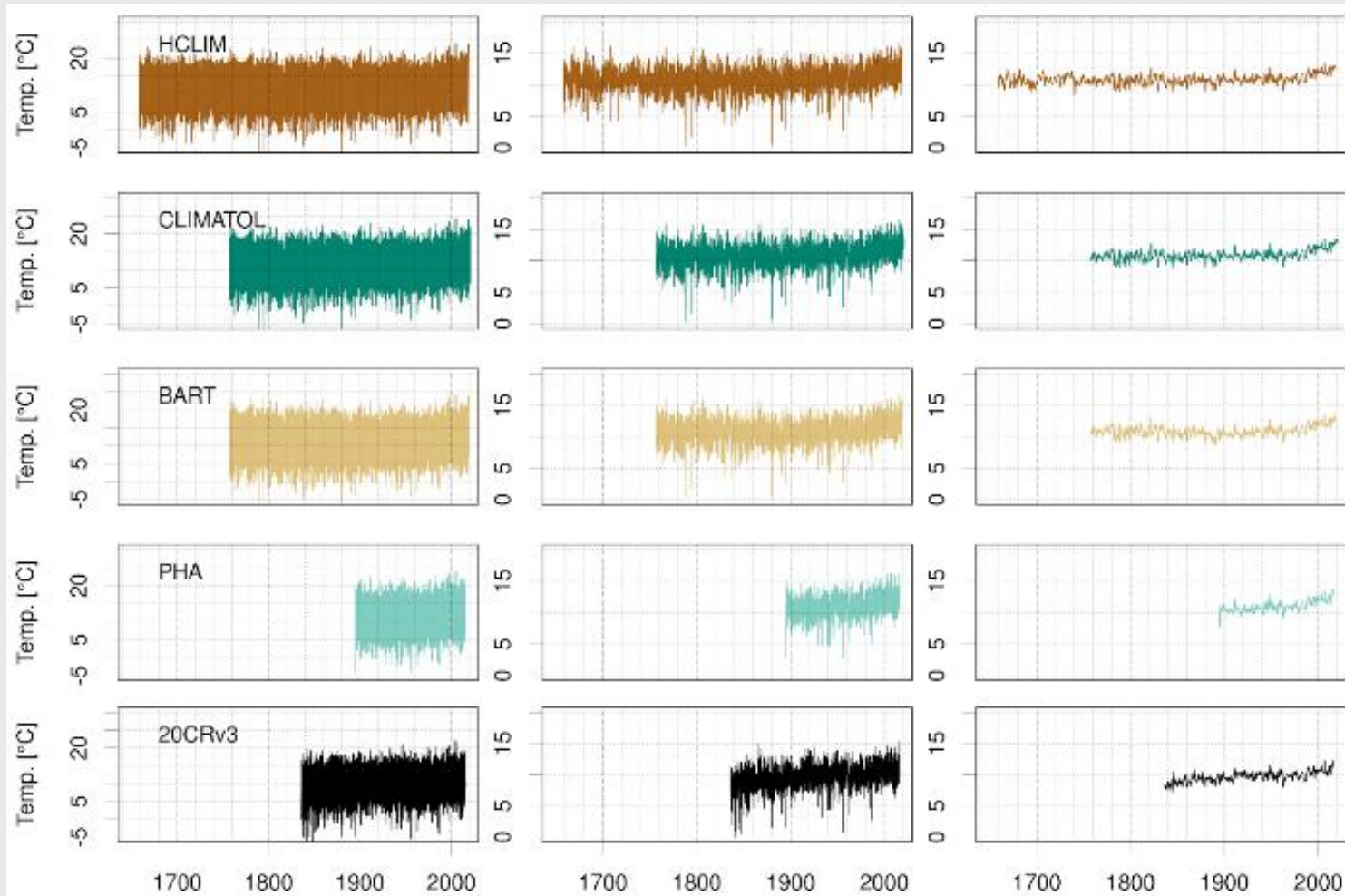
- Large methodological spread
- Higher uncertainty
  - Network density matters critically



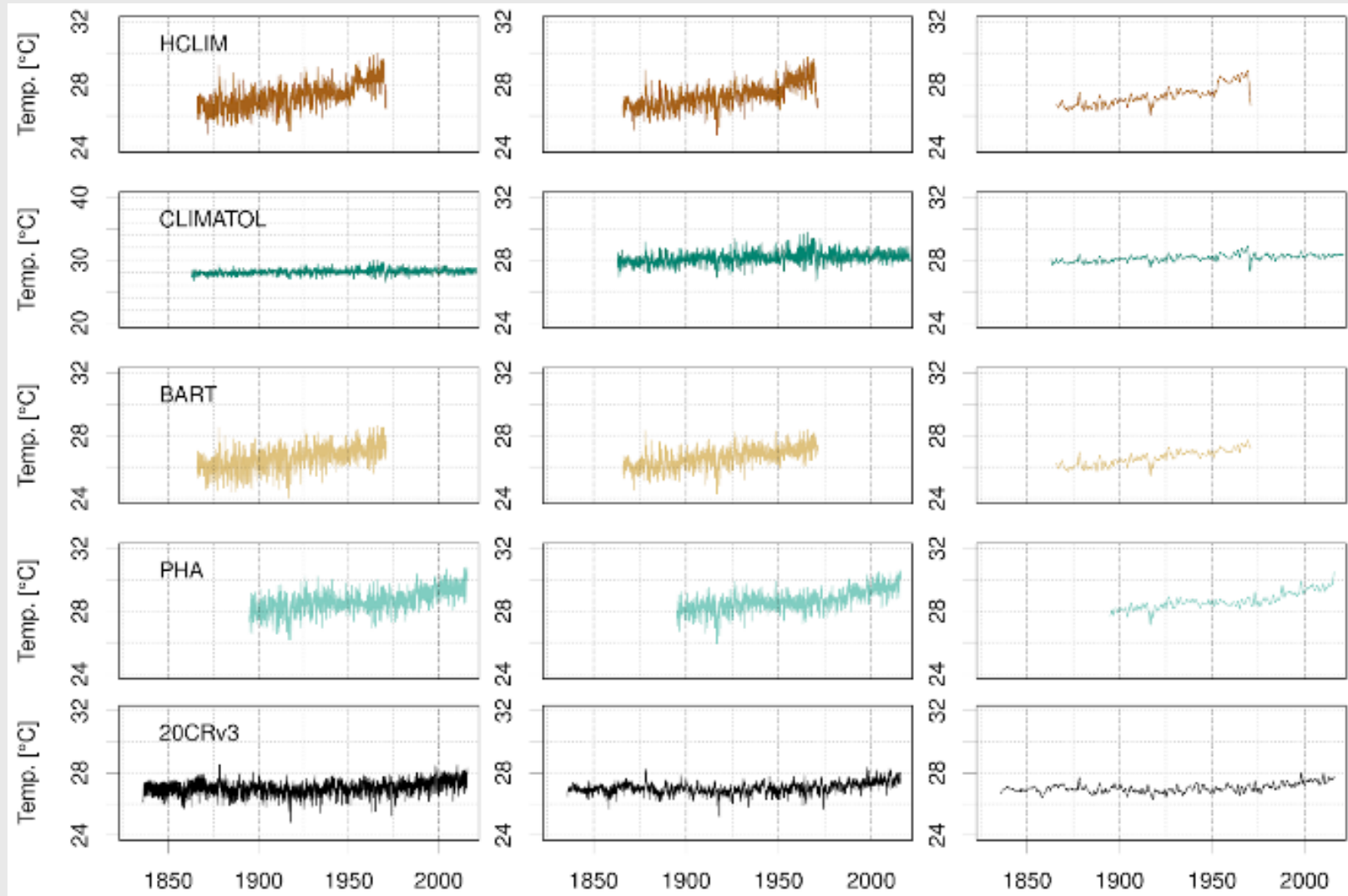
# Comparison with 20CRv3



Paris

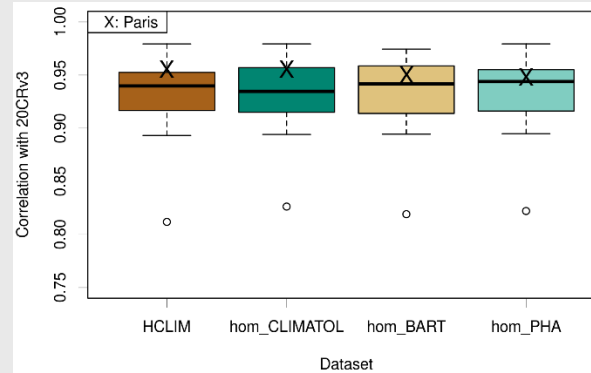
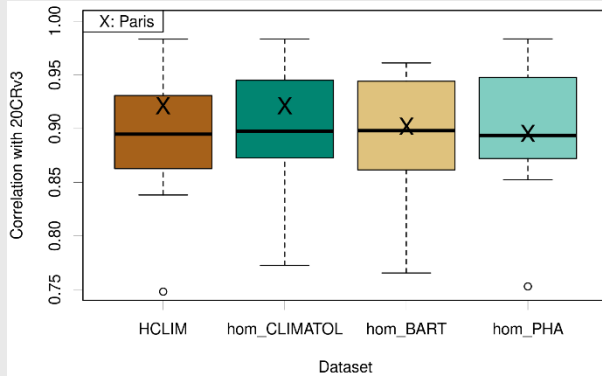


# Comparison with 20CRv3

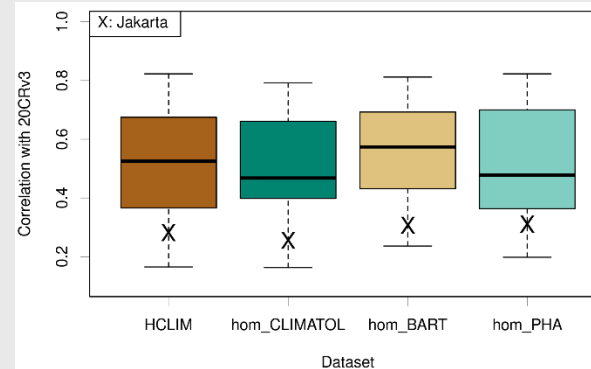
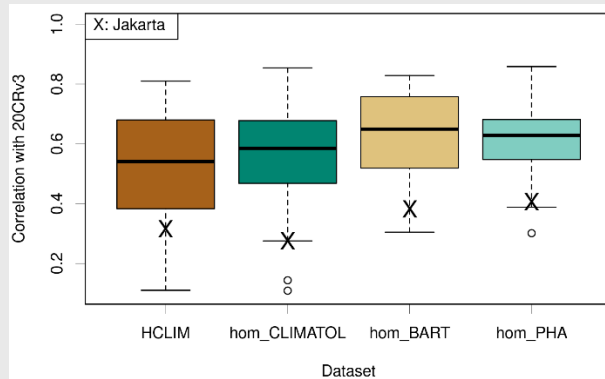


# Comparison with 20CRv3

Paris

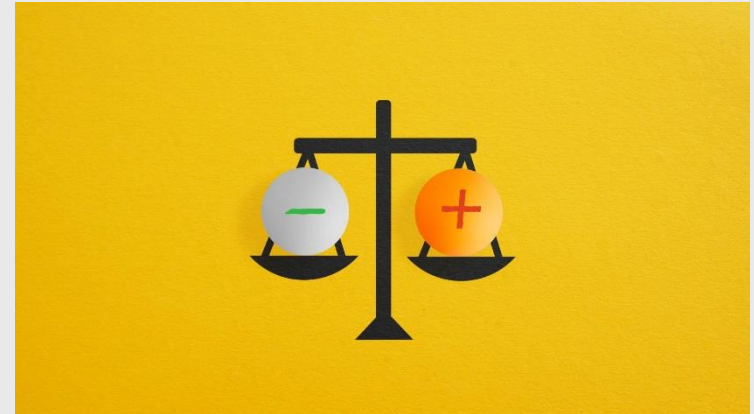


Jakarta



# Reliability Assessment

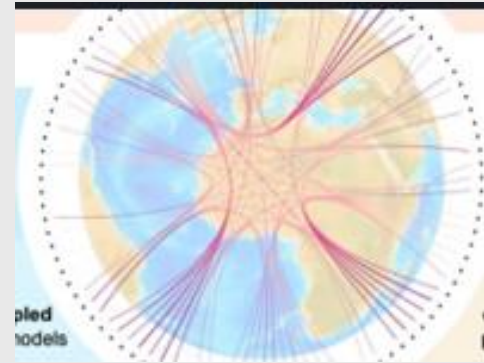
- Automated methods are:
  - **Useful**
  - **Scalable**
  - **Systematic**
  - **Reproducible processing**
- BUT:
  - sensitive to network structure
  - method-dependent outcomes
- No single "best" solution exists



**Method choice matters, but data structure matters more**

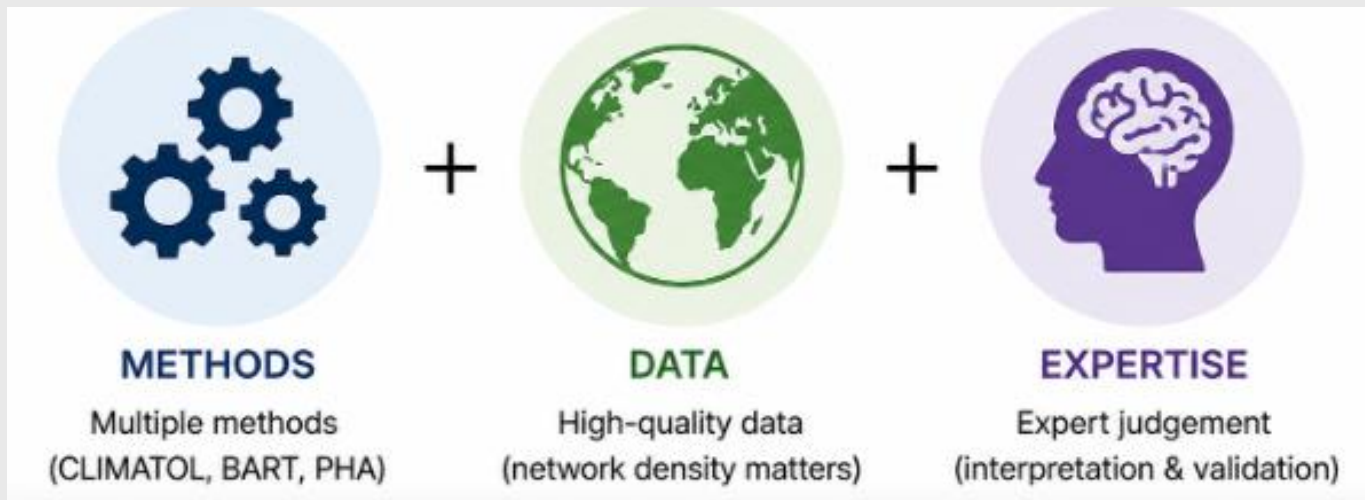
# Implications

- Early climate trends depend on:
  - method choice
  - network density
- Uncertainty must be:
  - quantified
  - communicated
- Multi-method approaches recommended



# Conclusions

- Automated homogenization:
  - works reasonably in dense networks
  - large uncertainty in sparse regions
- Method choice matters
- No objective “correct” solution exists  
→ Expert judgement remains essential



# Outlook

- Expand benchmarking with:
  - synthetic tests
  - multi-method ensembles
- Better metadata rescue
- Hybrid approaches:
  - statistics + expert knowledge
  - cautious ML integration


## ► Key message

**Robust homogenisation requires combining methods, data, and expertise**

# Hclim

## Database

### *Citation:*

**Lundstad, Elin; Brugnara, Yuri; Brönnimann, Stefan (2022):** Global Early Instrumental Monthly Meteorological Multivariable Database (HCLIM) [dataset bundled publication]. *PANGAEA*,  <https://doi.org/10.1594/PANGAEA.940724>

## Article

Lundstad, E., Brugnara, Y., Pappert, D. *et al.* The global historical climate database HCLIM. *Sci Data* **10**, 44 (2023). <https://doi.org/10.1038/s41597-022-01919-w>



Norwegian  
Meteorological  
Institute

# Thanks for your attention!

T.           +47 94 88 10 28  
E.           [elinl@met.no](mailto:elinl@met.no)  
BS.         elinlun.bsky.social



Norwegian  
Meteorological  
Institute