Challenges in Homogenizing Climate Data and Assessing Trend Representativeness

Xiaolan L. Wang

Climate Research Division, Environment and Climate Change Canada

12th Seminar for Homogenization and
Quality Control in Climatological Databases
Budapest, Hungary
4-8 May 2026

Outline

Challenges common to all variables

Metadata-based versus undocumented changepoint detection

Mean-adjustments *vs* Quantile-Matching (QM) methods

Additional challenges in QC & homogenization of non-Gaussian variables (e.g., precipitation, wind speed)

Challenges common to all variables

The first challenge is reference series selection or construction

- Best practice for reference selection:
  - ➢ Correlation matters more than distance

    A developing urban station series should not be used as reference to homogenize a rural station series, and vice verse, even if they are located close to each other.
  - ➢ Prioritize correlation significance (α), not correlation value:

    N1=40,   R1=0.3, p = 1- α < 0.95

    N2=100, R2=0.2, p = 1- α >0.95

    R2 is more significant than R1, although R2 < R1

Common Critical Values (α = 0.05, two-tailed)

n = 10	r_{crit} = ±0.632
n = 20	r_{crit} = ±0.444
n = 30	r_{crit} = ±0.361
n = 50	r_{crit} = ±0.279
n = 100	r_{crit} = ±0.197

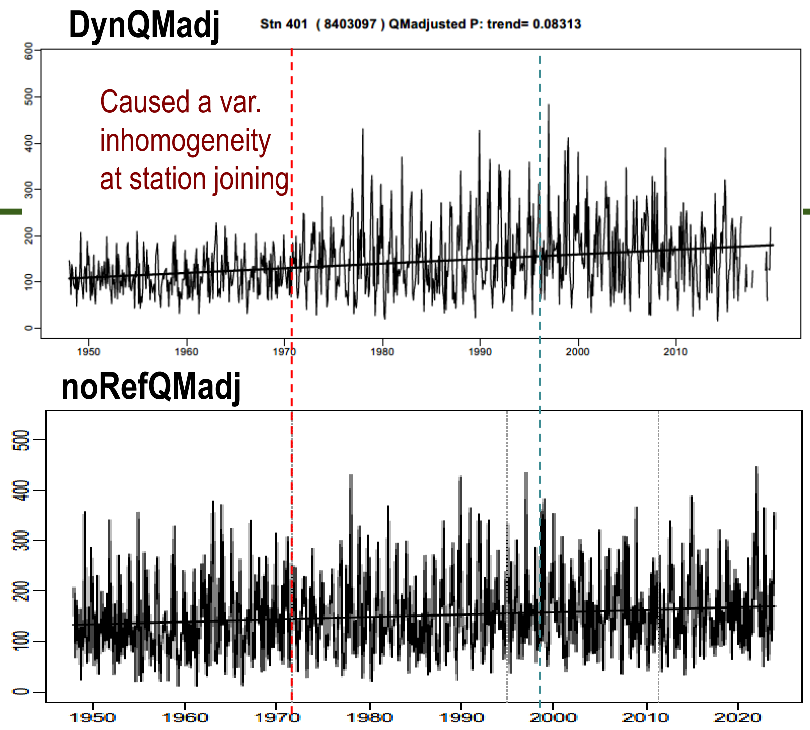
- **Reference series should faithfully represent the climate signal at the base-station location.**

  Climate signal includes trend and low-frequency variability (e.g., ENSO, PDO, …)

  Trend could be linear or non-linear.

When base and reference series share the same climate signal, the base-reference difference is free of climate signal and can be tested using the RHtests package (https://github.com/ECCC-CDAS). This ensures that the climate signal itself is not mistakenly adjusted (see next slide).

An example showing that use of a poor reference series can introduce new artifacts into the adjusted series →



In Wang et al. (2023; https://doi.org/10.1175/JCLI-D-23-0193.1):

DynQMadj: QM adjustment using the best-correlated homogeneous reference segments

(possible to use diff. Refs for diff. Cs in the same base series)

noRefQMadj: QM adjustment applied without using a reference series

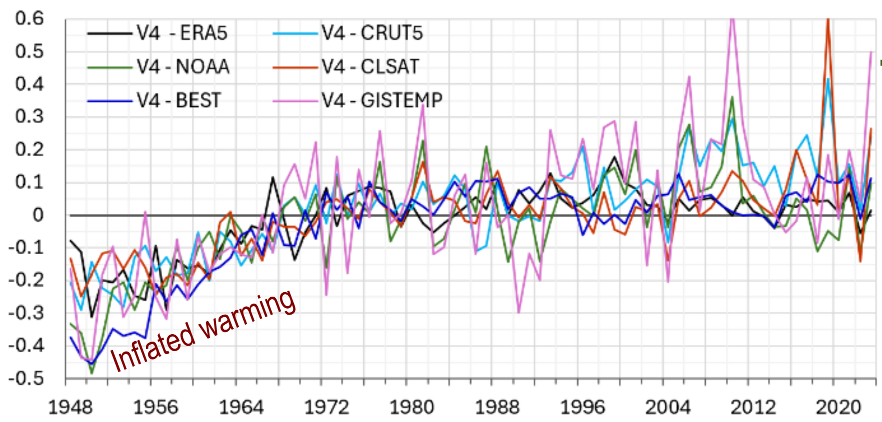
DynQMadj shows no performance gain because the Canadian precipitation network is too sparse to support a robust reference series. This limitation is specific to this variable and network density; other variables (e.g., temperature) or denser precipitation networks may yield different results.

Tip: If the reference series has a different linear trend (i.e., an imperfect reference), the base-minus-reference series can be treated as a new base series for running **FindU** and **StepSize** in RHtests.

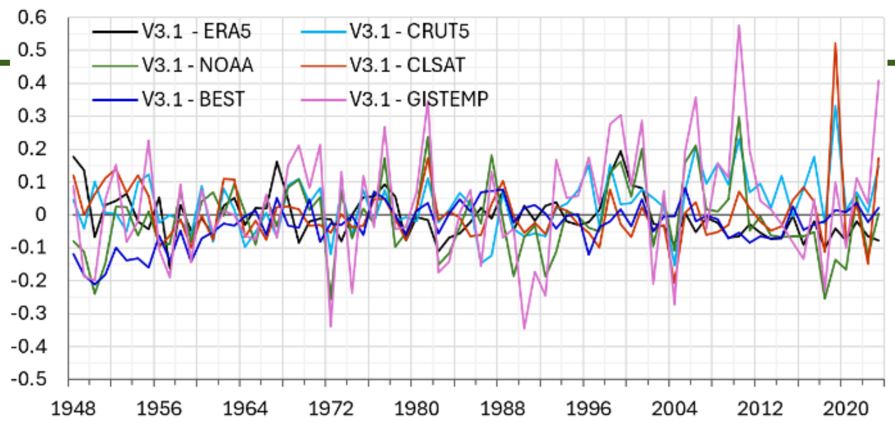
Homogenization without reference data is generally more uncertain than reference-based homogenization and can introduce substantial biases in estimated trends (see example on the next slide).

V4 (no Ref) **and V3.1 in comparison with six international/global datasets** (Zwiers & Wang, 2026; https://doi.org/10.1080/07055900.2026.2657986)

V4 compared with six international datasets



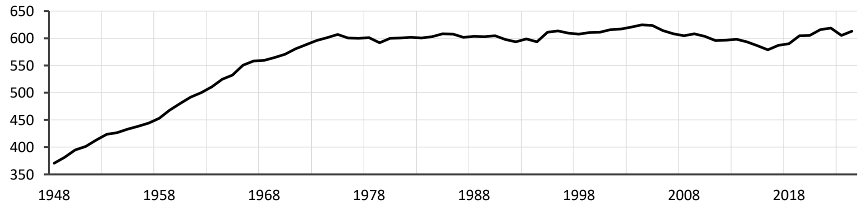
V3.1 compared with six international datasets



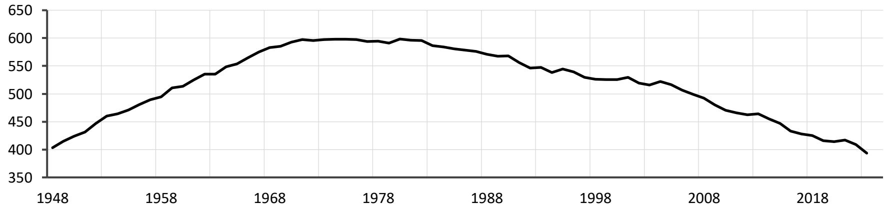
1948-2023
trends: C/10yr
(outlier) V4: 2.25
V3.1: 1.95
Avg6:1.94
ERA5: 1.95
CRUT5: 1.88
NOAA: 2.03
CLSAT: 1.95
BEST:1.88
GISST: 1.92

Differences (°C) between anomalies in annual average daily mean temperature (relative to the 1961−1990 baseline normal) for Canada as a whole during the 1948−2023 period for (a) CanGridT mlyV4 and ERA5, CRUT5, NOAA-MLOST,and C-LSAT2.0, BEST, GISST, and (b) CanGridT mlyV3.1 and those same six datasets.

V4 data availability



V3.1 data availability



Annual count of stations with non-missing monthly mean temperature data in (a) the CanHomT mlyV4 dataset (excluding estimated values) and (b) the HomT mlyV3.1 dataset. A complete record with no missing data would show 651 stations for all years in (a) and 632 stations in (b).

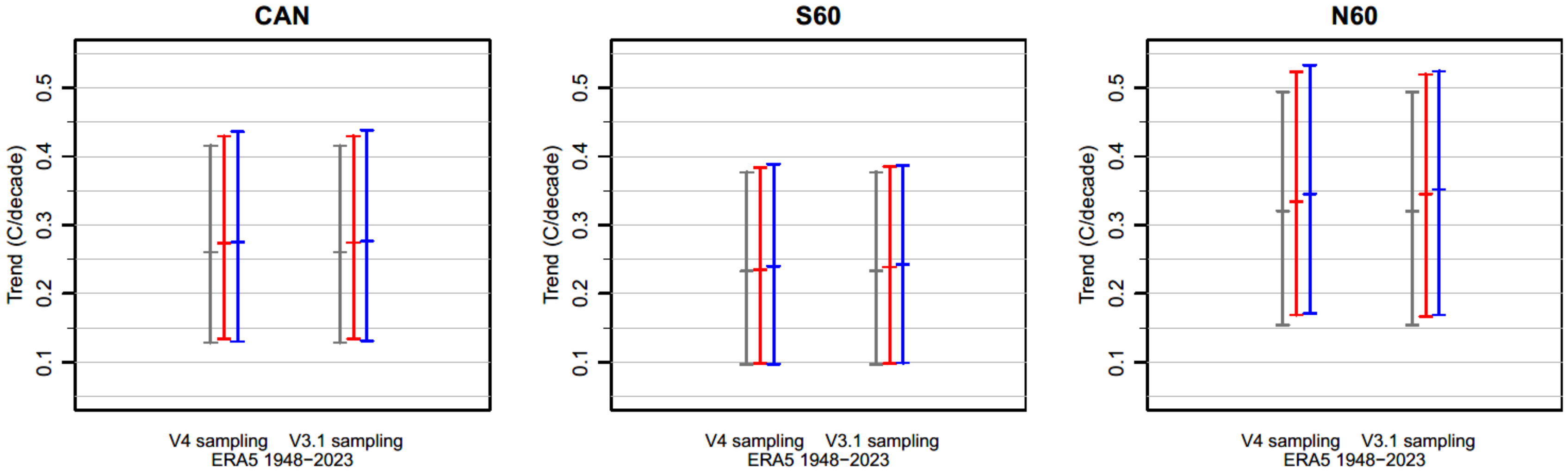
The warming difference between V4 and V3.1 is not driven by differences in data availability (see next slide)

Environment and Climate Change Canada
Environnement et Changement climatique Canada

The additional missing data in V3.1 led to a slightly higher overestimate rather than an underestimate, which is inconsistent with the lower warming in V3.1 compared with V4.

Both V4 and V3.1 station network configurations represent Canada's warming reasonably well

Without missing values or homogeneity issues, V4 and V3.1 would **overestimate** Canada-wide warming over 1948-2023 by 0.11 and 0.12 °C (5.7% and 6.1%), respectively. Accounting for missing months reduces these overestimates slightly to 0.10 and 0.11 °C (5.0% and 5.4%).

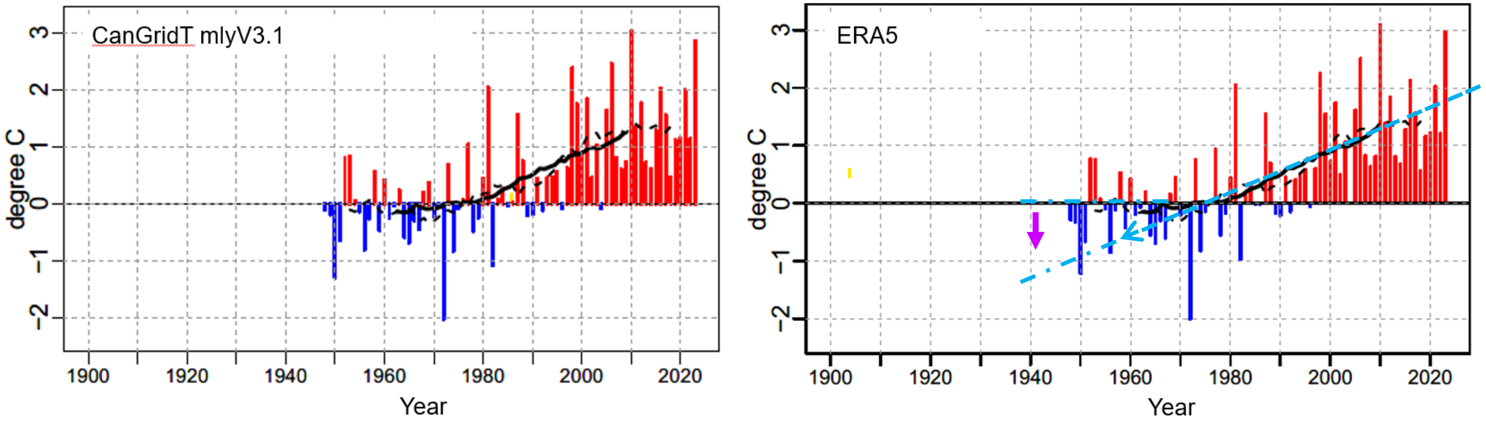


Regional average trends over 1948–2023 for a) Canada (CAN), b) southern Canada (S60; south of 60° N), and c) northern Canada (N60), estimated from complete ERA5 reanalysis data (**grey bars; benchmark trends**) and from reconstructed ERA5 data using V4 or V3.1 station sampling, with matching months of missing data (red bars) and with no missing data (blue bars). Bar lengths indicate the 95% confidence intervals of the trend estimates (central dash). Source: Zwiers & Wang (2026); https://doi.org/10.1080/07055900.2026.2657986.

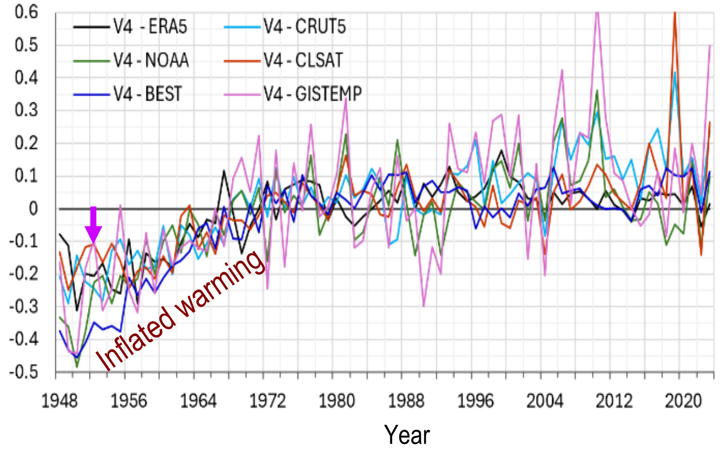
Adjusting temperature data without reference series is likely the primary cause of the substantially higher warming in V4

This is because past warming in Canada is highly non-linear, with most warming since 1948 occurred after 1970.

This pattern of warming is consistent with that observed in global mean surface temperature (Gulev, et al., 2021, Fig. 2.11).



Anomalies in Canada's spatially averaged annual average daily mean temperature (°C) relative to its baseline (1961-1990) average value, as derived from the CanGridT mlyV3.1 and ERA5 data. The black solid and dashed lines represent the 30-year and 11-year running averages, respectively.



Differences (°C) between anomalies

Without a reference series, adjustments (mean or QM) assume a linear trend over the full record and do not capture nonlinear trends. Using a reference series with the same nonlinear trend implicitly accounts for trend nonlinearity, because the trend is separated from the estimation of adjustments.

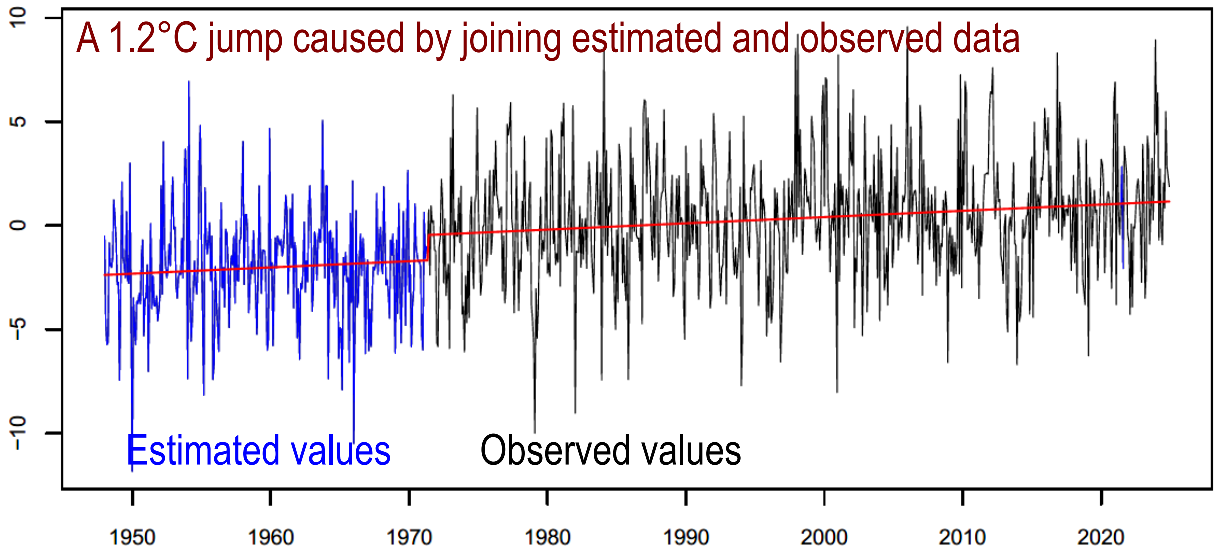
Because temperature series were adjusted to the most recent homogeneous segment (a common practice), the faster post-1970 warming rate was extrapolated backward to the pre-1970 period. This inflates the estimated warming prior to 1970 and, consequently, the overall trend for 1948-2023.

Metadata is essential for data homogenization but is often incomplete or unavailable

Changepoints with known causes must be treated differently from undocumented or uncertain changepoints

- For changepoints with known causes (type-0), it is sufficient to test only the statistical significance of the changepoint; there is no need to detect the most probable time of change. As a result, the adjustments are less uncertain.
  Changepoints with a clearly established cause can therefore be adjusted even when they are only marginally significant.

- If changepoints with known causes are treated as unknown changepoints (type-1), they are unlikely to be identified or adjusted. This can result in large biases in trend estimates, as demonstrated in CanHomT V4 (Wan et al. 2025):

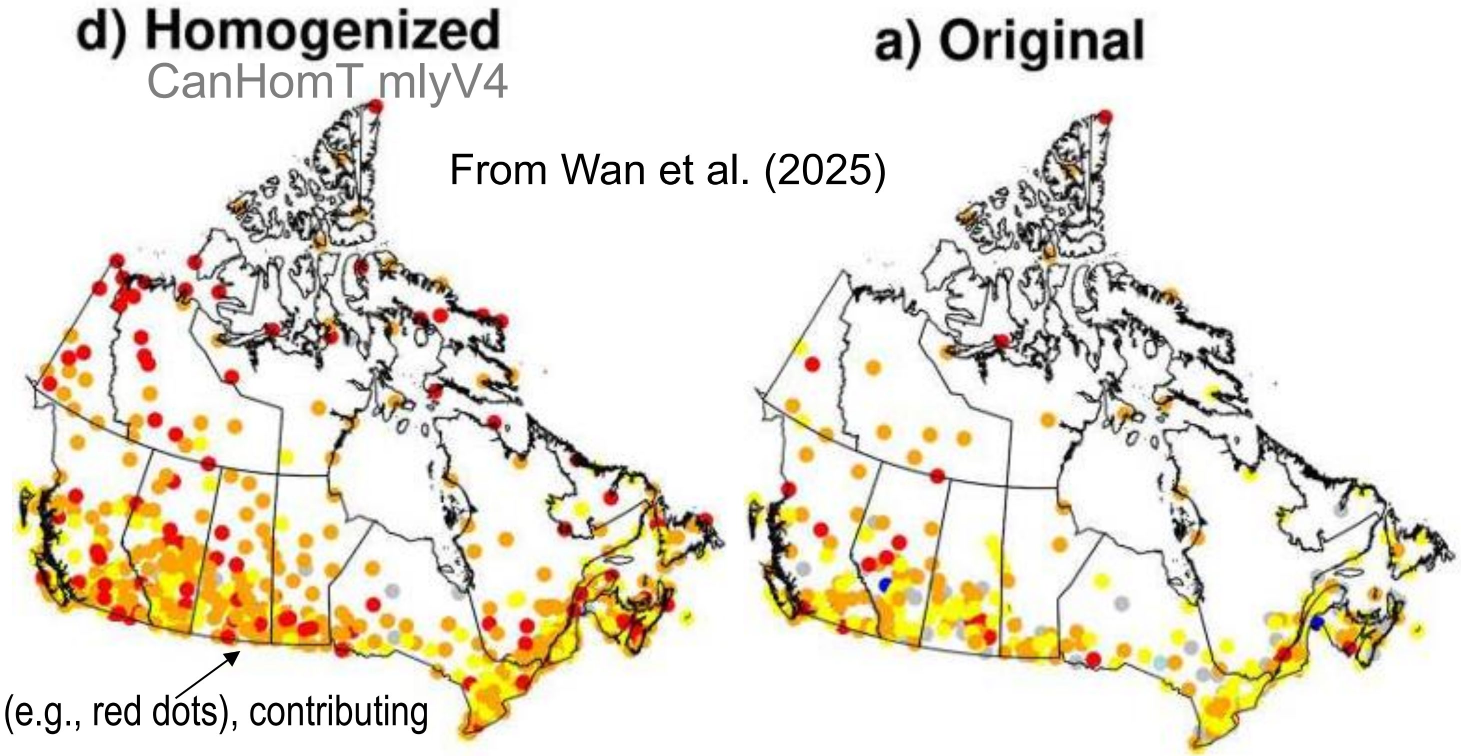


Type-0 test: 1-α =0.9998 (highly significant)
Type-1 test: 1-α =0.94 (insignificant at 5%)

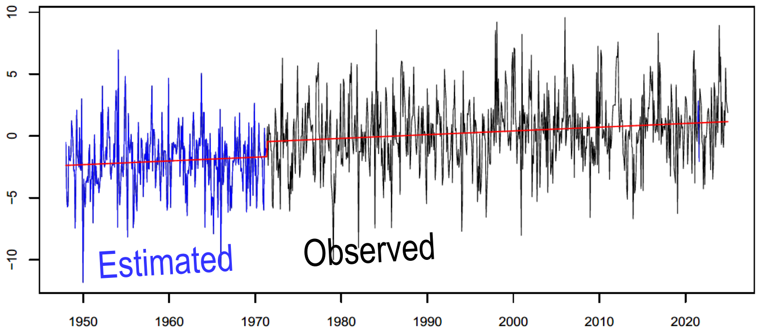
Failure to detect and correct this artificial jump inflates the trend from 0.256 to 0.493 °C per decade—nearly doubling the estimated warming.
This bias is present in V4.

Post-adjustment checks for atypical station-level trends should be an integral part of the homogenization workflow

Homogenization results require careful verification when spatial consistency of trends after homogenization is lower than the original data.



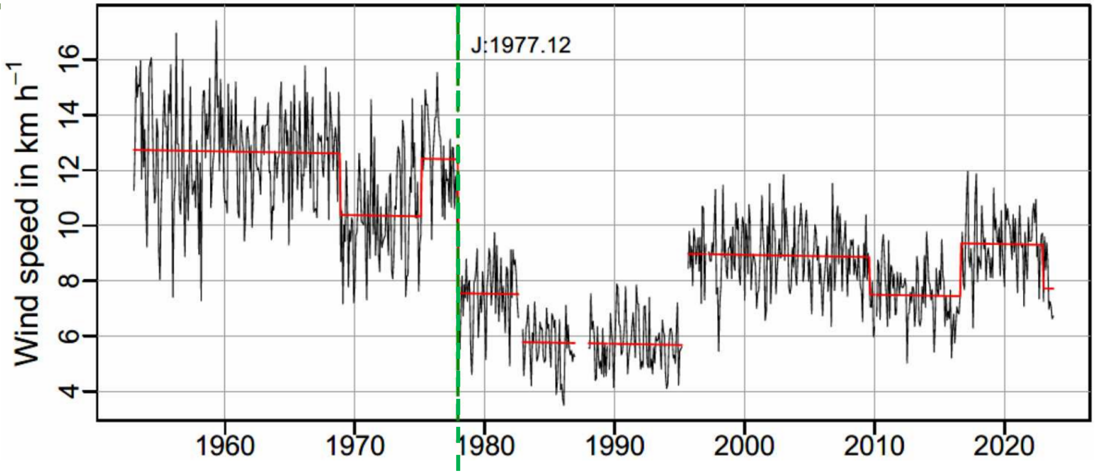
A case of improper homogenization, in which estimated data are not made homogeneous with observed data:



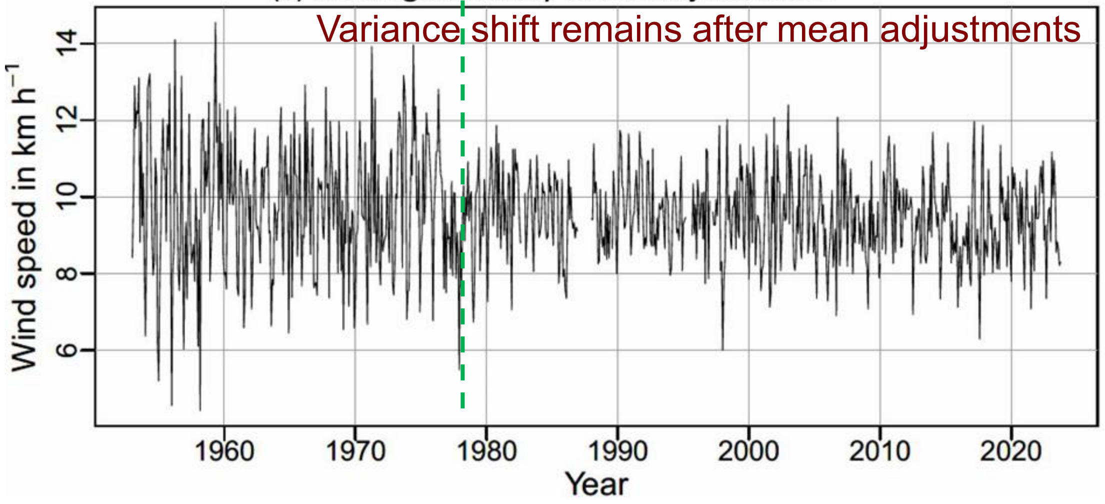
Non-climatic jumps like this occur at many stations (e.g., red dots), contributing to the substantially higher warming in V4.

Mean adjustment versus Quantile-Matching (QM) methods

(a) Unhomogenized

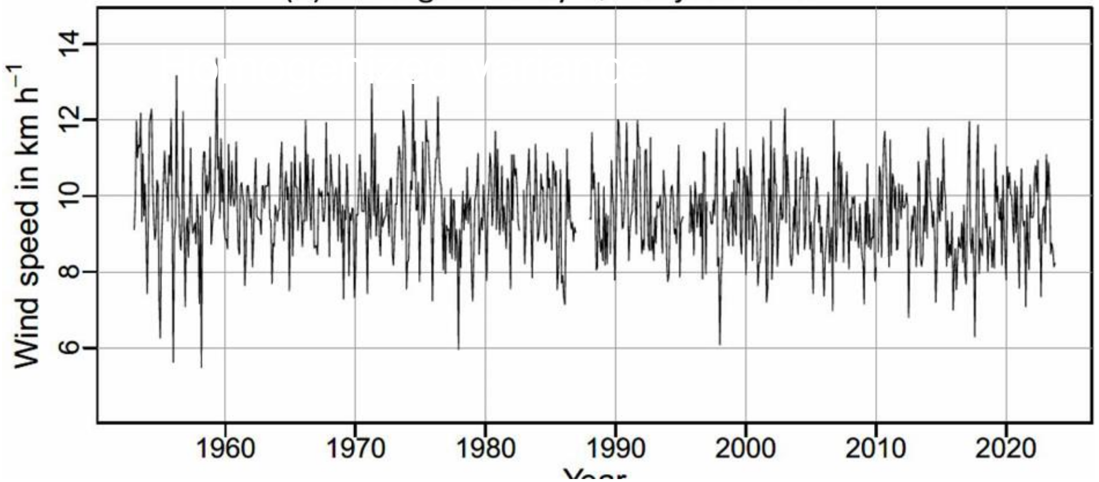


(c) Homogenized by mean adjustments



QM is superior to mean adjustments when enough data are available (i.e., segments are not too short) to estimate QM adjustments reliably.

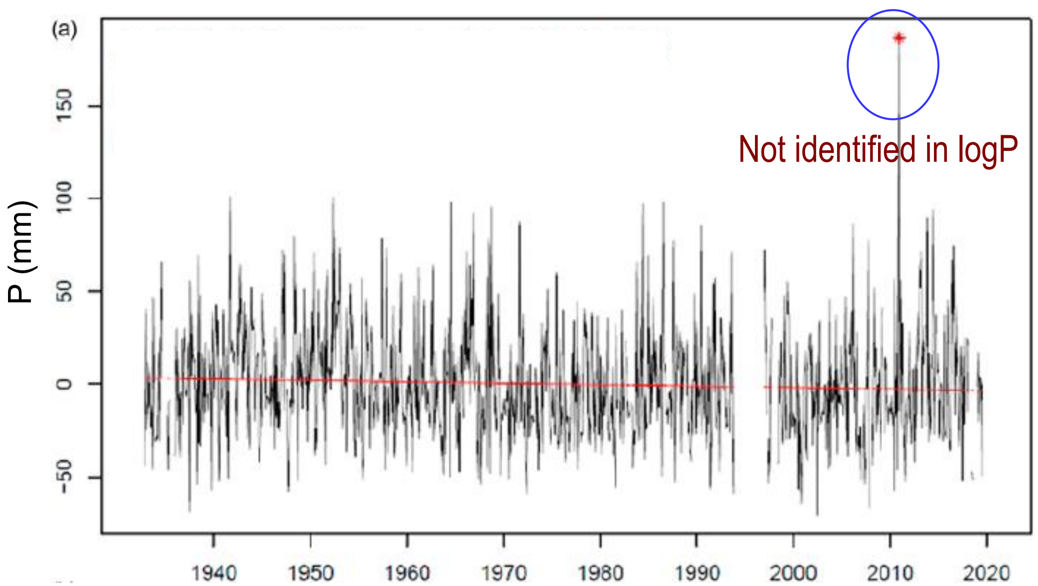
(b) Homogenized by QM adjustments

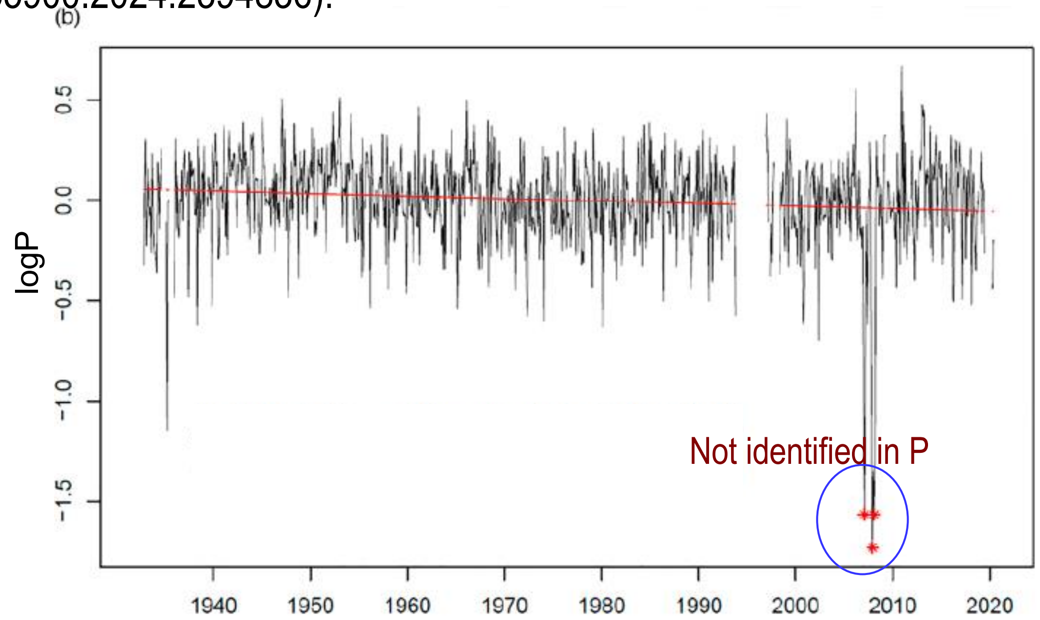


Additional challenges in precipitation and wind data homogenization

- Precipitation and wind data are non-Gaussian, non-negative; therefore, tests designed for Gaussian variables are not directly applicable.
- Best practice: Apply QC and changepoint detection tests to both the original precipitation (P) series and its log-transformed form (logP).
  - Testing P is effective for detecting large erroneous values or inhomogeneities that primarily affect high values
  - Testing logP is effective for detecting zero or near-zero values (i.e., small values)

Example (Cheng, Wang and Feng, 2024; https.//doi.org/10.1080/07055900.2024.2394836):





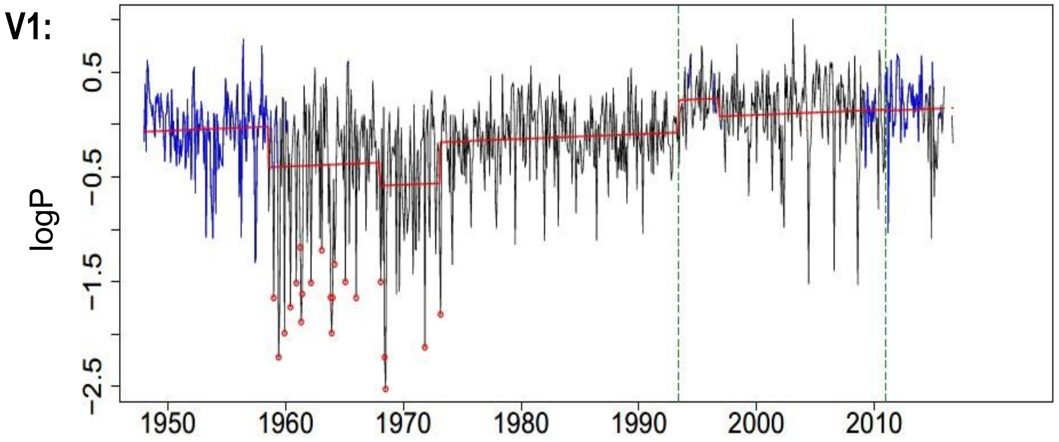
Source: Cheng, Wang & Feng (2024); https.//doi.org/10.1080/07055900.2024.2394836

Critical to identify and treat false zero values in monthly precipitation data

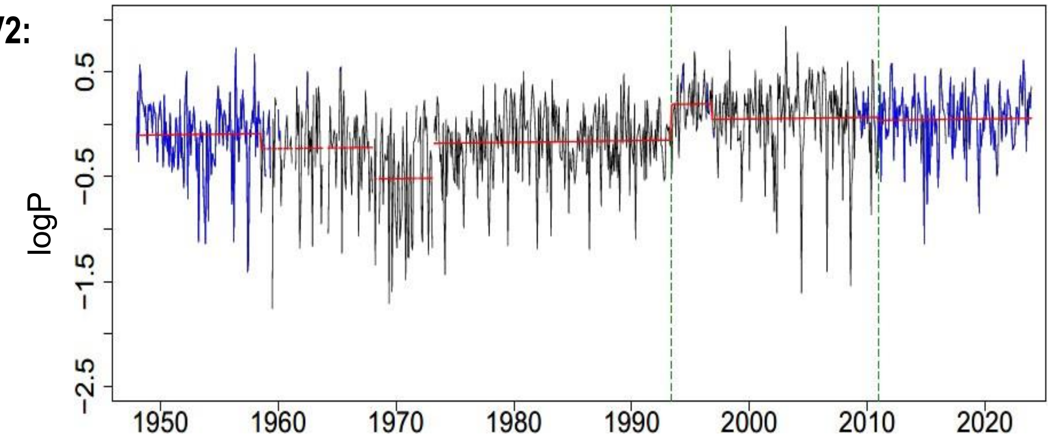
Example (CanHomP V1 *v.s.* V2):

The V1 series below contains 20 months of mis-recorded missing values (red circles) prior to 1973; these were identified and set to missing in V2.

V1:



V2:

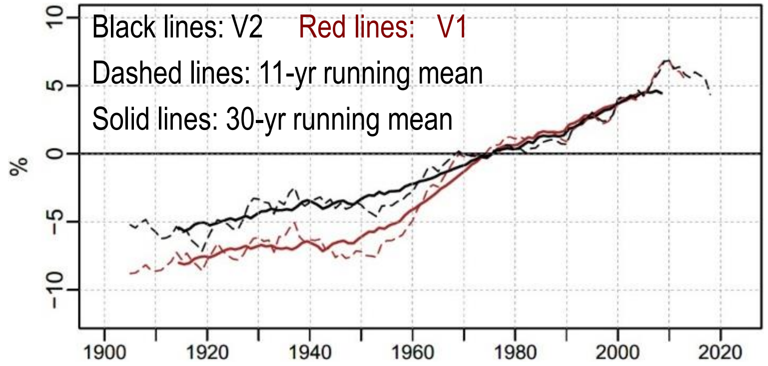


This issue occurred at many stations prior to 1970, particularly in the North.

Untreated false zero months are interpolated over station-sparse regions

→ Low bias in Canada's average annual total precipitation

→ Artificial rapid increase (1950-1970; red dashed line) -->

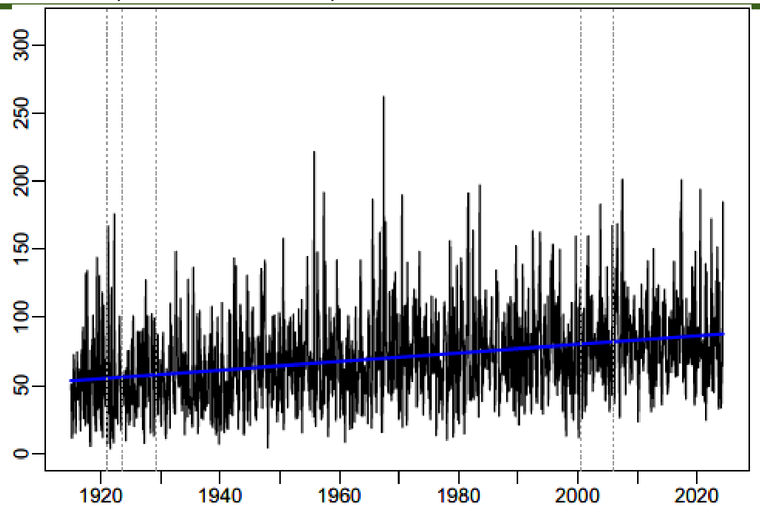


11-year and 30-year running means of anomalies in Canada's average annual total precipitation, expressed relative to the 1961-1990 climatological mean.

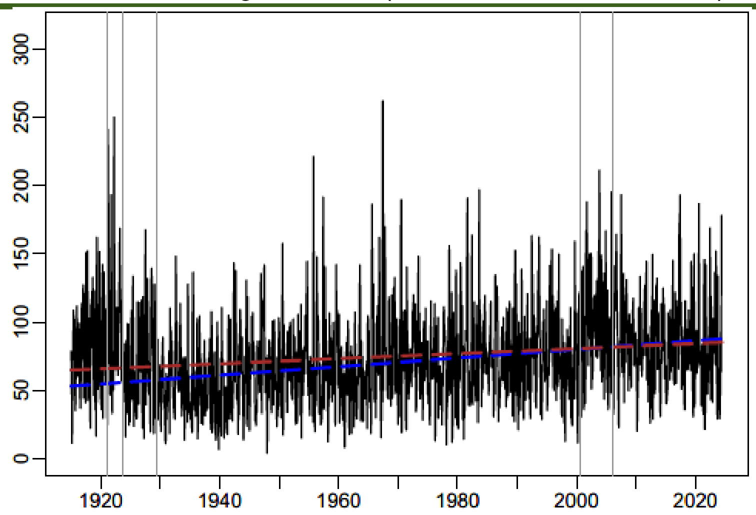
Homogenize untransformed precipitation series and enforce non-negativity in adjusted values

A pitfall to avoid: Do not homogenize logP and then back-transform to obtain HP (i.e., HP_fromHlogP).

HP (trend = -0.0186)



Raw/unhomogenized P (blue line is the HP trend)

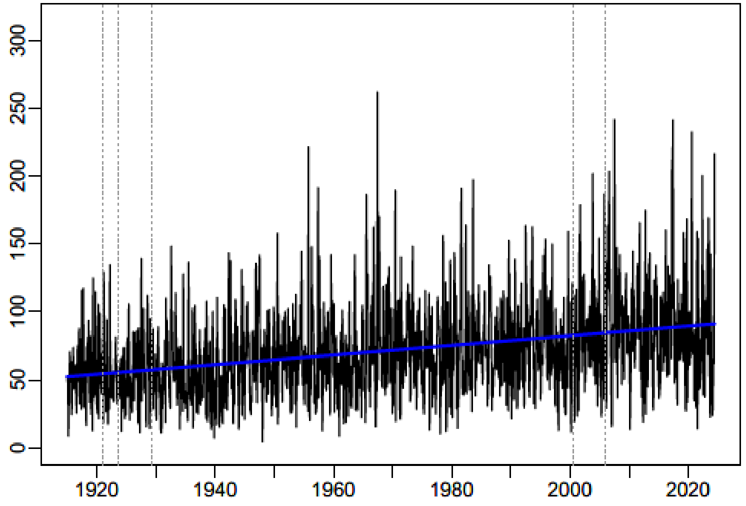


HP_fromLogP approach can introduce artificial variance inhomogeneity in adjusted series

- Occurs with QM or mean adjustments

  Example: QM adjustments -->

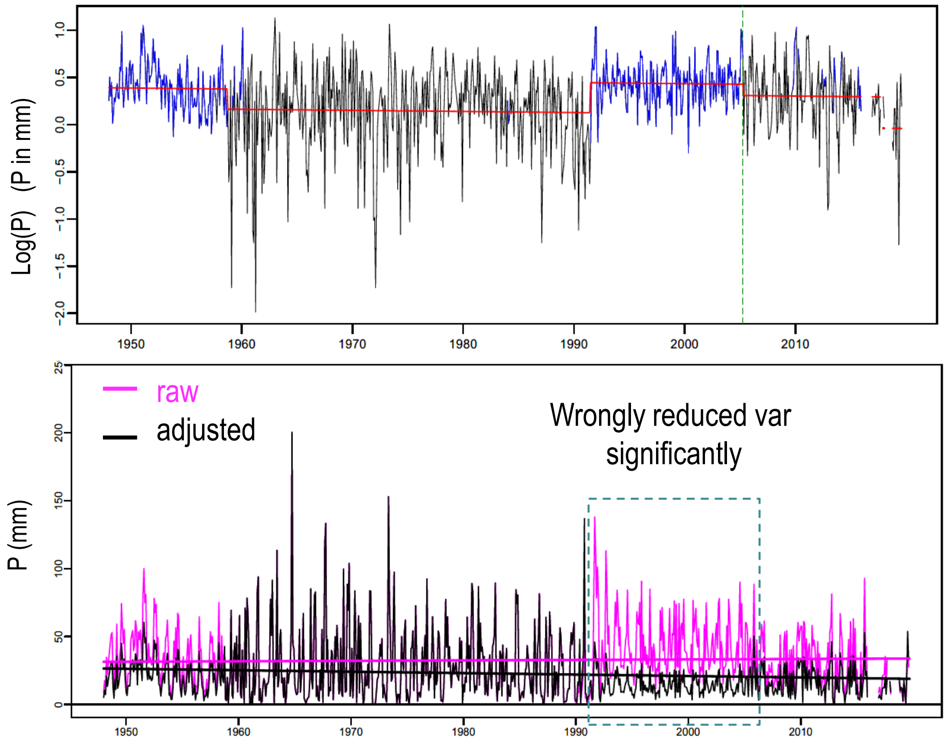
HP_fromHlogP (trend = -0.0296): the variance increase over time is not present in the raw series



Similarly, **do not apply mean-adjustments (Δ) to logP to obtain HP**

$$P_t^a = e^{\log(P_t)+\Delta} = e^{\Delta} P_t$$

- Var(P_t^a) is multiplied by $e^{2\Delta}$
  - $\Delta < 0$: variance decreases
  - $\Delta > 0$: variance increases
- This occurs regardless of any true variance change (including no change) associated with the mean shift



Homogenizing log-transformed ratios is mathematically equivalent to deriving HP via homogenization of logP (HP_fromHlogP) using a reference series:

$$\log(B_t/R_t) = \log(B_t) - \log(R_t)$$

This approach must be avoided (more on this to follow).

Homogenization of daily precipitation data

- More challenging than monthly precipitation due to frequency zeros and non-negative values.
- Adjustments must ensure that no negative values are introduced.
- Frequency discontinuities must be addressed before estimating and applying QM adjustments

Our approach (Wang & Feng, 2026; https.//doi.org/10.1016/j.wace.2026.100860)**:**

Replied on homogenized monthly P series and identified changepoints to guide daily homogenization. Specifically,

RHtests_dlyPrcp (https://github.com/ECCC-CDAS) used to

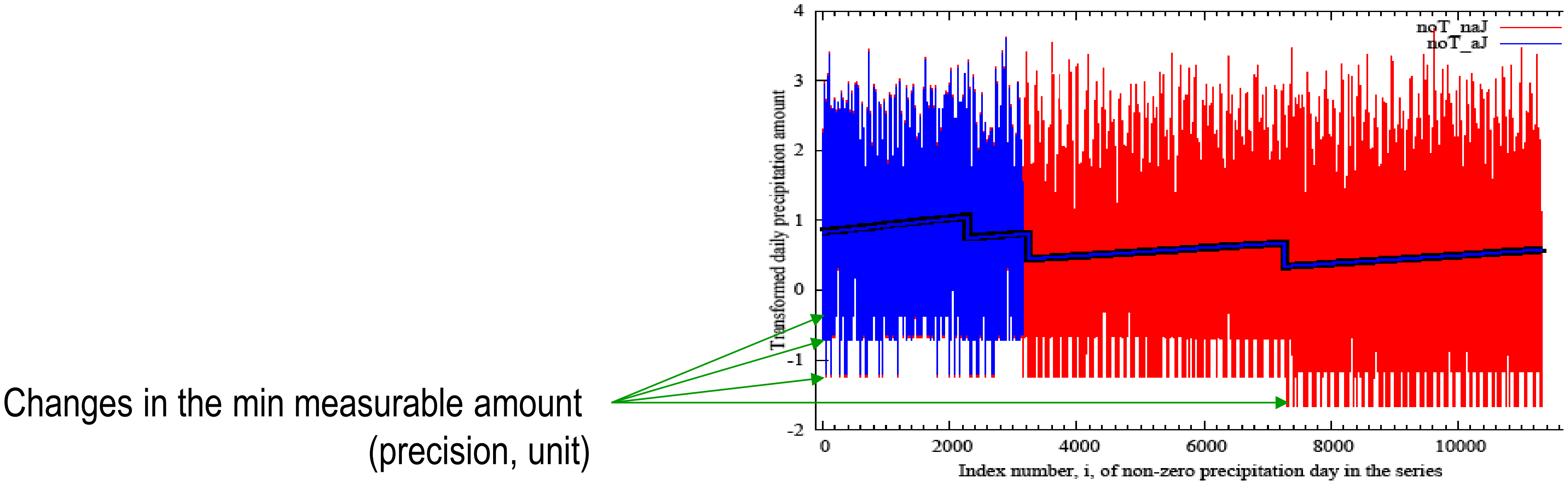
- Identify the most likely day of change within months with significant monthly changepoints.
- Estimate daily adjustment proportions so that adjusted daily values sum to the adjusted monthly totals

Missing daily precipitation values are infilled using ANUSPLIN, an advanced spatial interpolation method applied to a substantially larger dataset

Frequency discontinuity

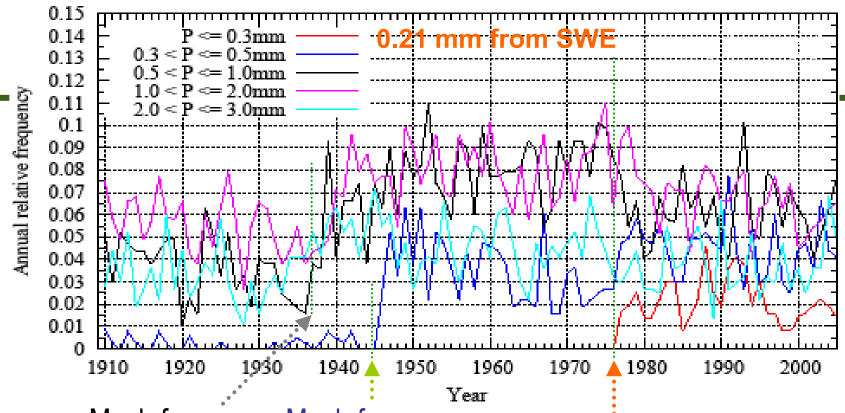
- Precipitation is non-continuous variable and may exhibit **frequency discontinuities**
- These must be addressed before applying QM adjustments
- Otherwise, frequency mismatch will be introduced in the adjusted data series

Example: frequency discontinuity caused by a change in measurement precision:



Ratio-based adjustments failed to homogenize the series (see next slide)

Discontinuities occur primarily in small precipitation (P ≤ 3 mm), especially in the frequency of occurrence:



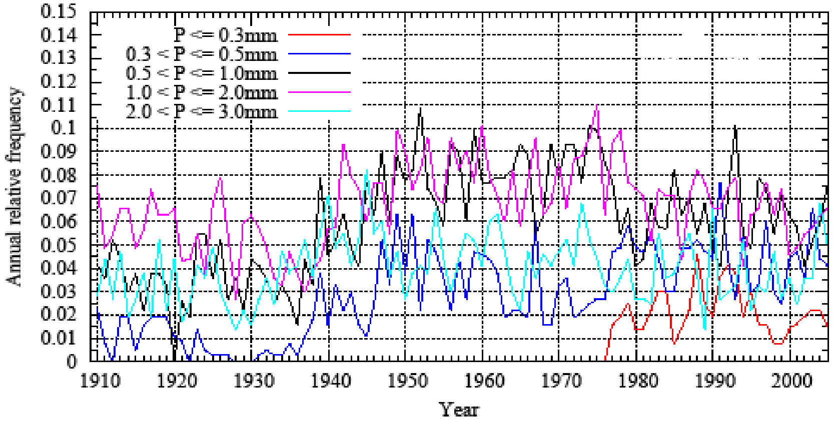
Much fewer 0.5 ~ 3 mm until 1937

Much fewer 0.3 ~ 0.5 mm until 1945 - joining

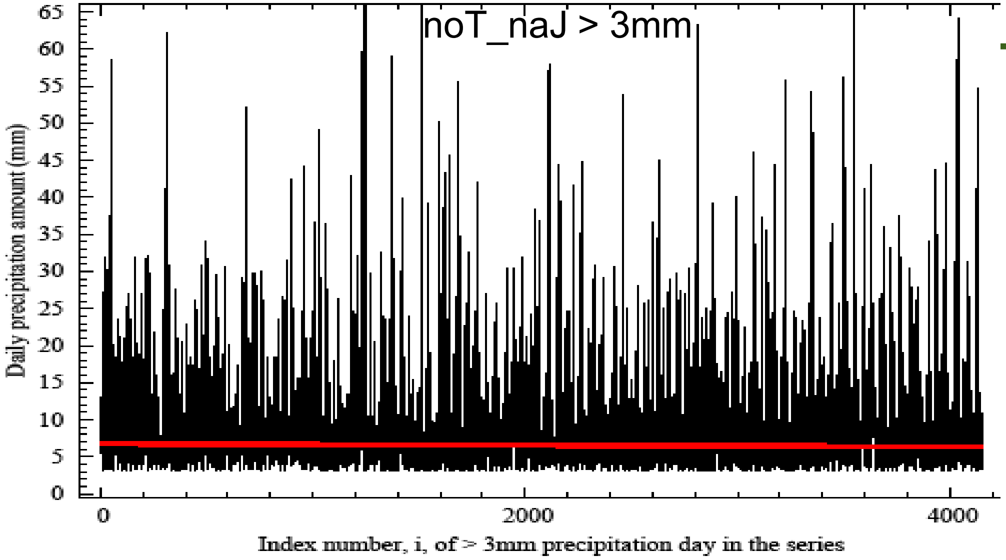
No P < 0.3 mm until 1976

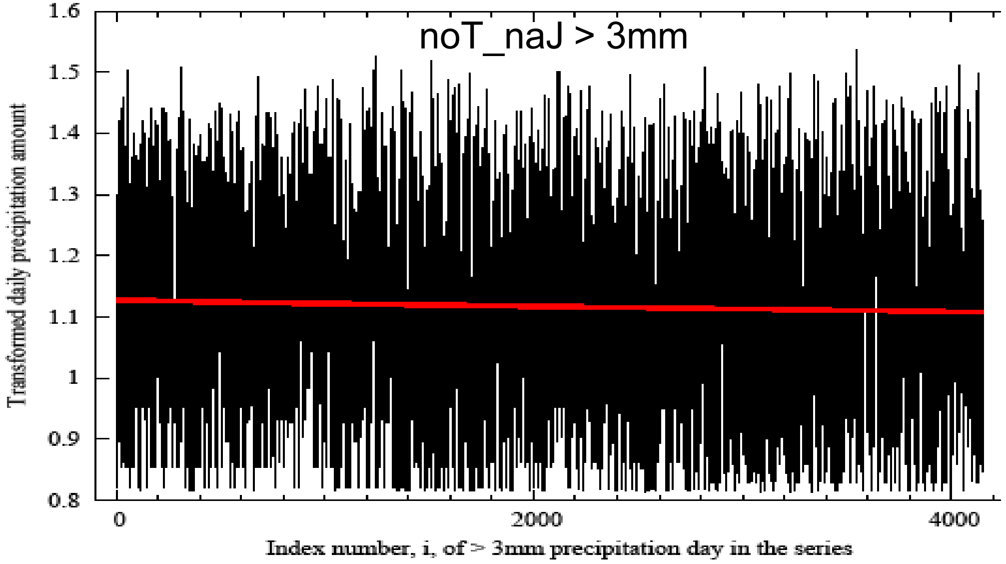
Ratio-based adjustments are **inappropriate** in this case, as they disproportionately adjust larger precipitation values that do not require adjustment

The above frequency discontinuities largely remain in the adjusted series:



Series of daily P > 3 mm – homogeneous!





Thank you very much for your attention!

Questions/comments?

Xiaolan L. Wang