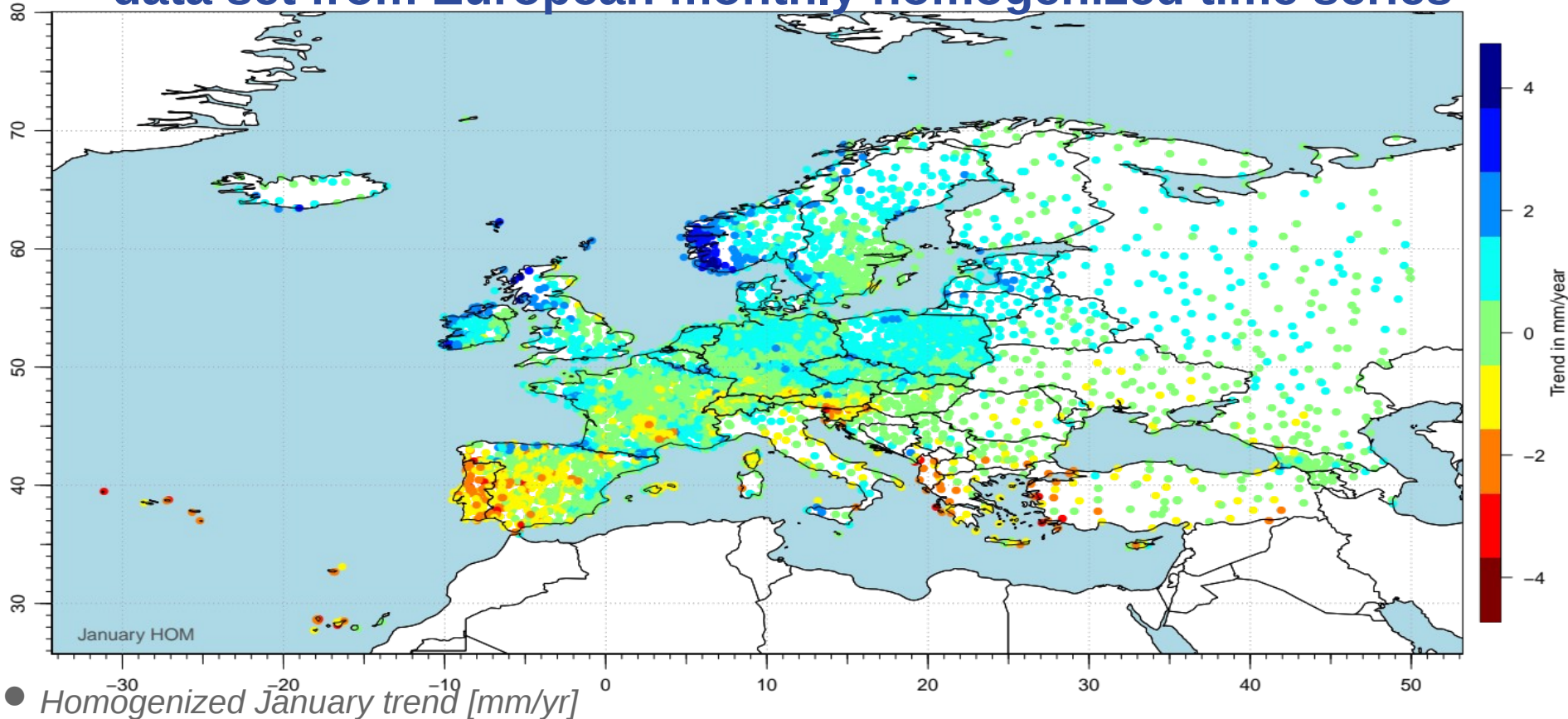


# HOMPRA Europe 2 – An update of a gridded precipitation data set from European monthly homogenized time series

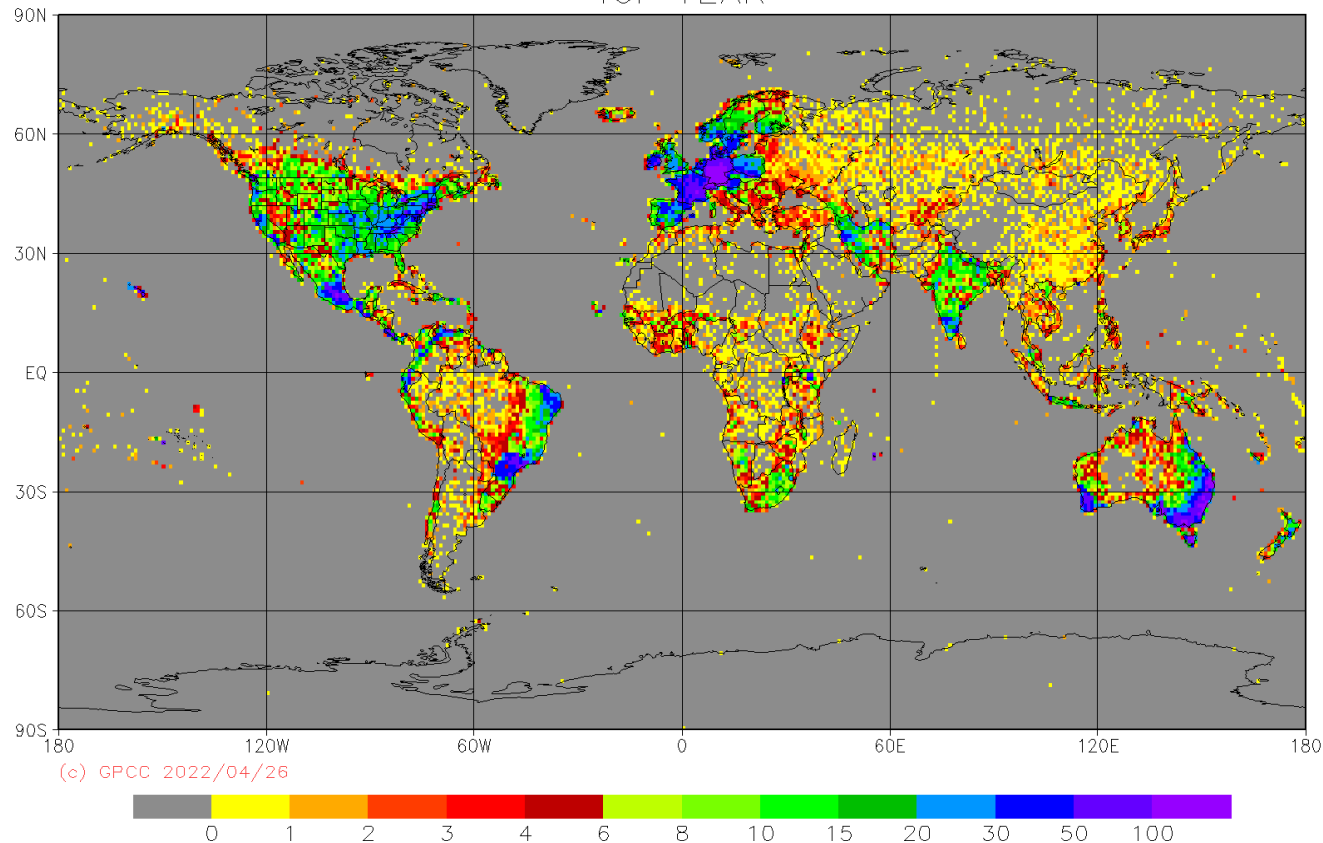


Elke Rustemeier, Peter Finger, Udo Schneider, Markus Ziese, and Stephanie Hänsel

Global Precipitation Climatology Centre, Deutscher Wetterdienst, Offenbach am Main, Germany

# Global precipitation Climatology Centre

GPCC Precipitation Normals (Version 2022)  
number of stations per 1.0 degree grid  
for YEAR



# Outline

- ➔ Data base
  
- ➔ Actual homogenization
  - ➔ Networks of similar time series
  - ➔ Detection of break-points
  - ➔ Correction of breaks

## Carefully chosen subset

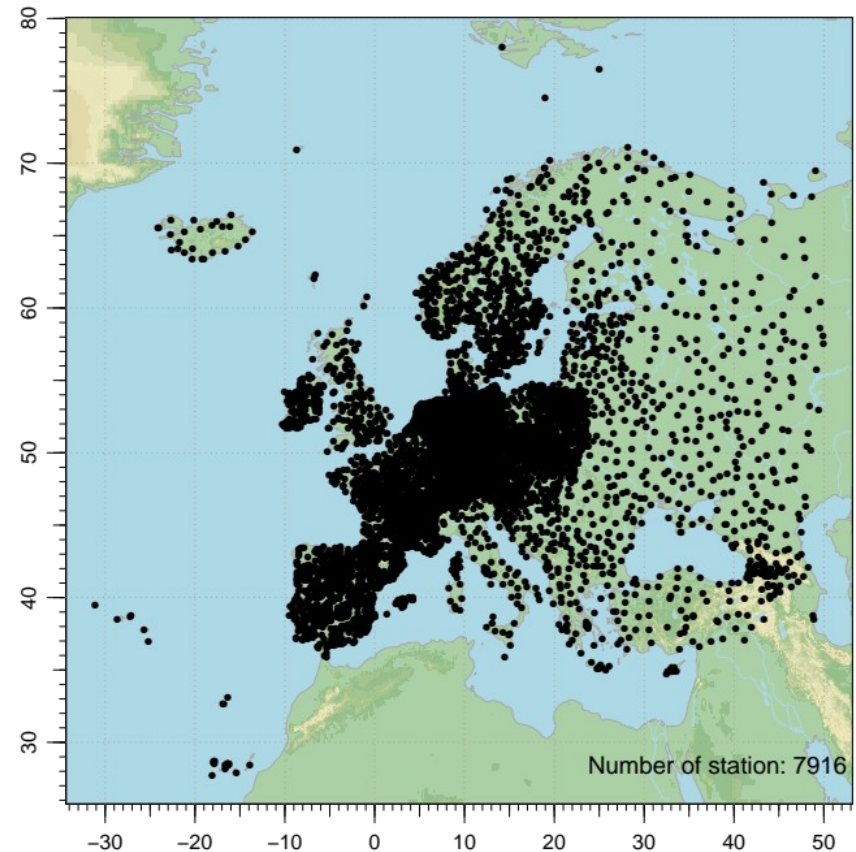
Monthly time series

Quality controlled data

Time periode 1951 – 2015

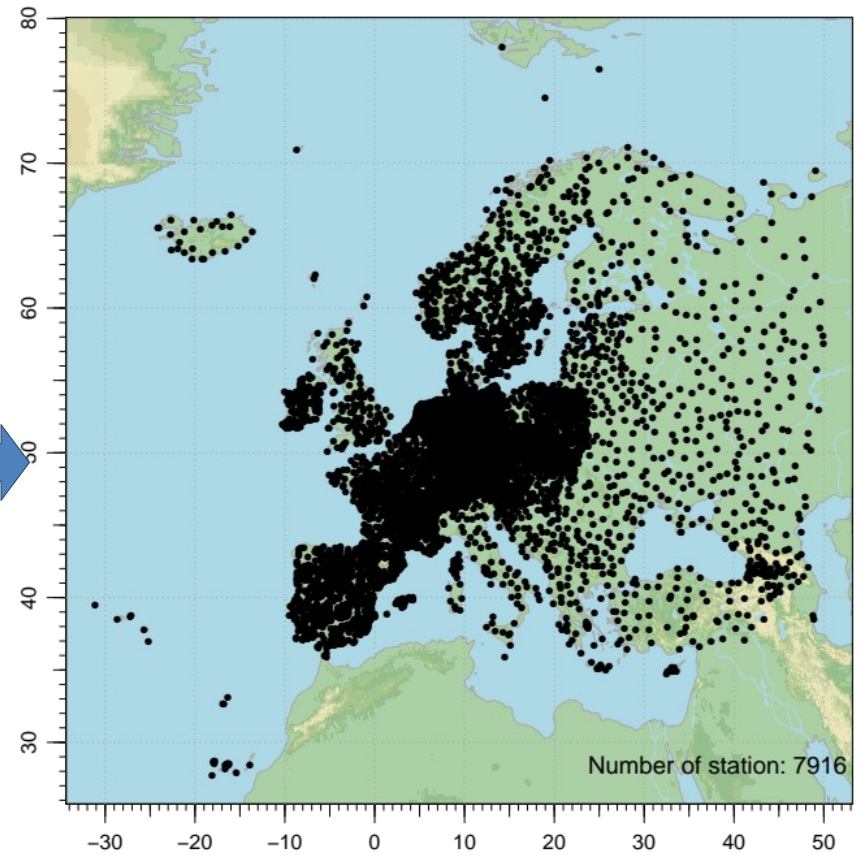
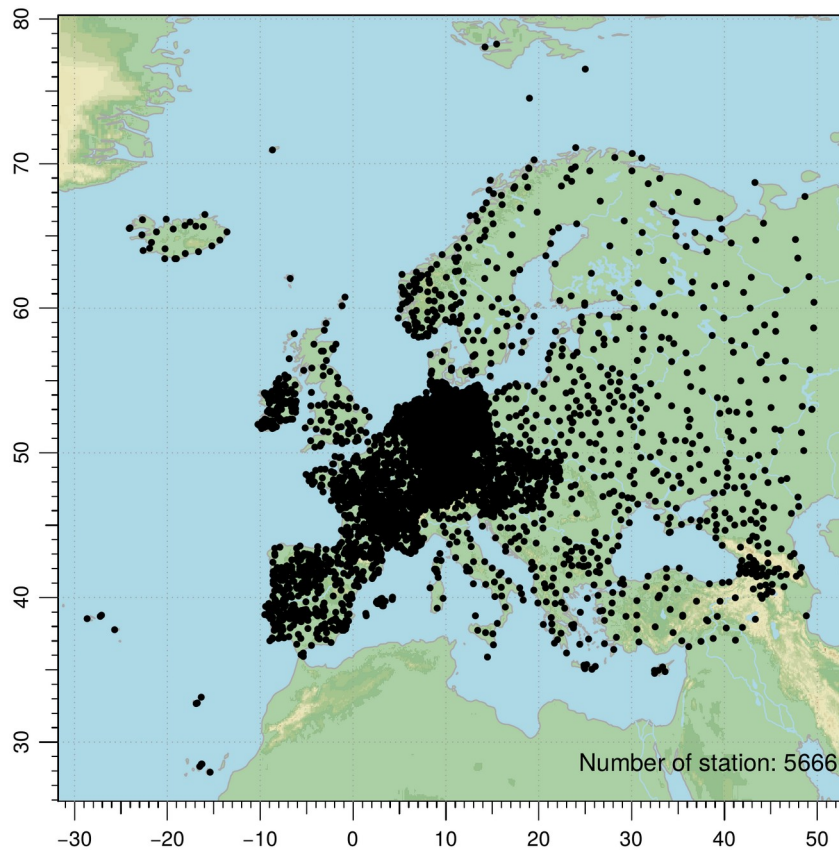
Not more than 20% of  
missing data

7916 time series

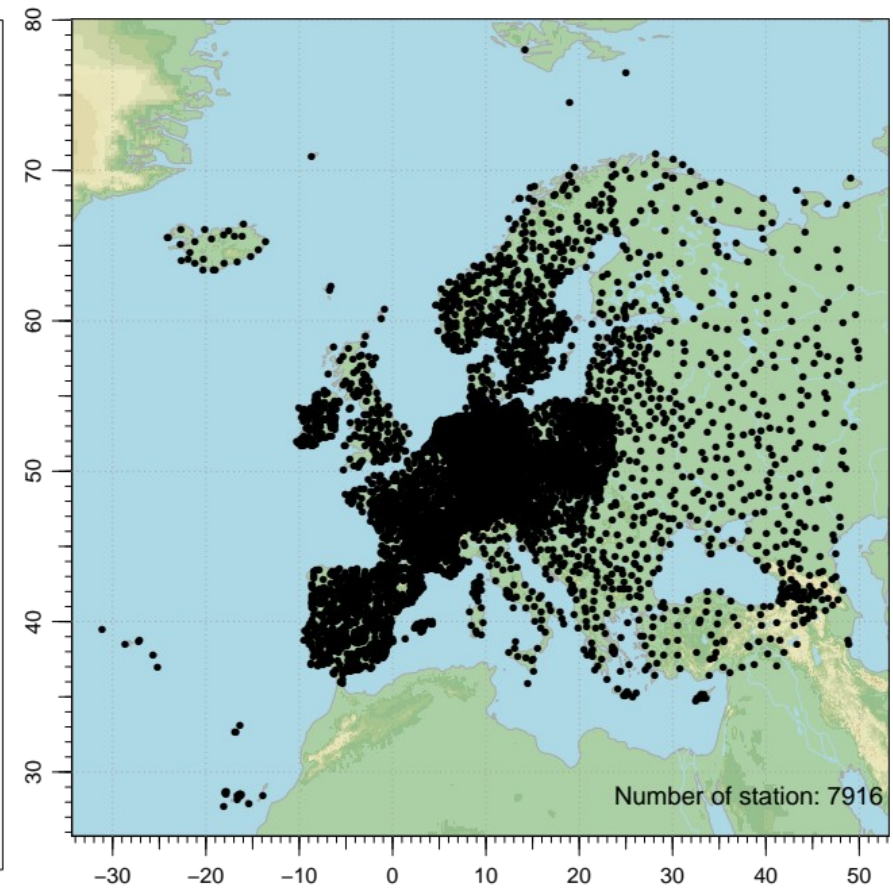
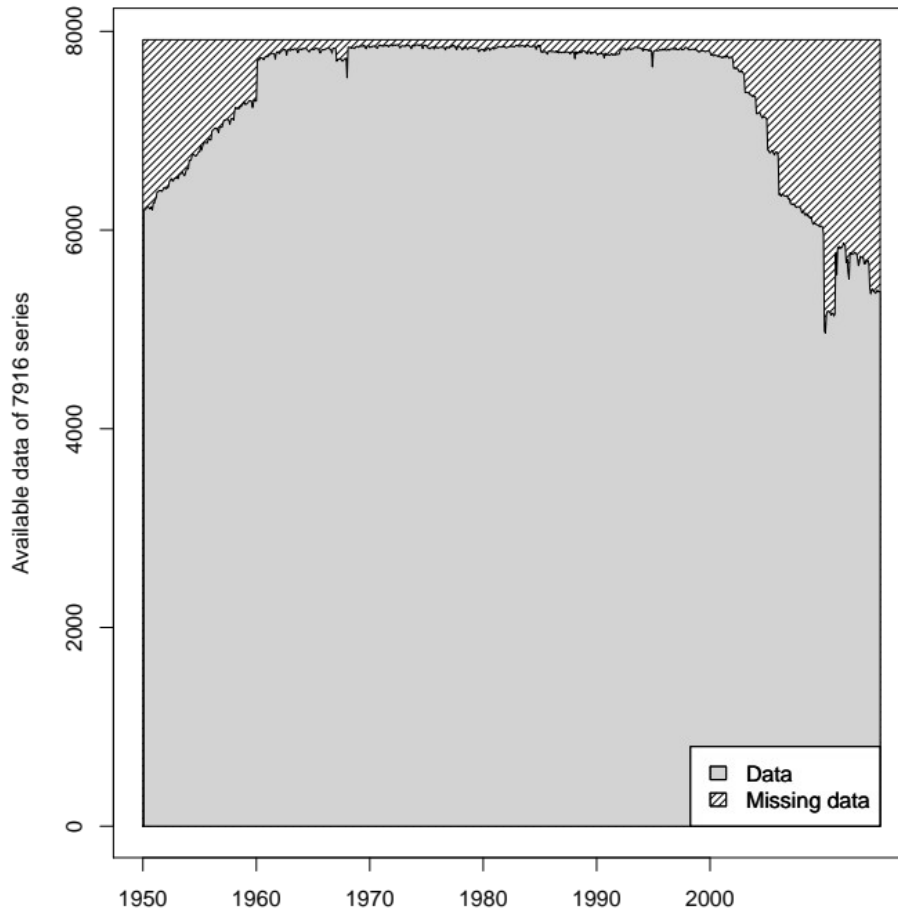




# Carefully chosen subset



# Carefully chosen subset



## Causes of inhomogeneity

The wind-induced error, which can be on average **2%-10% for rain** and **10%-50% for snow**, is the most important of systematic error. (Sevruk, 1985)

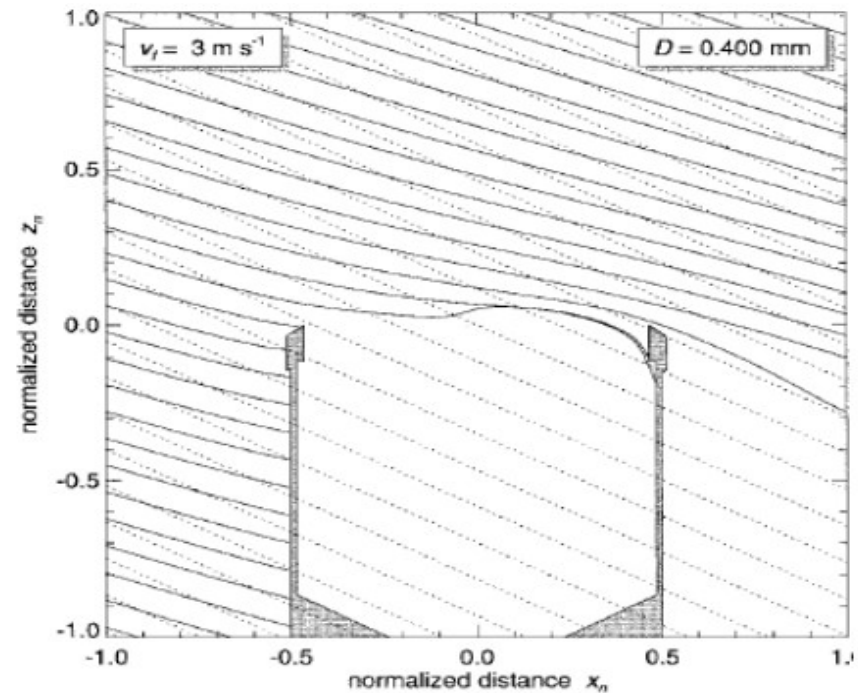


Figure: Computed trajectories of water drops (Nešpor and Sevruk, 1999)

# Causes of inhomogeneity

Table: Effect of modifications on different climate variables (Beaulieu, 2009)

Type of change	$T_{mean}$	$T_{min,max}$	$Prec_{tot}$	$Pres$	$Hum$	$v_{dir}$	$v_{force}$
Instrumentation	-	+	+	-	++	-	++
Instrument height	+	+	+	+++	+	+	++
Exposition	+	++	+++	-	+	++	++
Observation time	+	+	-	-	+	-	+
Calculation method	++	-	-	-	+	-	+
Relocation	++	+++	+++	-	+	++	+++
Environment	+	+	+	-	+	+	+
Observer	-	-	-	-	-	-	+

-	no major problems	$T_{mean}$	Average temperature	$Pres$	Air pressure
+	some inhomogeneities	$T_{max}$	Minimum temperature	$Hum$	Humidity
++	important breaks	$T_{min}$	Maximum temperature	$v_{dir}$	Wind direction
+++	very important breaks	$Prec_{tot}$	Precipitation	$v_{force}$	Wind force



# Causes of inhomogeneity

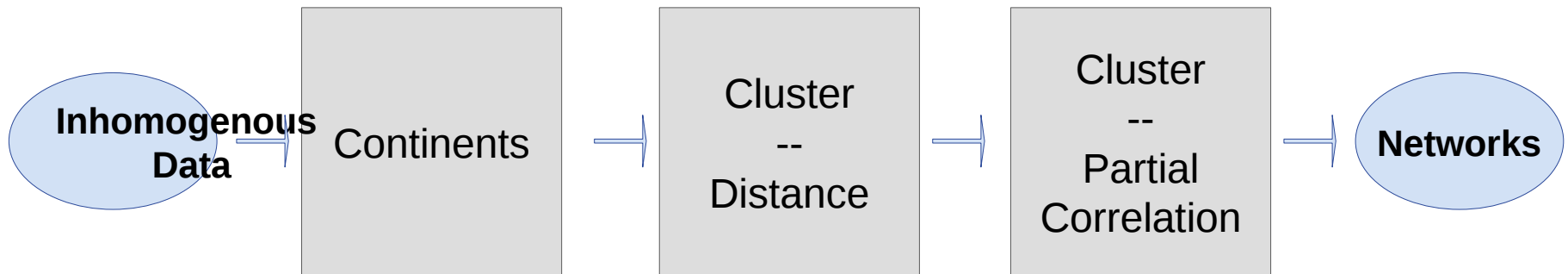
Table: Effect of modifications on different climate variables (Beaulieu, 2009)

Type of change	$T_{mean}$	$T_{min,max}$	$Prec_{tot}$	$Pres$	$Hum$	$v_{dir}$	$v_{force}$
Instrumentation	-	+	+	-	++	-	++
Instrument height	+	+	+	+++	+	+	++
Exposition	+	++	+++	-	+	++	++
Observation time	+	+	-	-	+	-	+
Calculation method	++	-	-	-	+	-	+
Relocation	++	+++	+++	-	+	++	+++
Environment	+	+	+	-	+	+	+
Observer	-	-	-	-	-	-	+

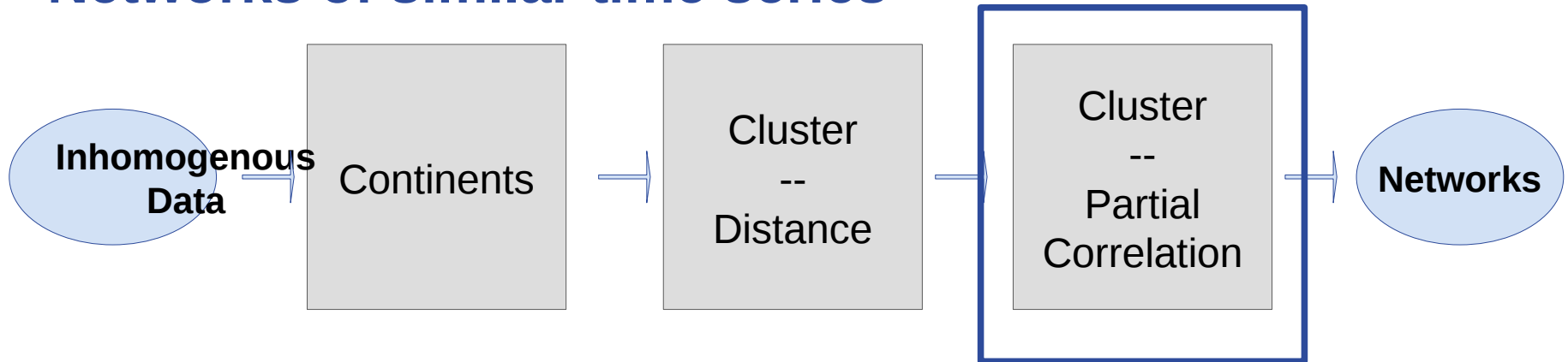
-	no major problems	$T_{mean}$	Average temperature	$Pres$	Air pressure
+	some inhomogeneities	$T_{max}$	Minimum temperature	$Hum$	Humidity
++	important breaks	$T_{min}$	Maximum temperature	$v_{dir}$	Wind direction
+++	very important breaks	$Prec_{tot}$	Precipitation	$v_{force}$	Wind force

## Networks of similar time series

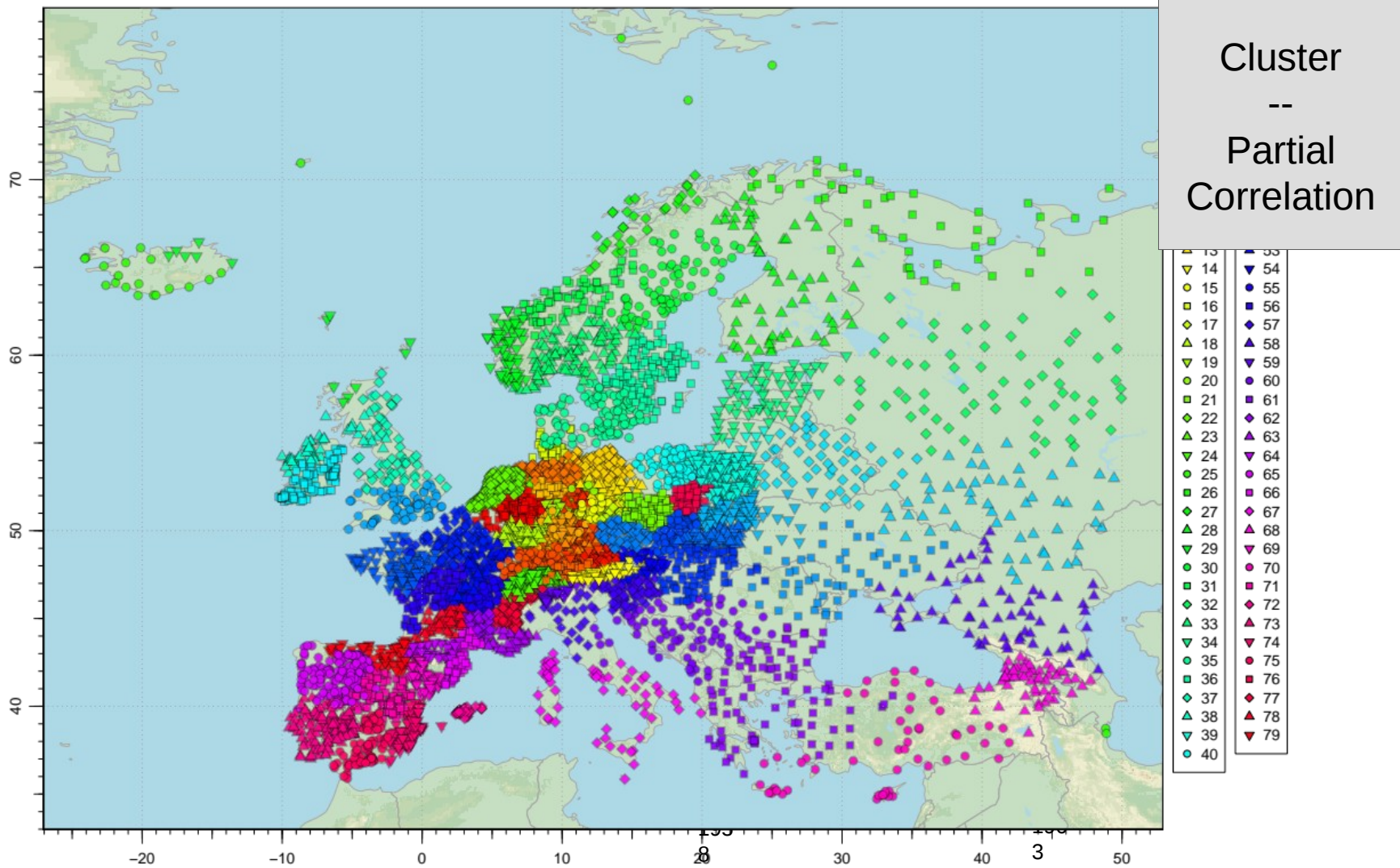


- Calculate **partial correlation** between the series
- (Minimal least squares)
  - Consecutive differences
  - Removal of the annual cycle
  - Calculation of the ranks
- **WARD CLUSTER**

## Networks of similar time series

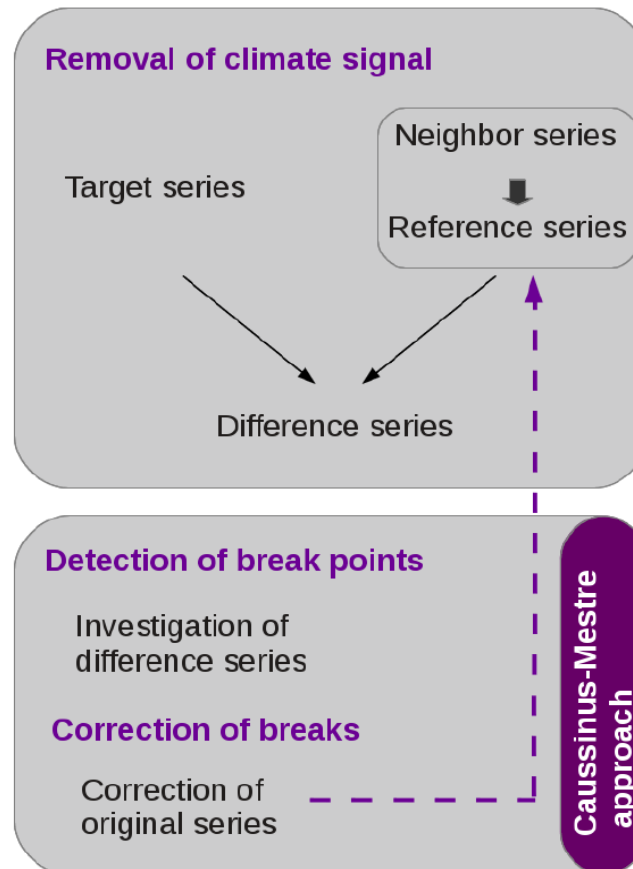


- Calculate **partial correlation** between the series
- (Minimal least squares)
  - Consecutive differences
  - Removal of the annual cycle
  - Calculation of the ranks
- **WARD CLUSTER**





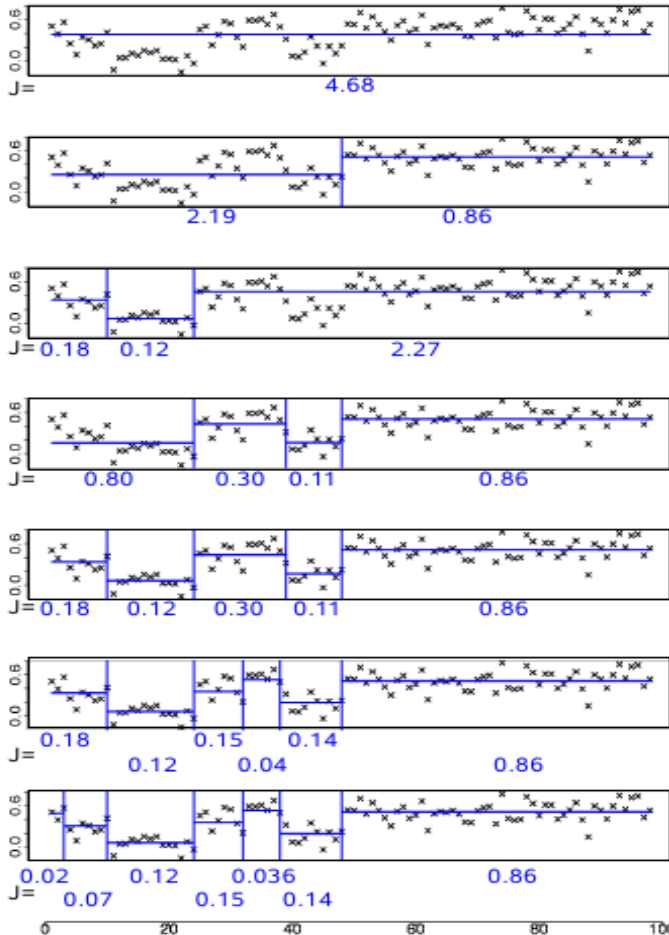
# Homogenization course



# Homogenization course: Break-point detection

Transformed time-series – transformed reference time-series

## Time-series – Reference



## Difference time-series

- Transformed time series – transformed reference time-series

## Log-likelihood

- Best break-point position for each number of breaks

## Penalty term

- Number of breaks

See **Caussinus and Mestre, 2004**

# Homogenization course: Correction

## → Box-Cox Transformation

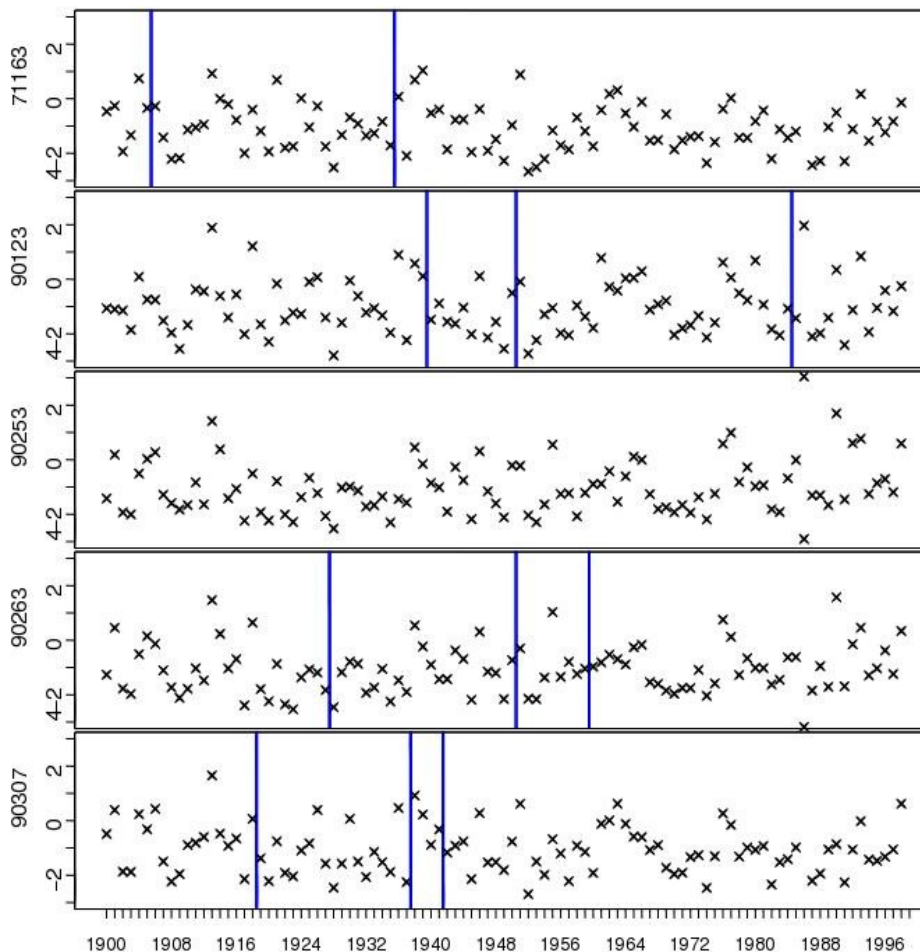
(Software requires normal distribution)

$$\rightarrow Y_{\text{new}} = \begin{cases} ((Y_t)^k - 1) / k & \text{for } k \neq 0.000 \\ \ln(Y_t) & \text{for } k = 0.000 \end{cases}$$

## → Reference series

– High correlated time series

# Homogenization course: Break correction

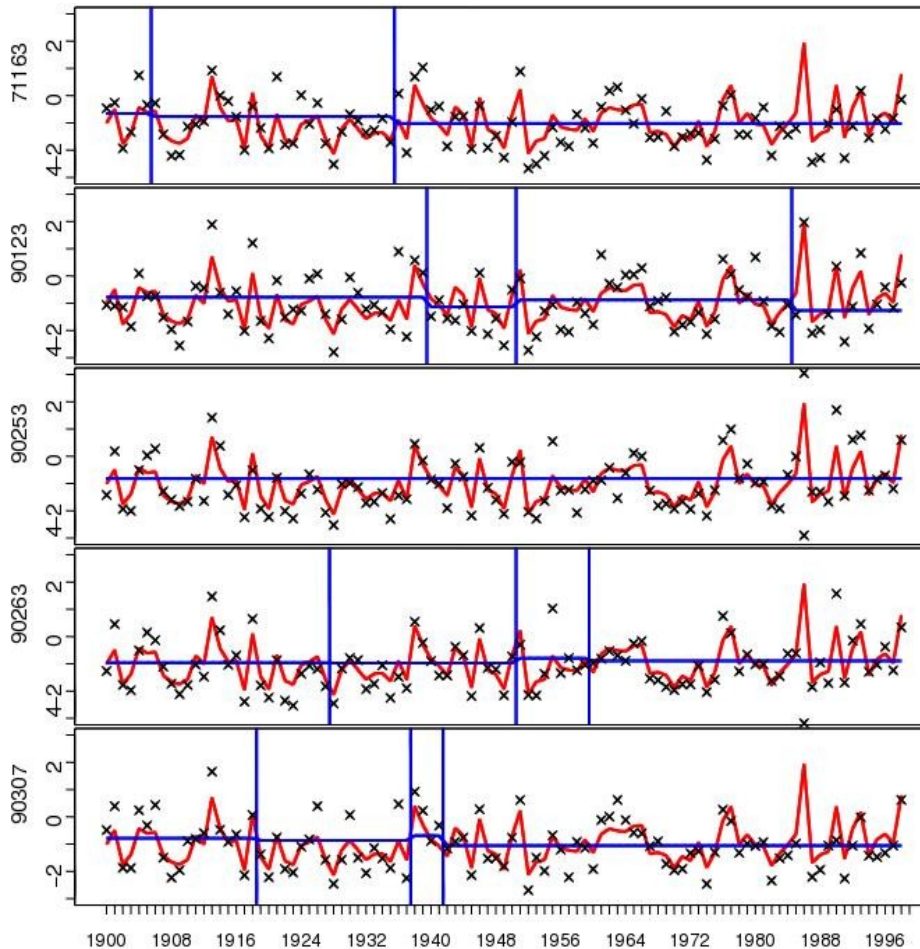


- 1) Binary coding of the series
  - 2) **Multiple linear regression** over homogeneous segments
  - 3) **Regression coefficients** indicate **break amplitude**
- See **Mestre (2003)**

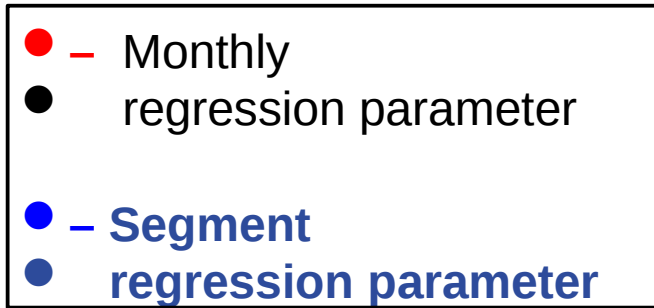
- x Power transformed monthly time-series
- | Detected breaks



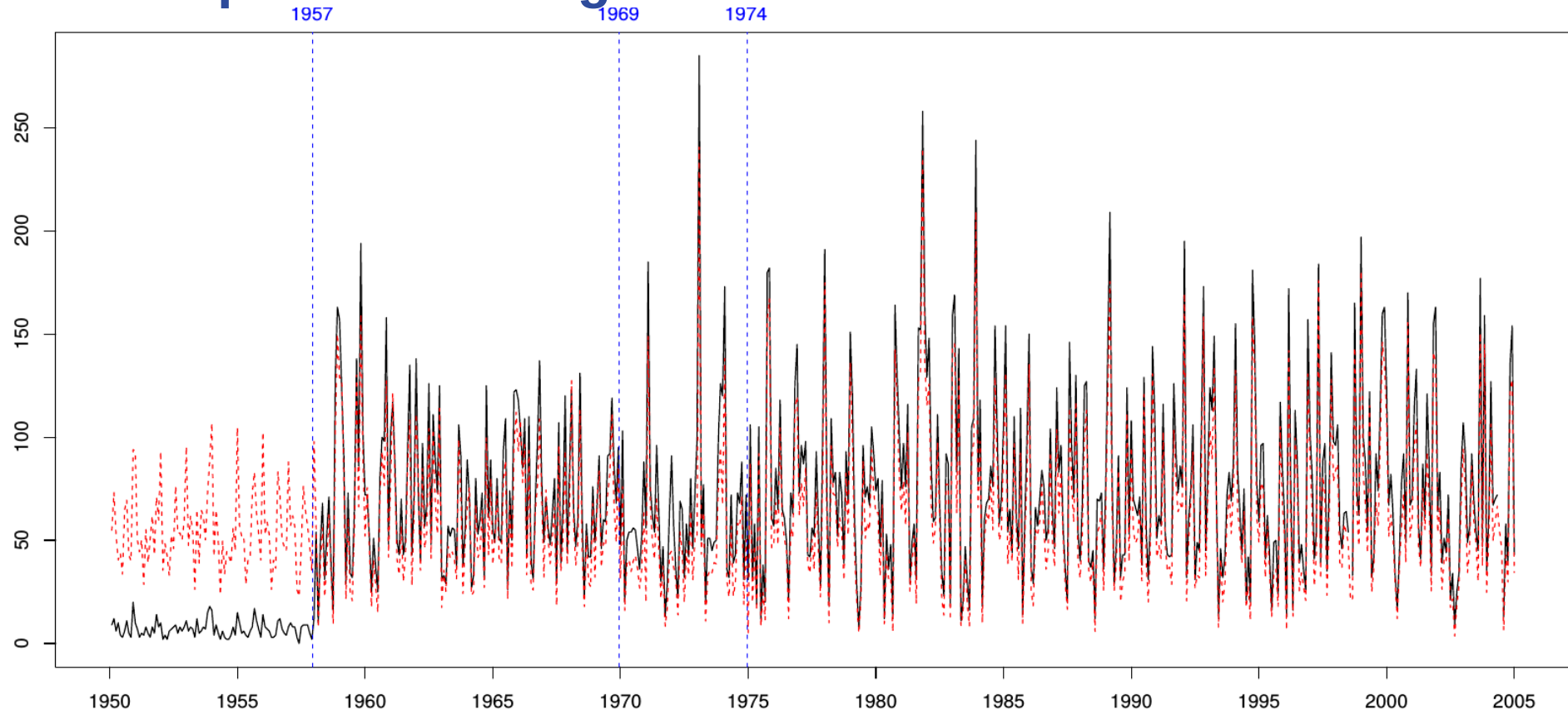
# Homogenization course: Break correction



- 1) Binary coding of the series
  - 2) **Multiple linear regression** over homogeneous segments
  - 3) **Regression coefficients** indicate **break amplitude**
- See **Mestre (2003)**



# Example of a homogenized time-series

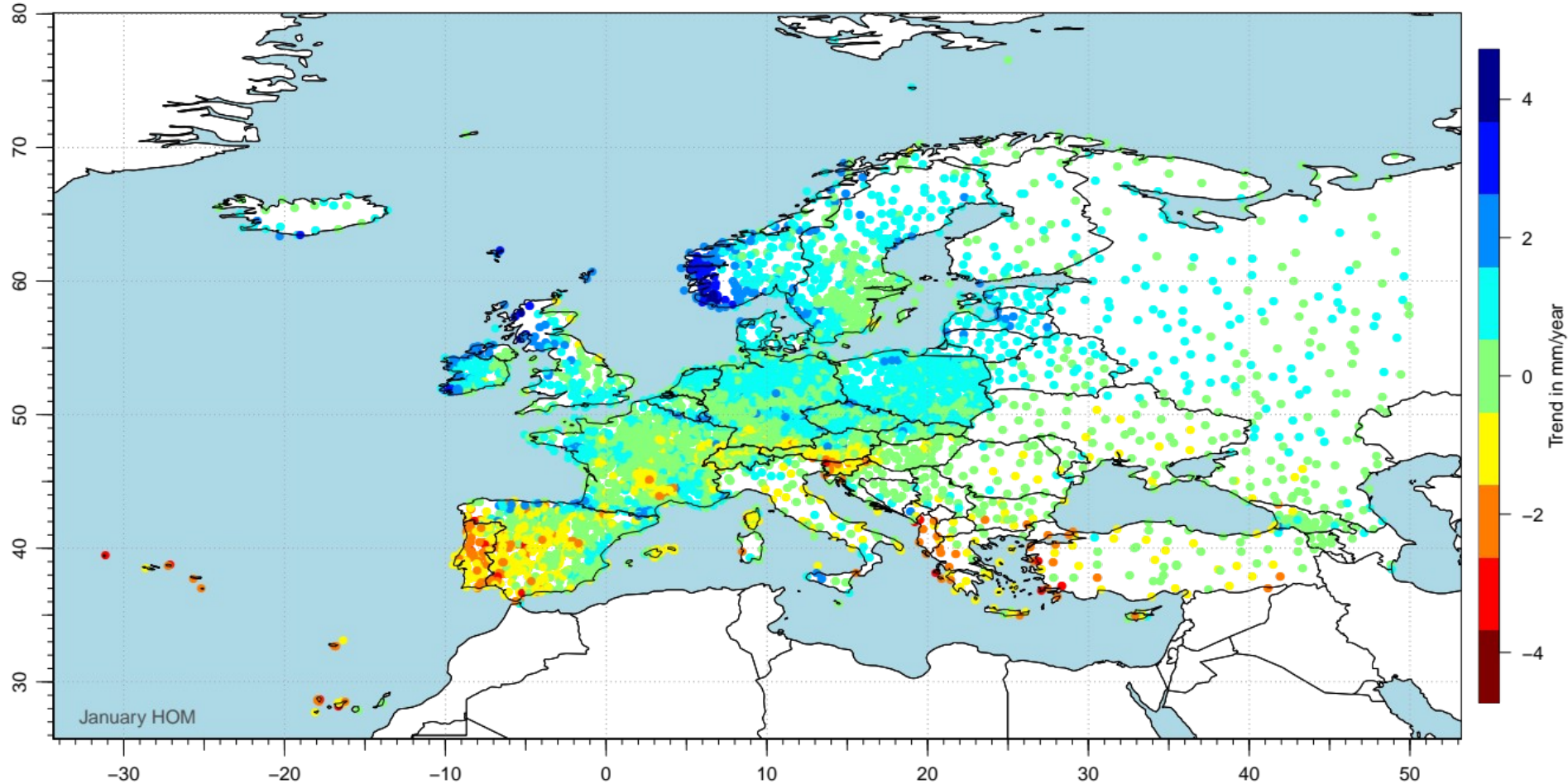


Inhomogenous  
Data

Networks

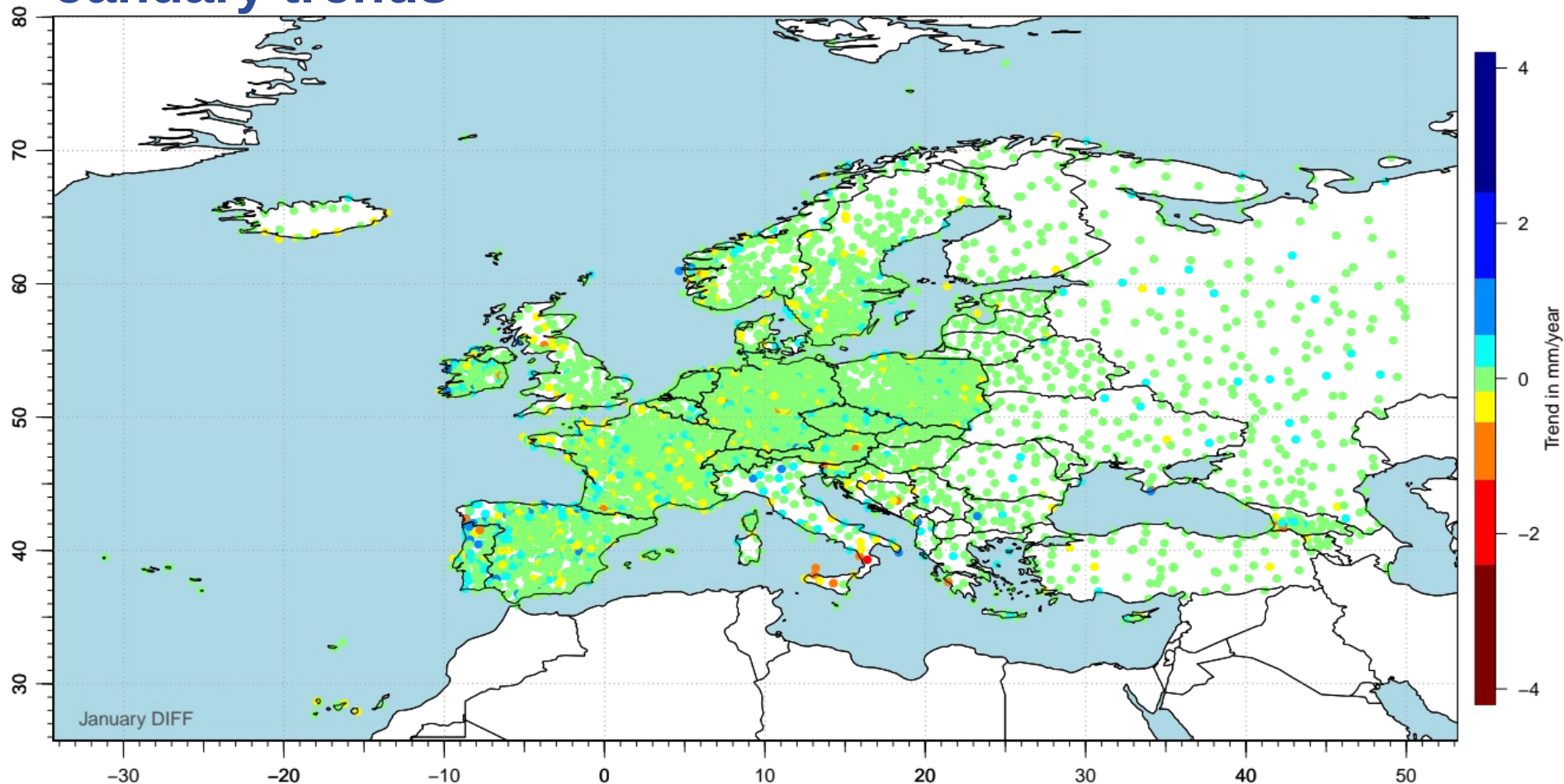
Homogenized  
Data

# Homogenized January trends

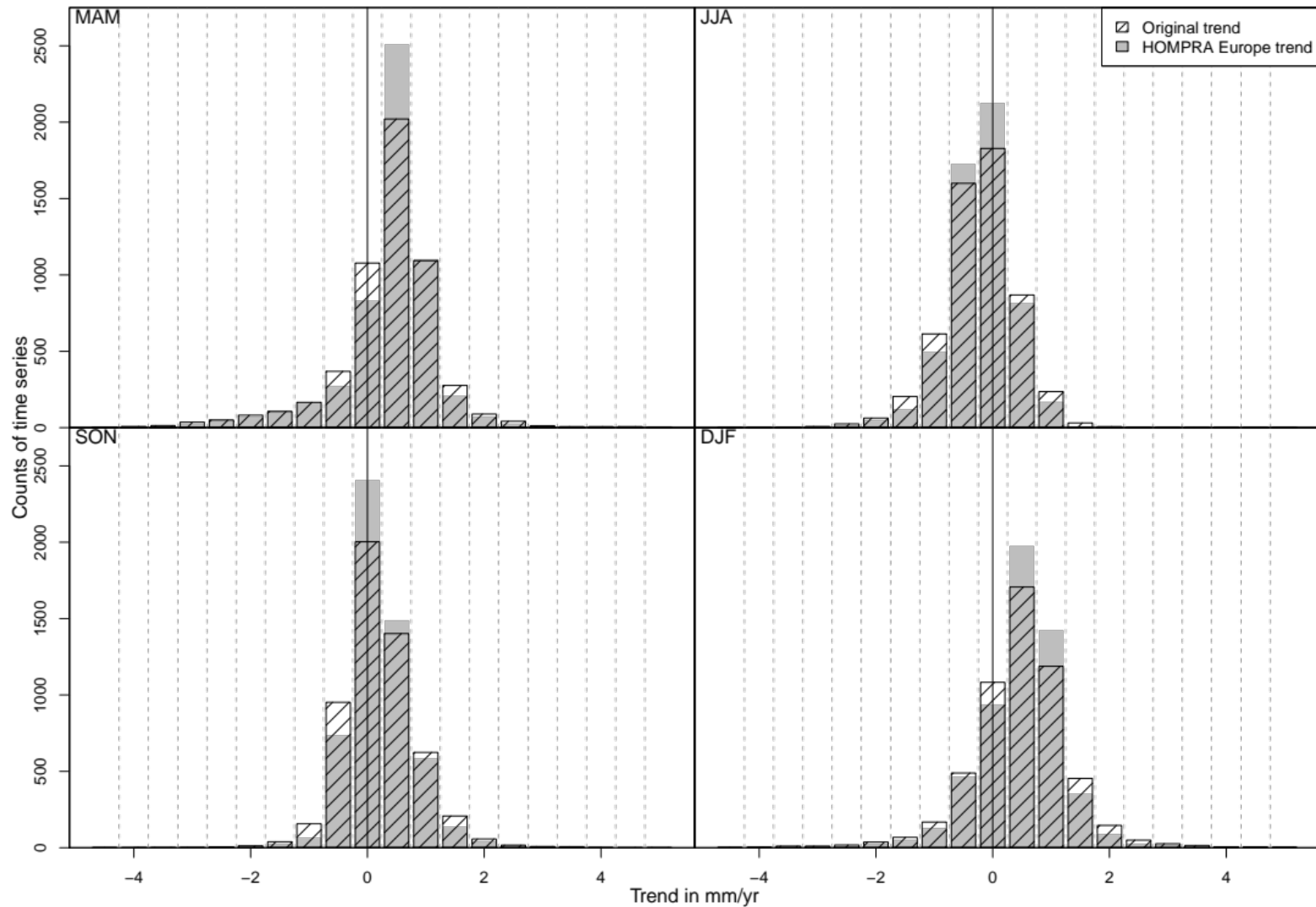




# Difference between raw and homogenized January trends



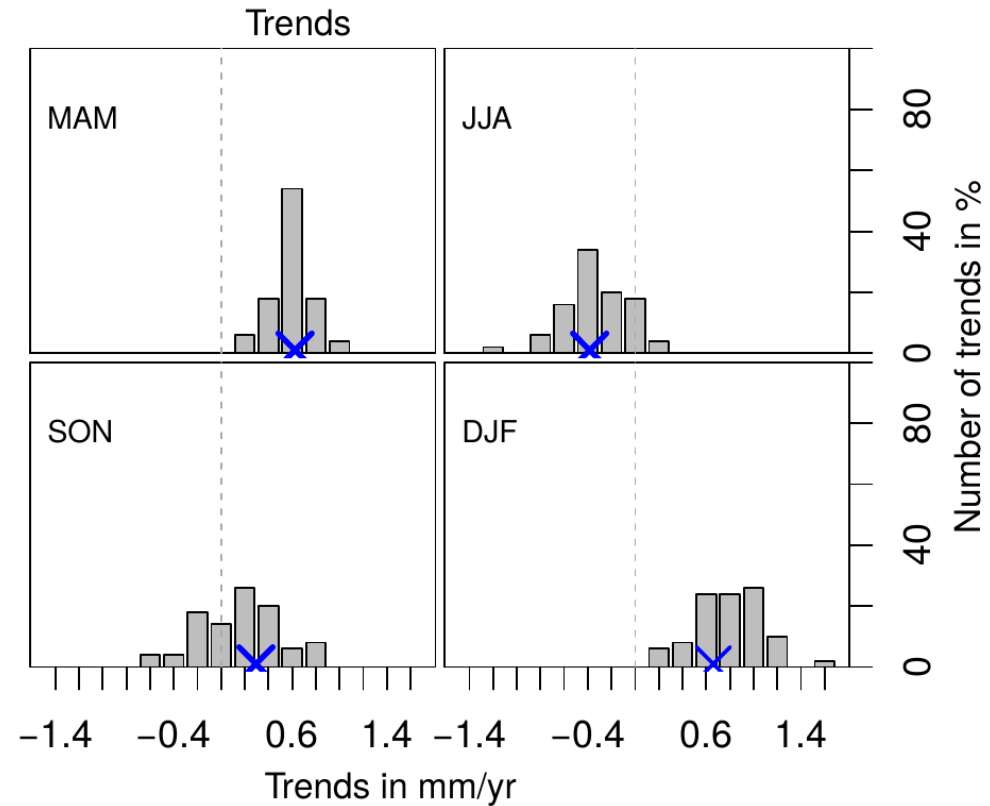
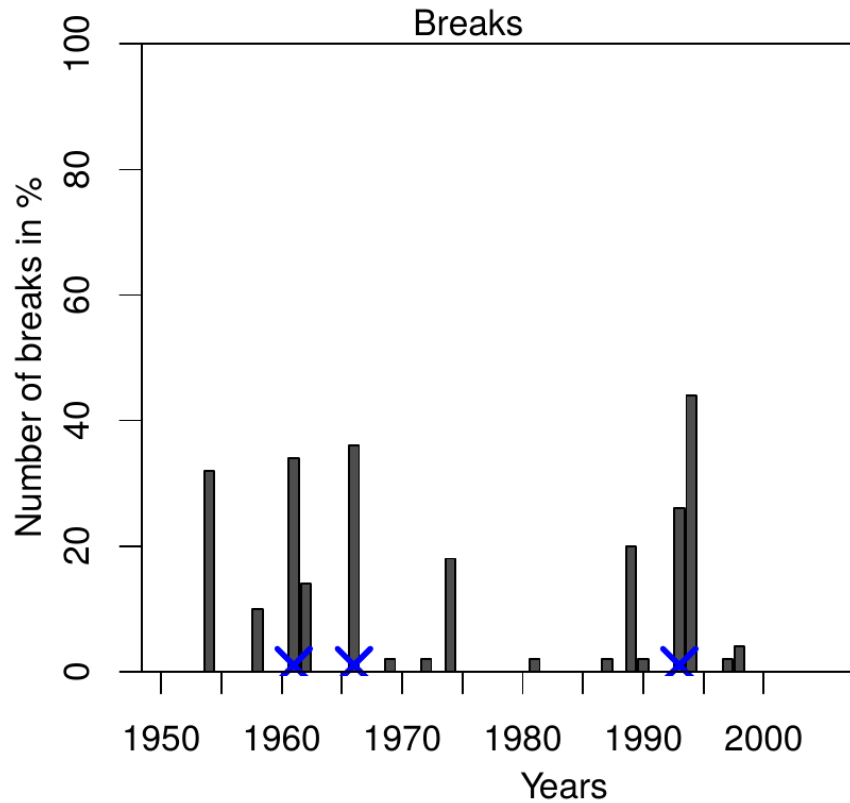




# Reporting back to quality control

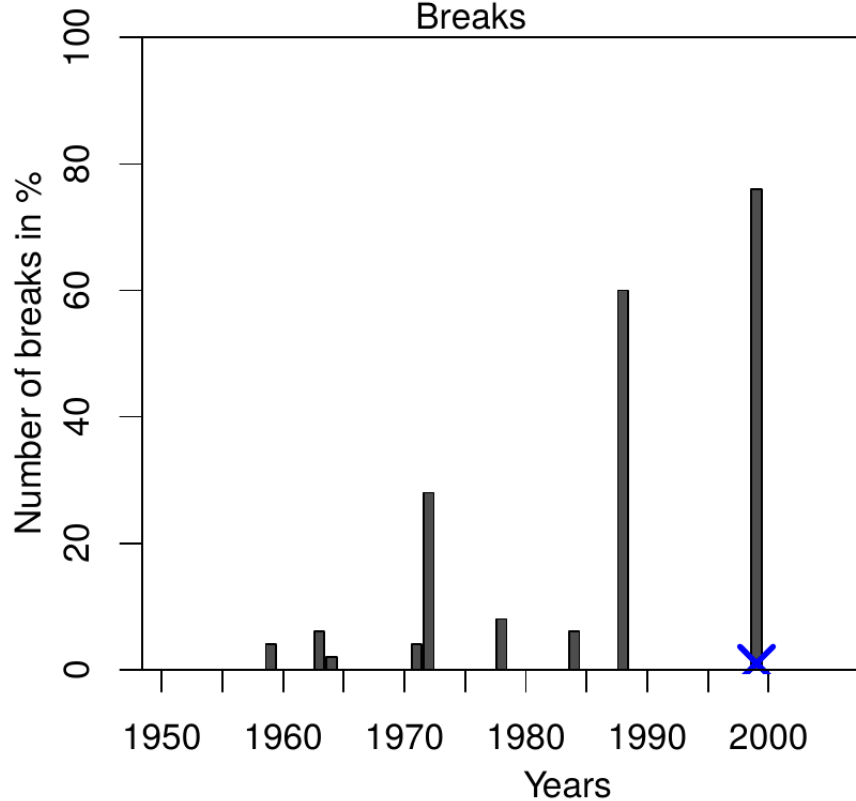
- **Correlation** between series **too high**
  - Duplicate stations
- **High correction factor**
  - May be factor 10 error
- **Too many zeros** compared to neighbor series

## 00000963 Diepholz

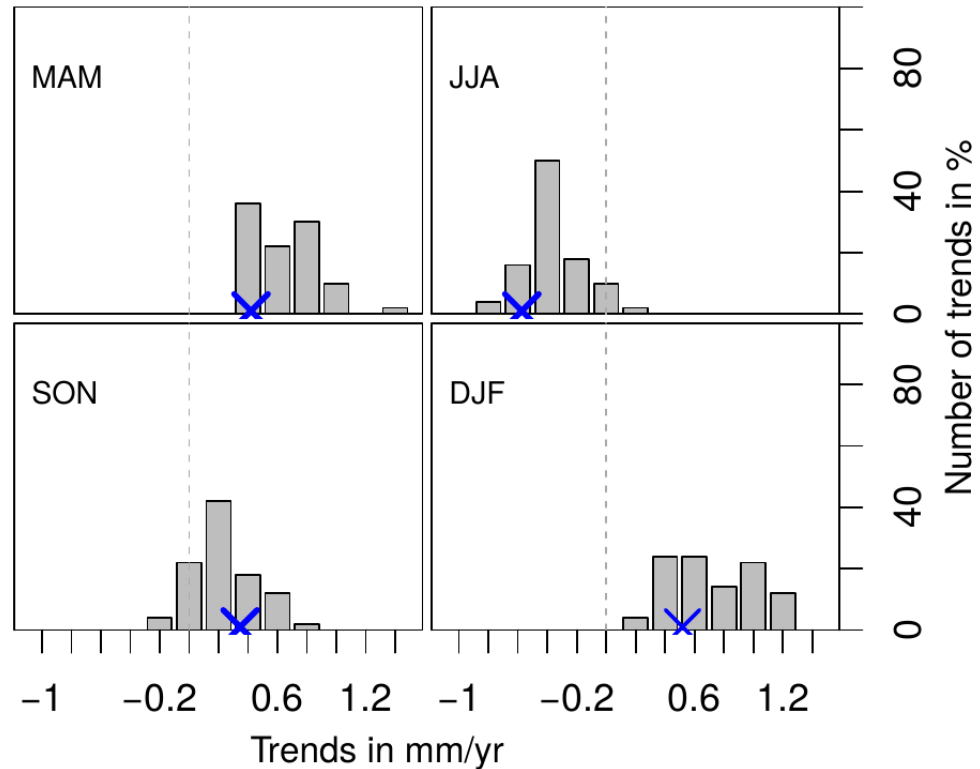


## 00002088 Heidenheim/Brenz

Breaks



Trends



# Verification

- **Especially important due to automation**

## Verification of the method

- Usually testing on independent data
  - **Artificial data**
- **Comparison with digitized meta data (Germany only)**

## Verification of the results

- **Suspicious series are controlled manually**
- Sensitivity study
  - **Variation of reference series**



# Todo list before publication of HOMPRA Europe 2

## Suspicious series are controlled manually

### Output for every time series

- **CRADDOCK test** on original and homogenized data
- Including neighbor series and detected breaks
- **Annual cycle** (including neighbor series )
- **Absolute** raw and corrected series
- **Relative** raw and corrected series

## Todo list before publication of HOMPRA Europe 2

### Blacklisting before interpolation

- Manually blacklisted series if necessary
  - **High correction factor**
  - Series without high **correlated** neighbors
  - Strong differences in the **annual cycle** between target and neighbor time series

# Interpolation

- **Modified SPHEREMAP**
- (Becker et al., 2013 and Schamm et al., 2014)
  - Distance and angle weighted, weighted average method
  - Applied on anomalies
  - One of the interpolation schemes that run operationally at the GPCC

## Summary and outlook

- Development of an **automatic algorithm**
  - Allows homogenization of large data sets
  - Homogenization of ~8000 monthly time series
  - Sensitivity study for the used neighbor series
  
- Comparison to HOMPRA Europe
- Publish HOMPRA Europe 2 ([gpcc.dwd.de](http://gpcc.dwd.de))
- Probably end of July

# Thank you for your attention!

## Homogenization

Start homogenization  
on networks

### Network I

#### Monthly totals

Missing values  
Box-Cox transformation  
(parallel)

- Target series and high correlated series
- Dummy codification
- Multiple linear regression

#### Annual totals

Detection of breakpoints  
Logarithmic transformation  
(parallel)

- Difference of target series and reference series
- Penalized log-likelihood
- Caussinus-Mestre criterion

#### Monthly totals

Correction of breaks  
Box-Cox transformation  
(parallel)

- Target series and high correlated series
- Dummy codification
- Multiple linear regression

### Network ...

### Sort into networks

#### Meta data

Sort stations to continents

#### Meta data

Ward cluster on  
Great circle distance

#### Monthly totals

Ward cluster on  
Partial correlation (parallel)