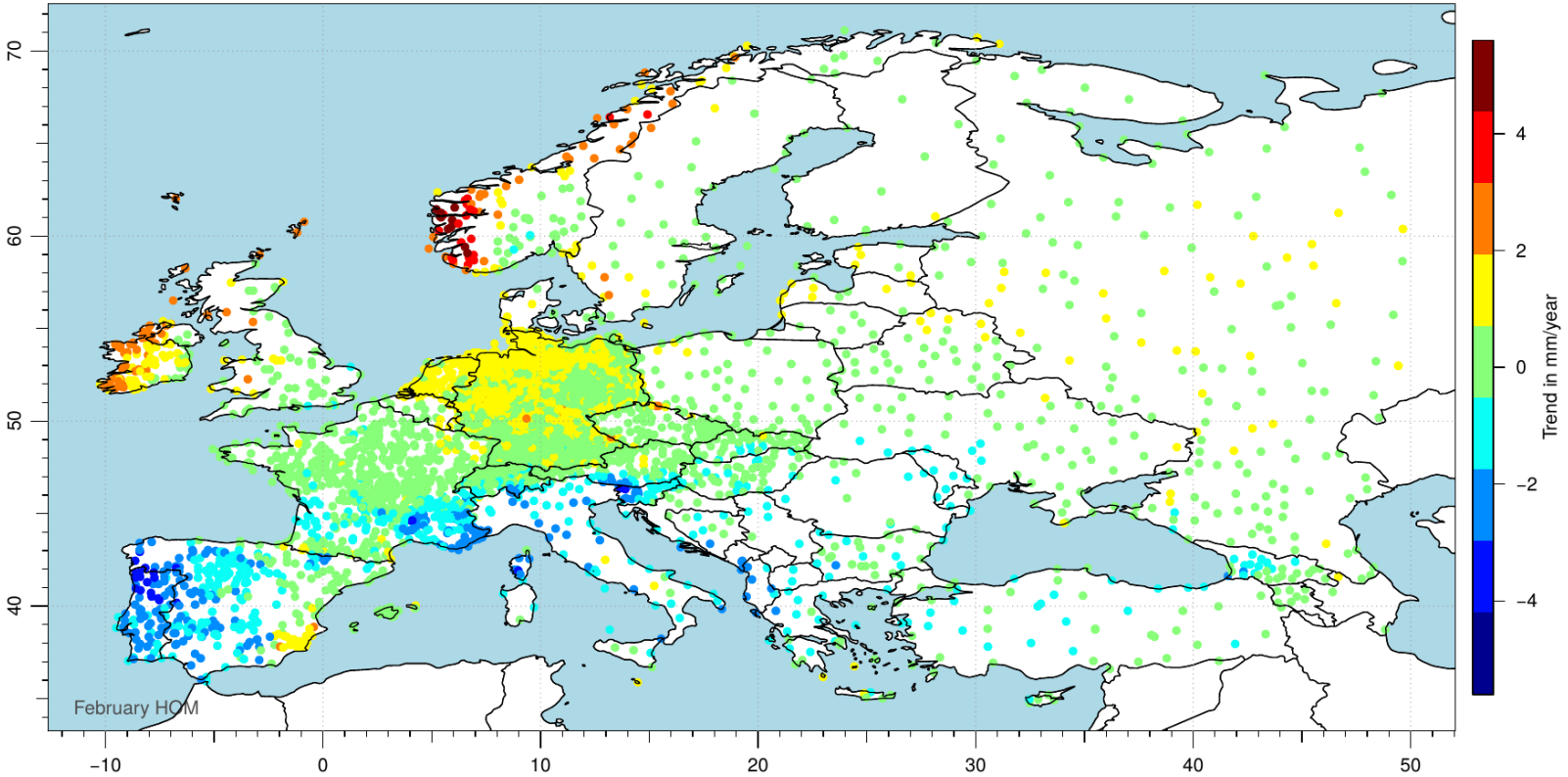
HOMPRA Europe – A gridded precipitation data set from European homogenized time series



Homogenized February trend [mm/yr]

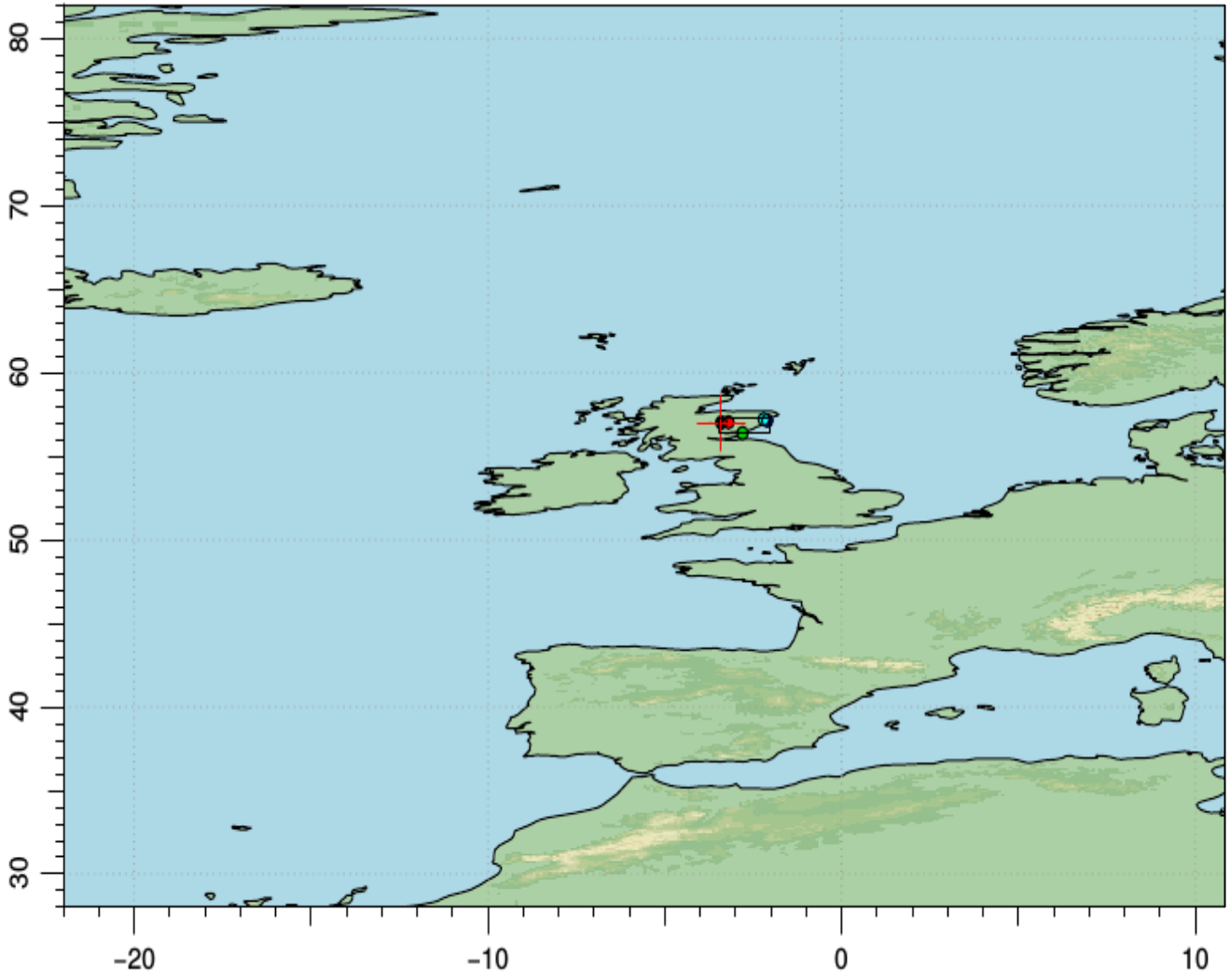
Elke Rustemeier[1], Alice Kapala[2], Anja Meyer-Christoffer[1], Peter Finger[1], Udo Schneider[1], Victor Venema[2], Markus Ziese[1], Clemens Simmer[2] and Andreas Becker[1]

[1]Deutscher Wetterdienst, Hydrometeorology, Offenbach am Main, Germany
[2]Meteorological Institute, University of Bonn, Bonn, Germany

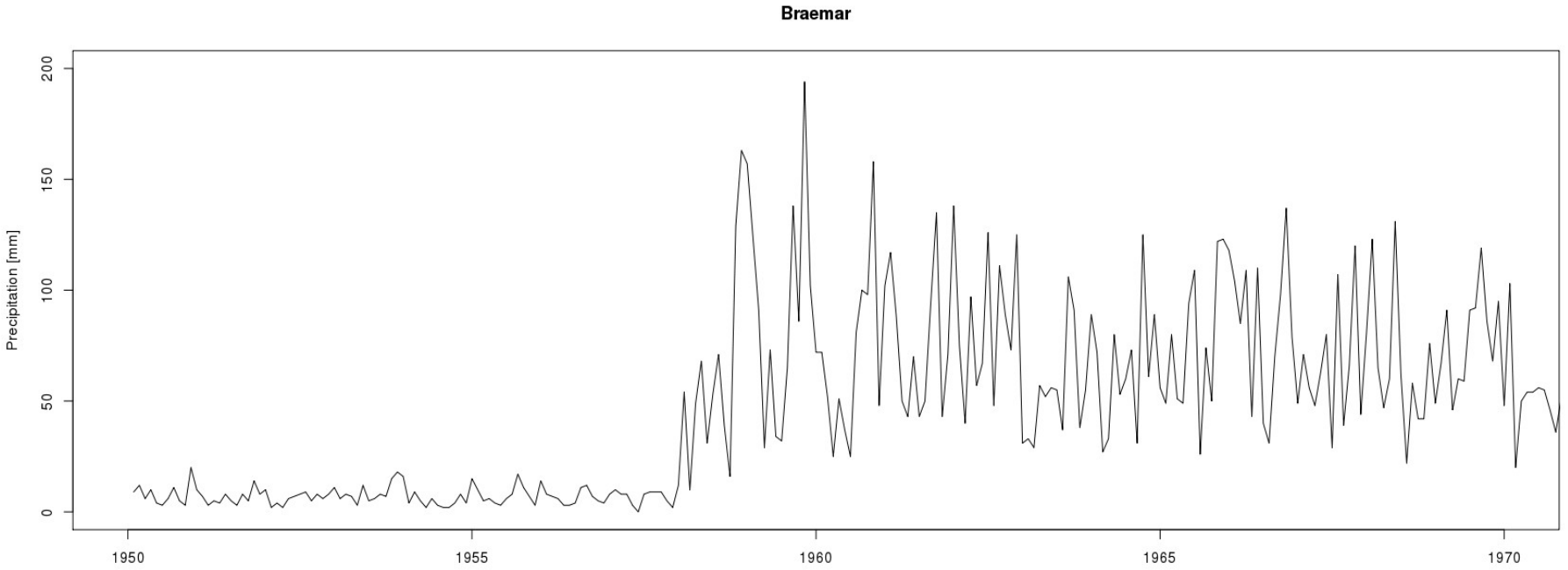
Braemar

Example:

Braemar in Scotland



Monthly homogenization aims for a correct trend



Example: Breamar in Scotland

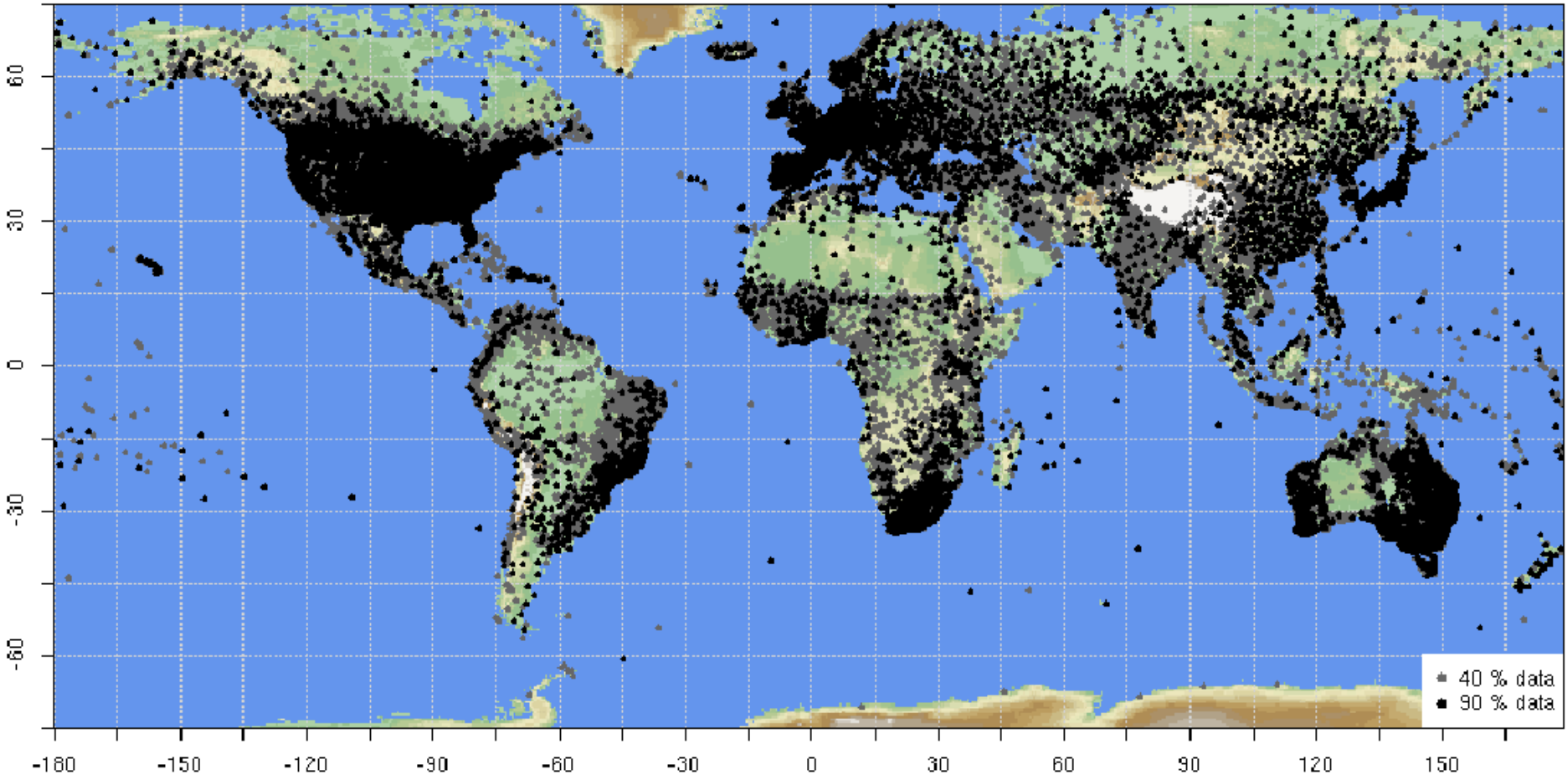
- A trend calculated from the raw data would obviously lead to wrong conclusions.

- Factor 10 error at the beginning of the time series
- Met Office and GPCC (DWD) corrected the data and hold raw and corrected data in their data bases

Overview

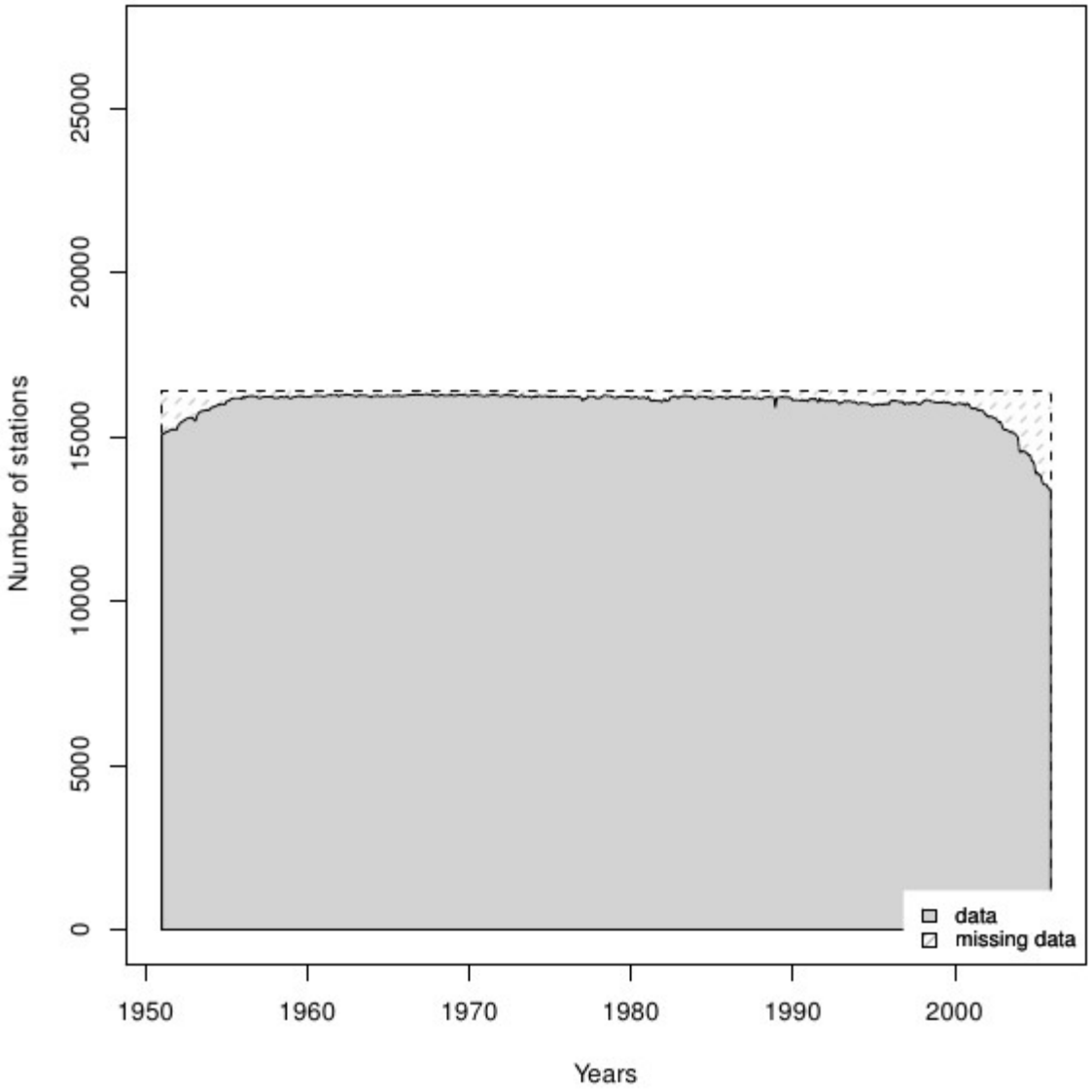
- Data base
- Actual homogenization
  - Networks of similar time series
  - Detection of break-points
  - Correction of breaks
- Interpolation

Data base



Data base

- **1951 – 2005**
- At most 10 % of missing values
- Quality controlled

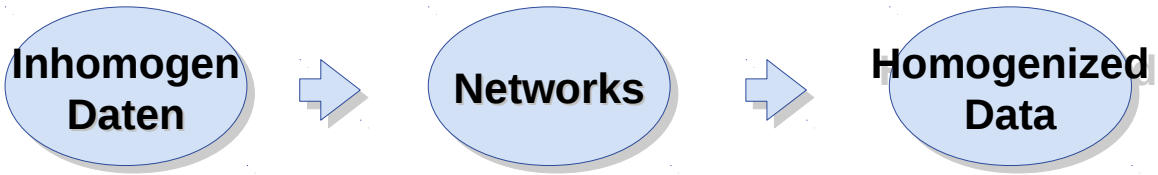


Homogenization of similar time series

- **Homogenization software can not handle more than 300 series at once (of more than 16000 series).**

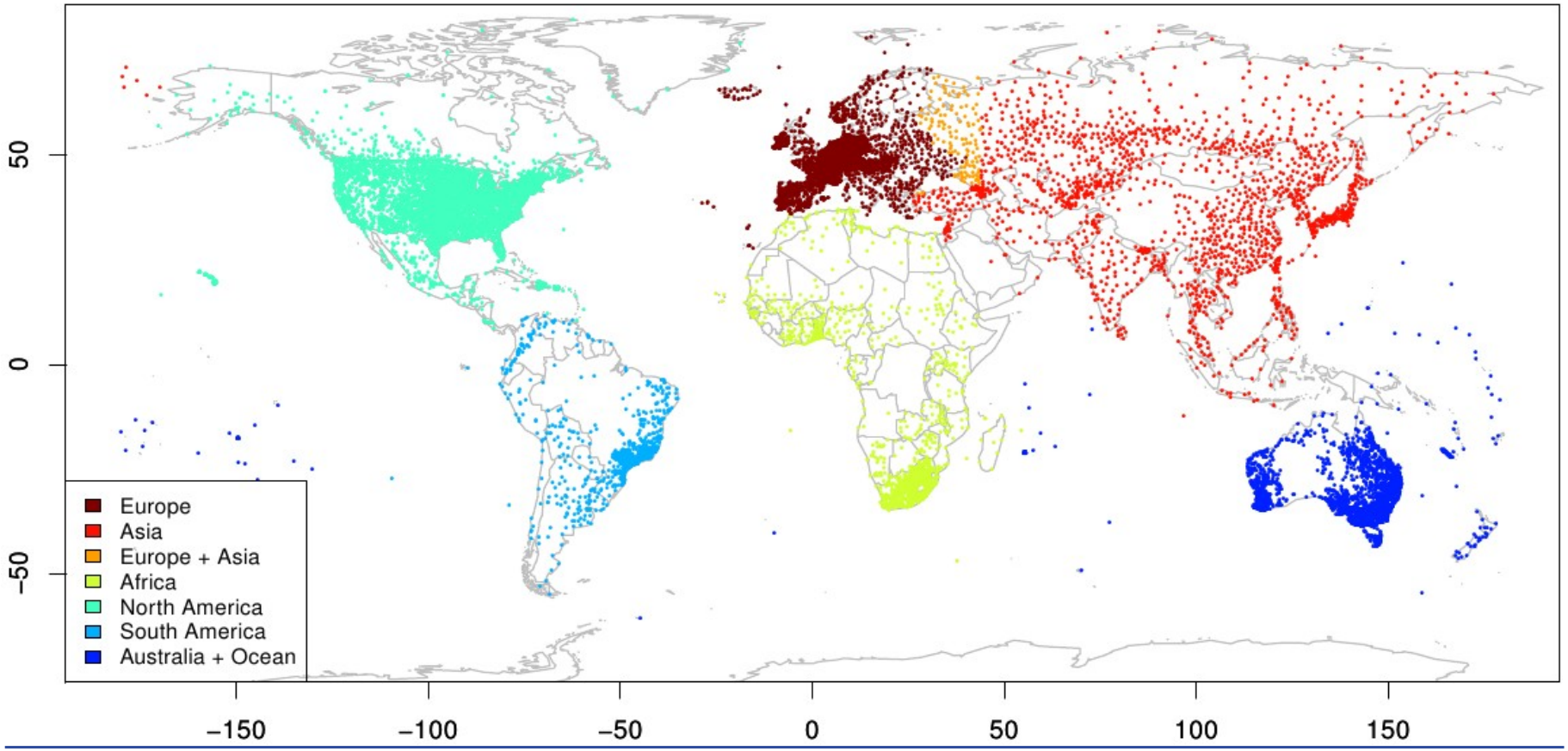
- Series have to be organized in networks of at most **300 series.**
- **Overlapping**

  since the homogenization algorithm depends on comparison with high correlated series.

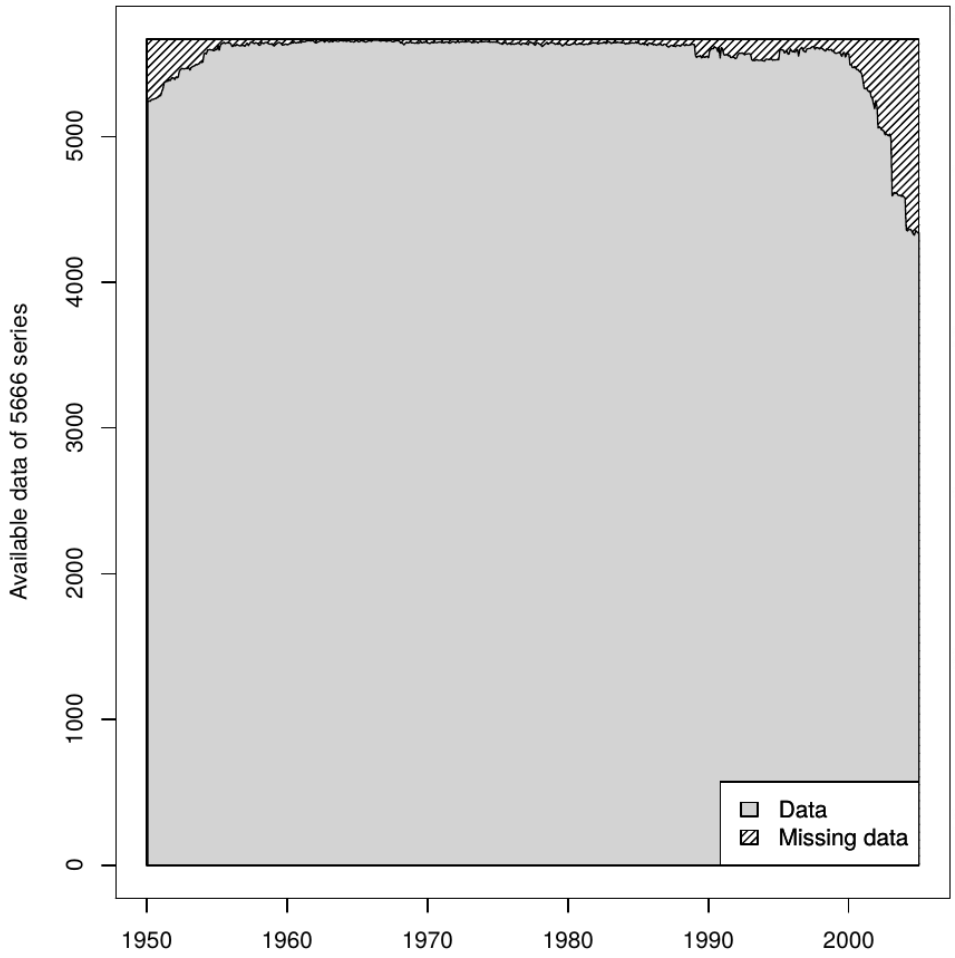


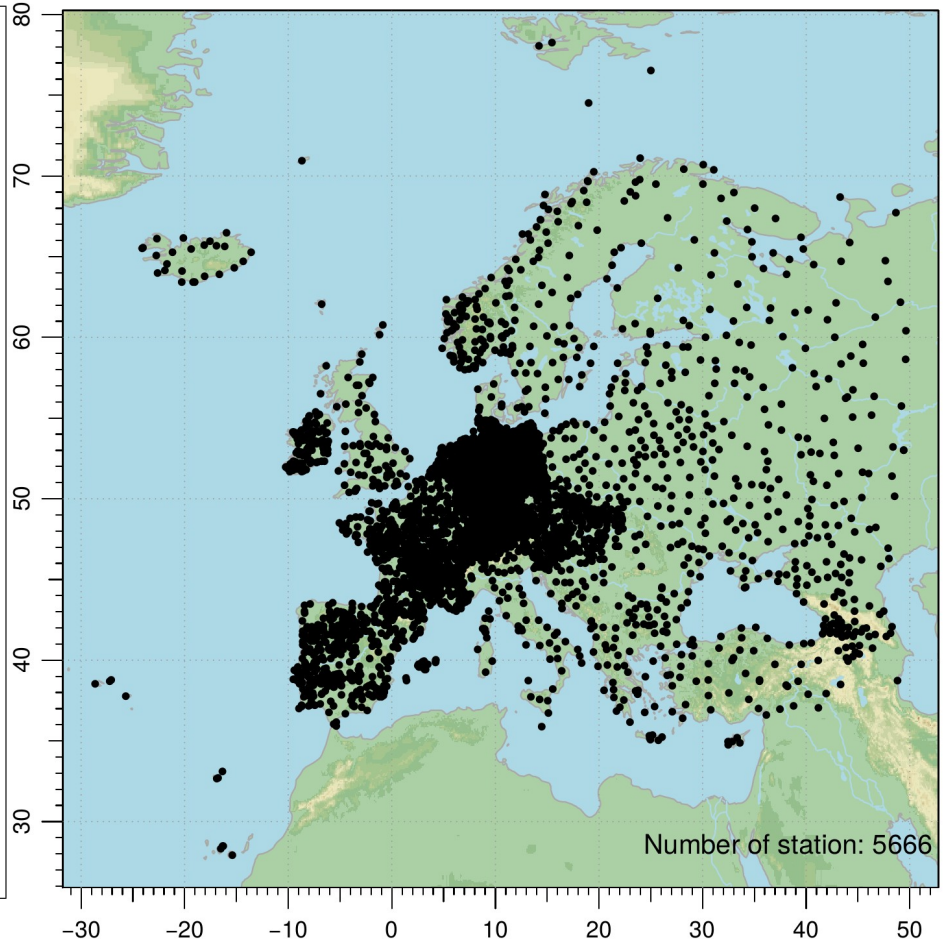
Networks of similar time series: Continents

- First subdevision: Continents
- Spatial distance to the closest continent



Networks of similar time series: Continents





Networks of similar time series

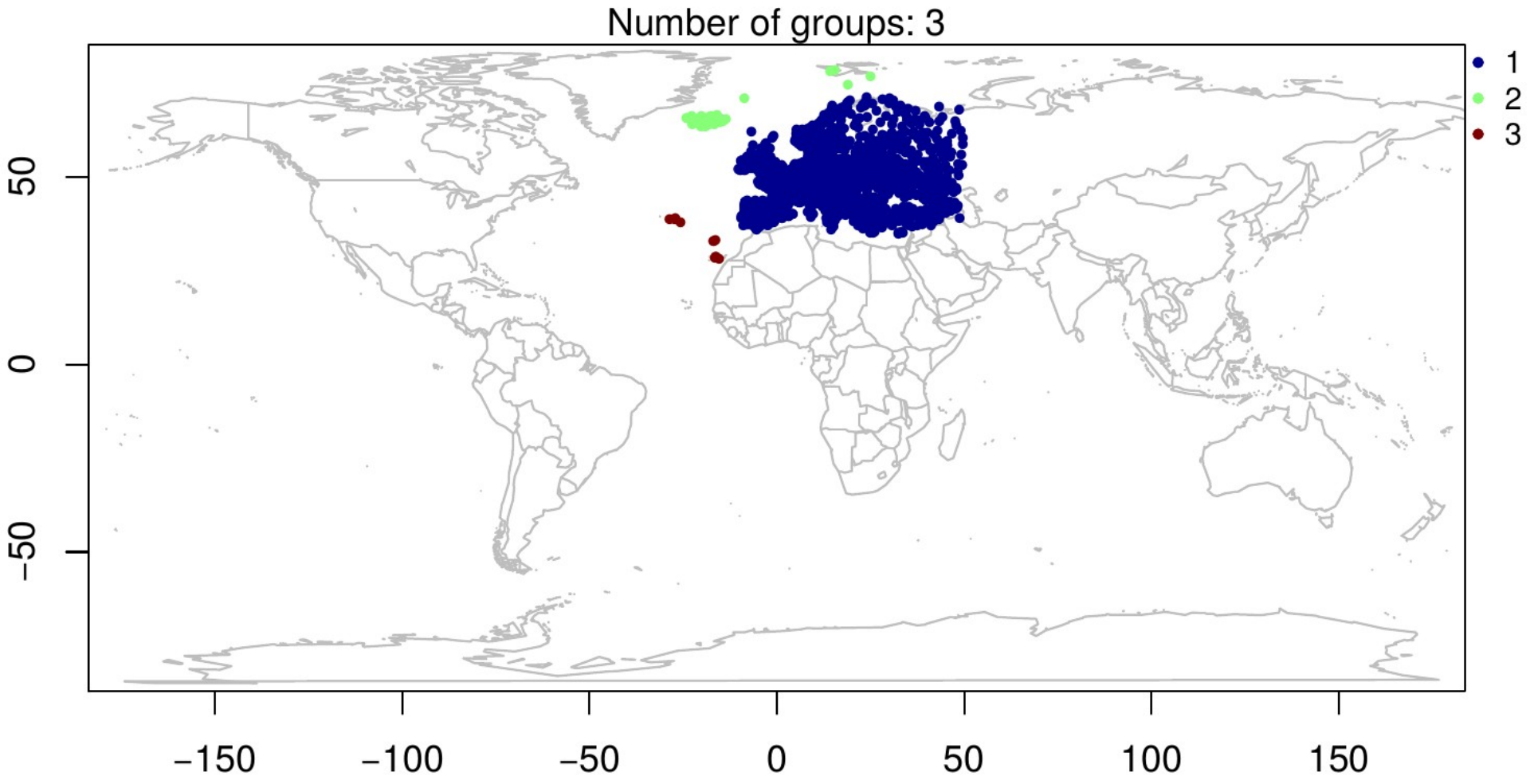
➔ **Second step within continents**

➔ Calculate **great circle distance** between the stations

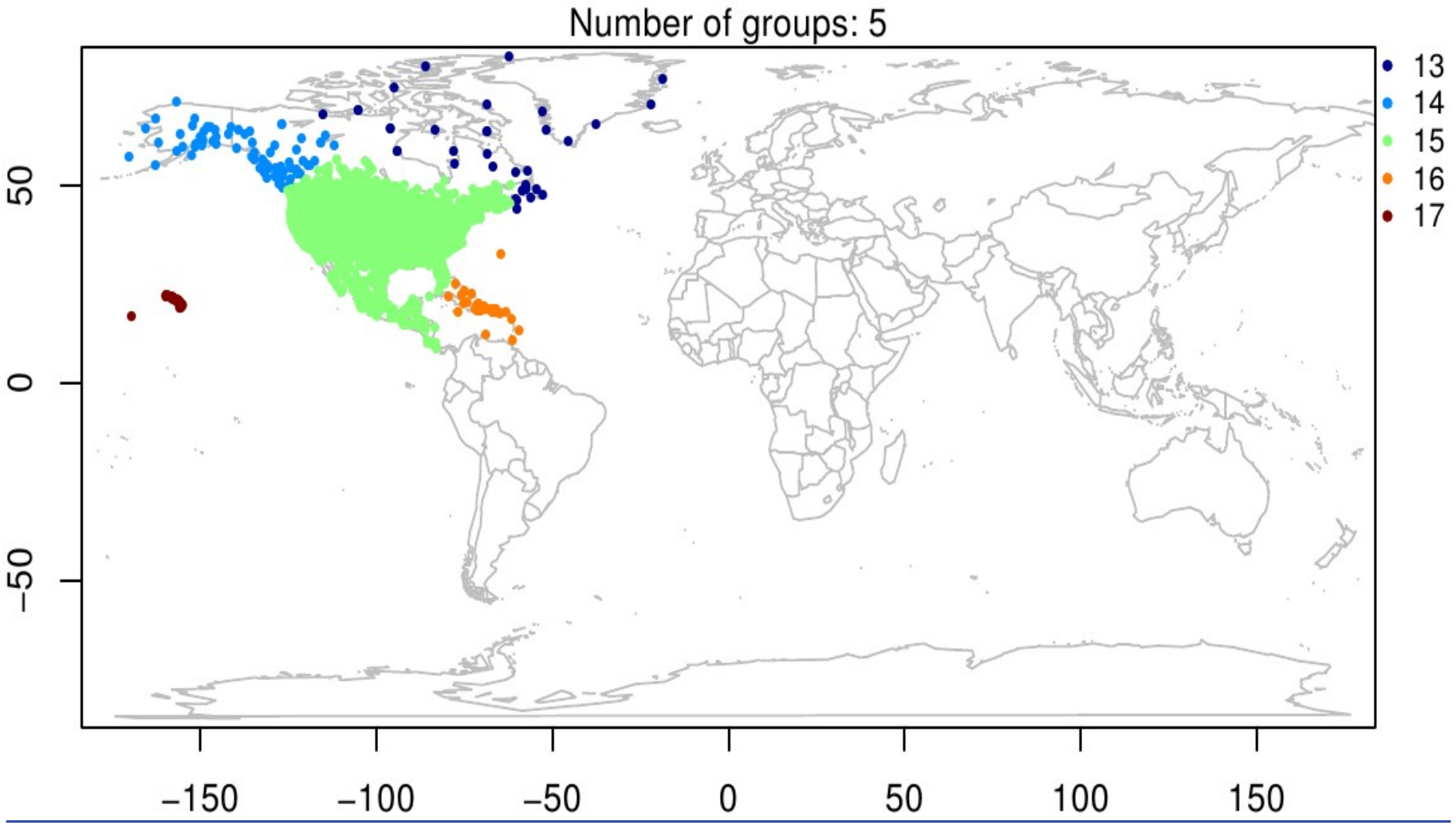
➔ **WARD CLUSTER**

- Hierarchial cluster analysis
- Minimum variance method
- Tends to produce clusters of equal size

Networks of similar time series



Networks of similar time series



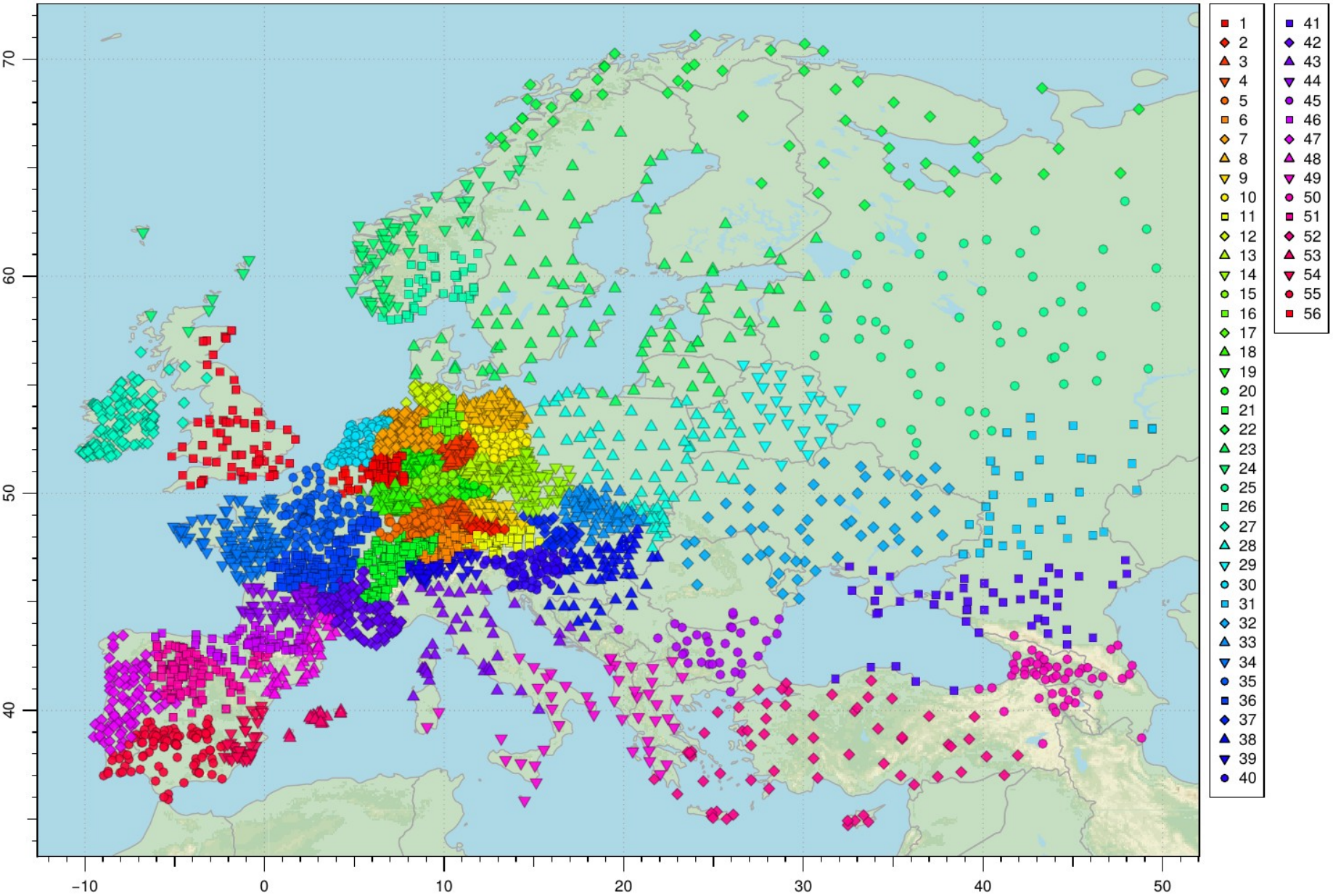
Networks of similar time series

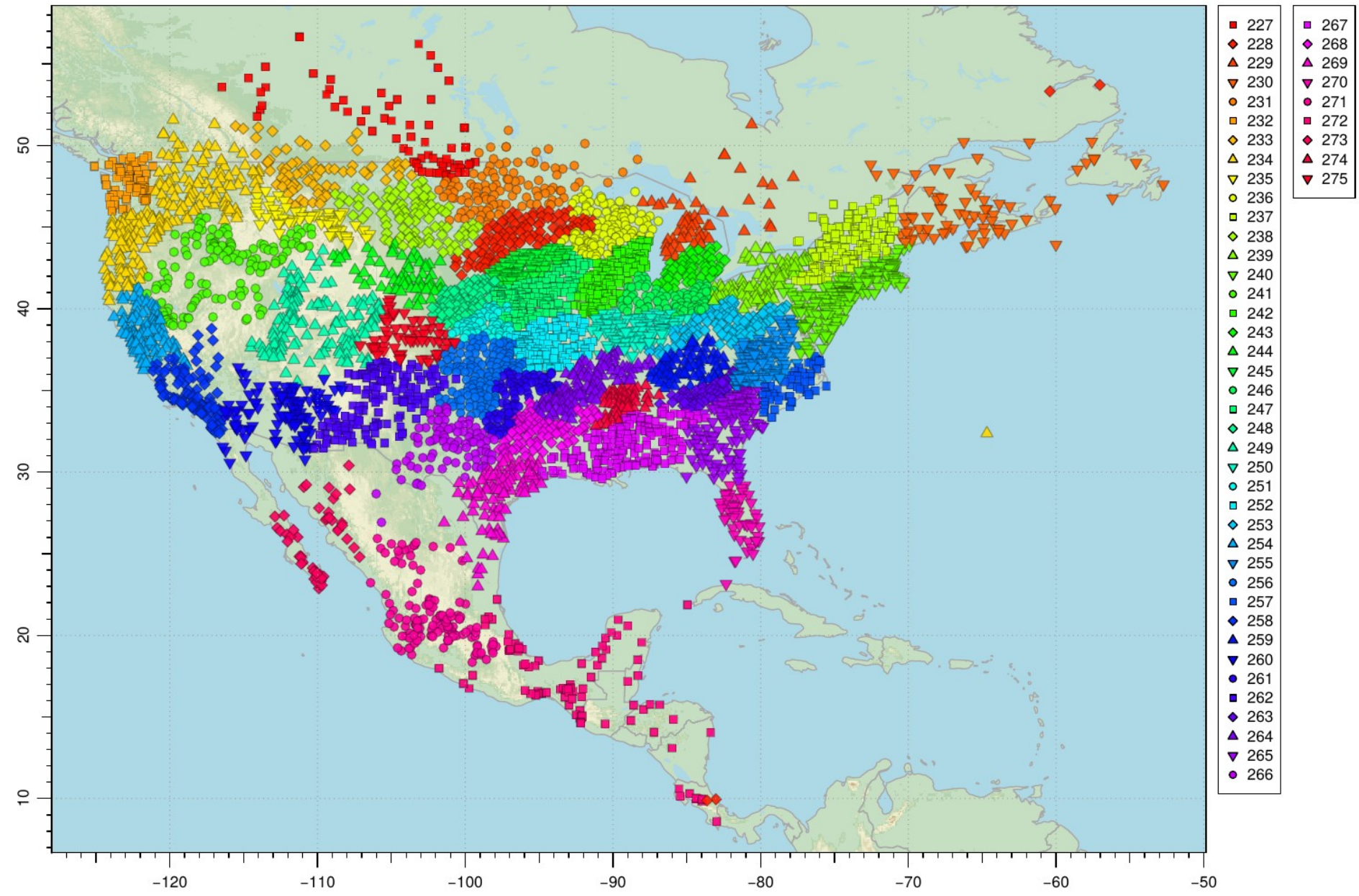
➔ **Third step**

➔ Calculate **partial correlation** between the series

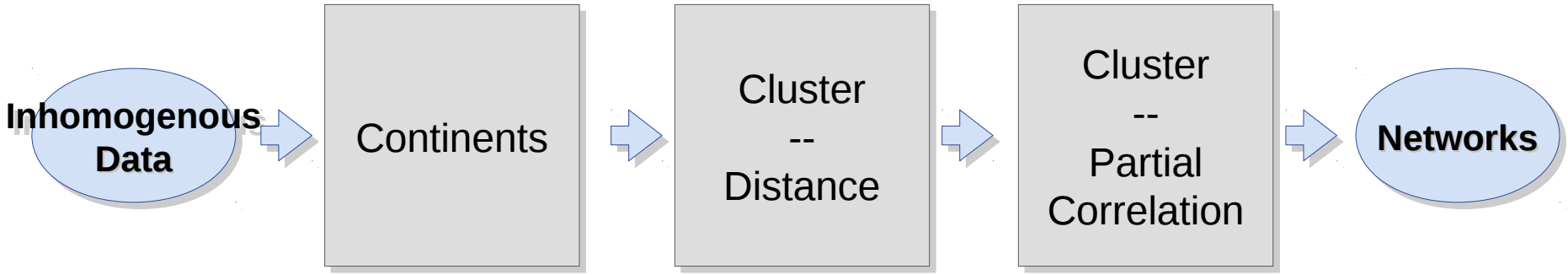
- Consecutive differences
- Removal of the annual cycle
- Calculation of the ranks

➔ **WARD CLUSTER**

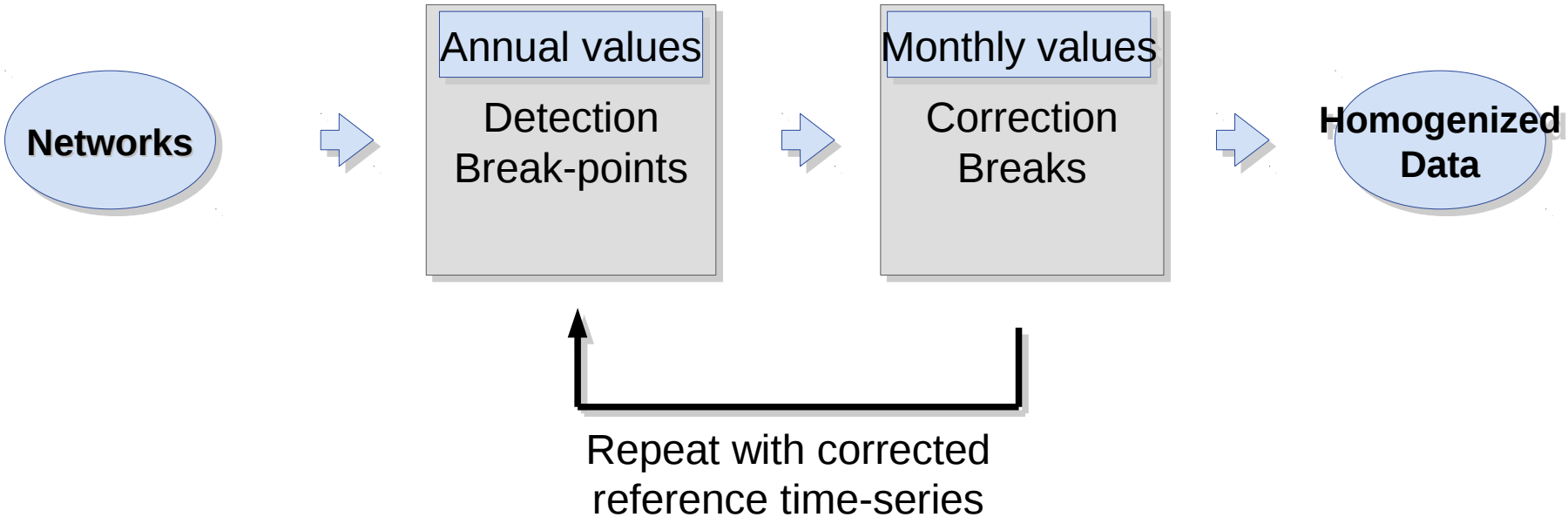

70
60
50
40
−10
0
10
20
30
40
50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56


227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
50
40
30
20
10
−120
−110
−100
−90
−80
−70
−60
−50

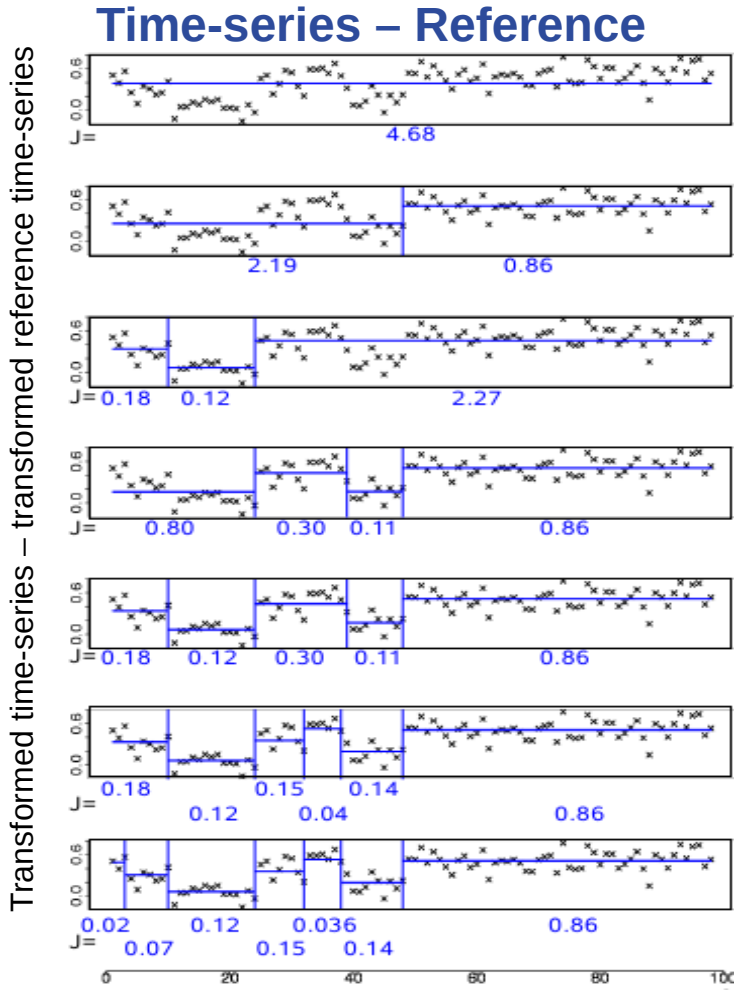
Networks of similar time series: Summary



Homogenization course



Homogenization course: Break-point detection



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

CAUSINUS-MESTRE

Homogenization course: Correction

➔ Box-Cox Transformation

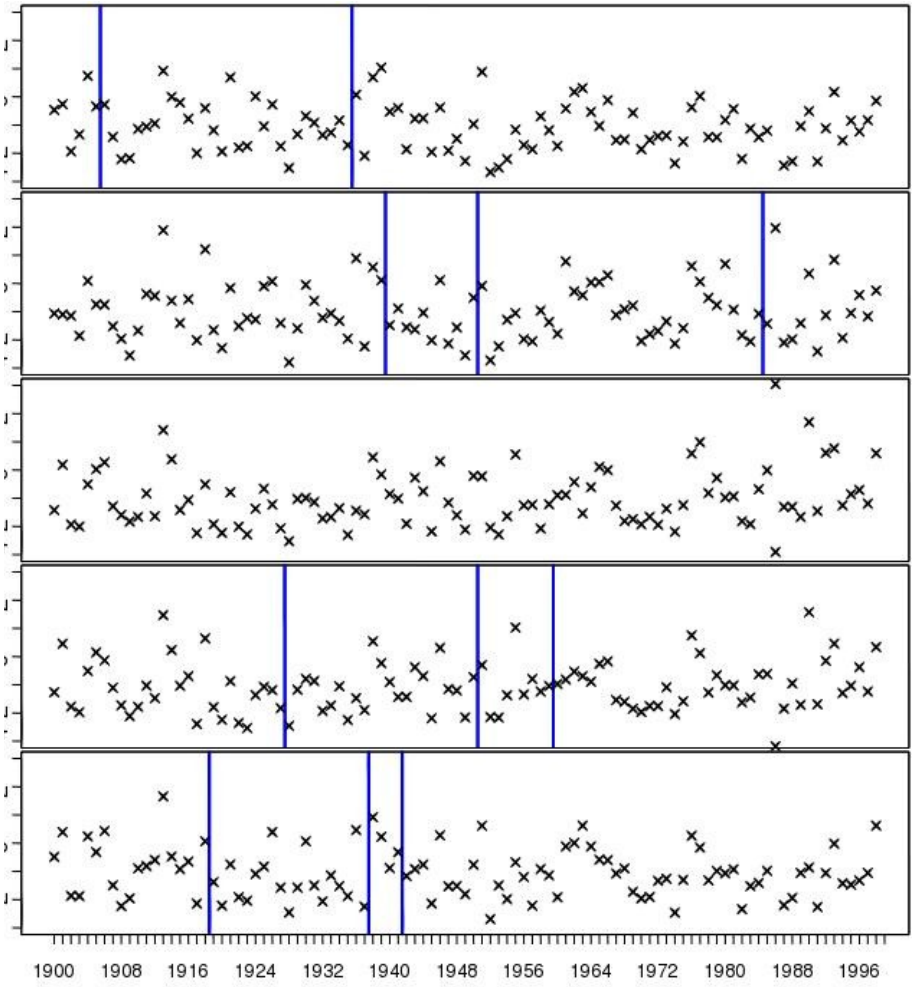
(Software requires normal distribution)

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

➔ Reference series

- High correlated time series

Homogenization course: Break correction



1) Binary coding of the series

2) **Multiple linear regression** over homogeneous segments

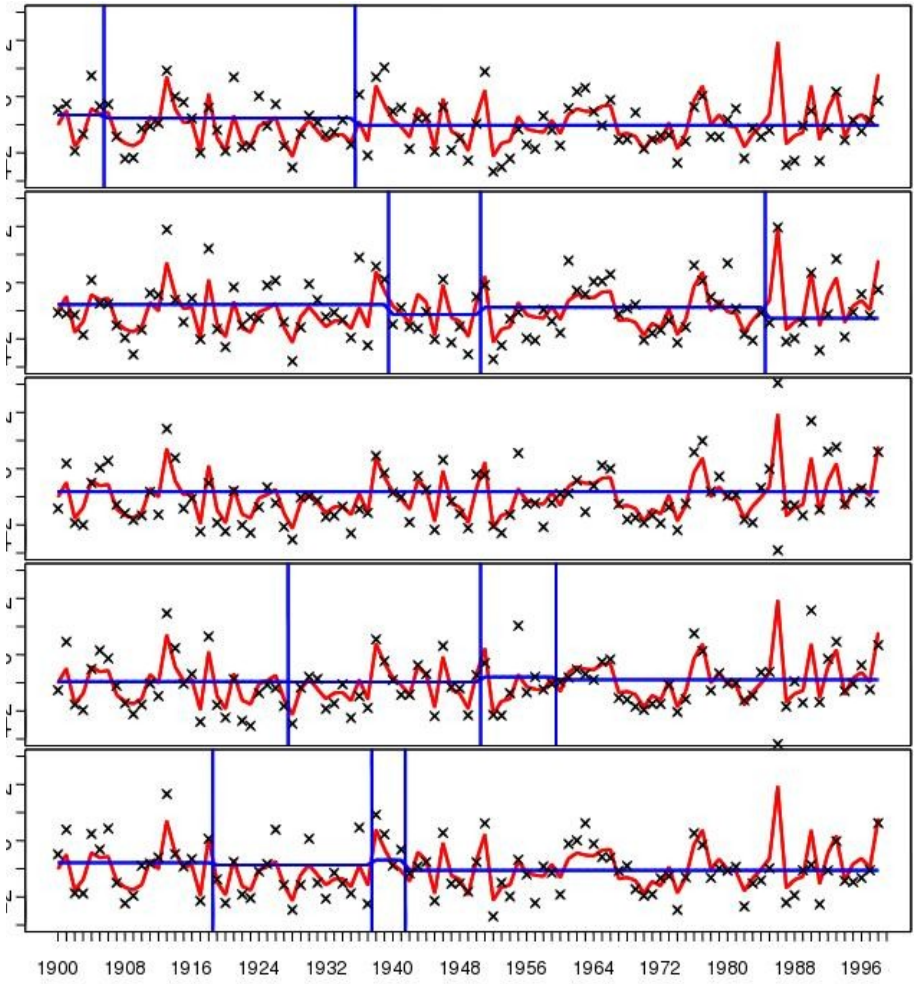
3) **Regression coefficients** indicate **break amplitude**

ANOVA

x Power transformed monthly time-series

| Detected breaks

Homogenization course: Break correction



1) Binary coding of the series

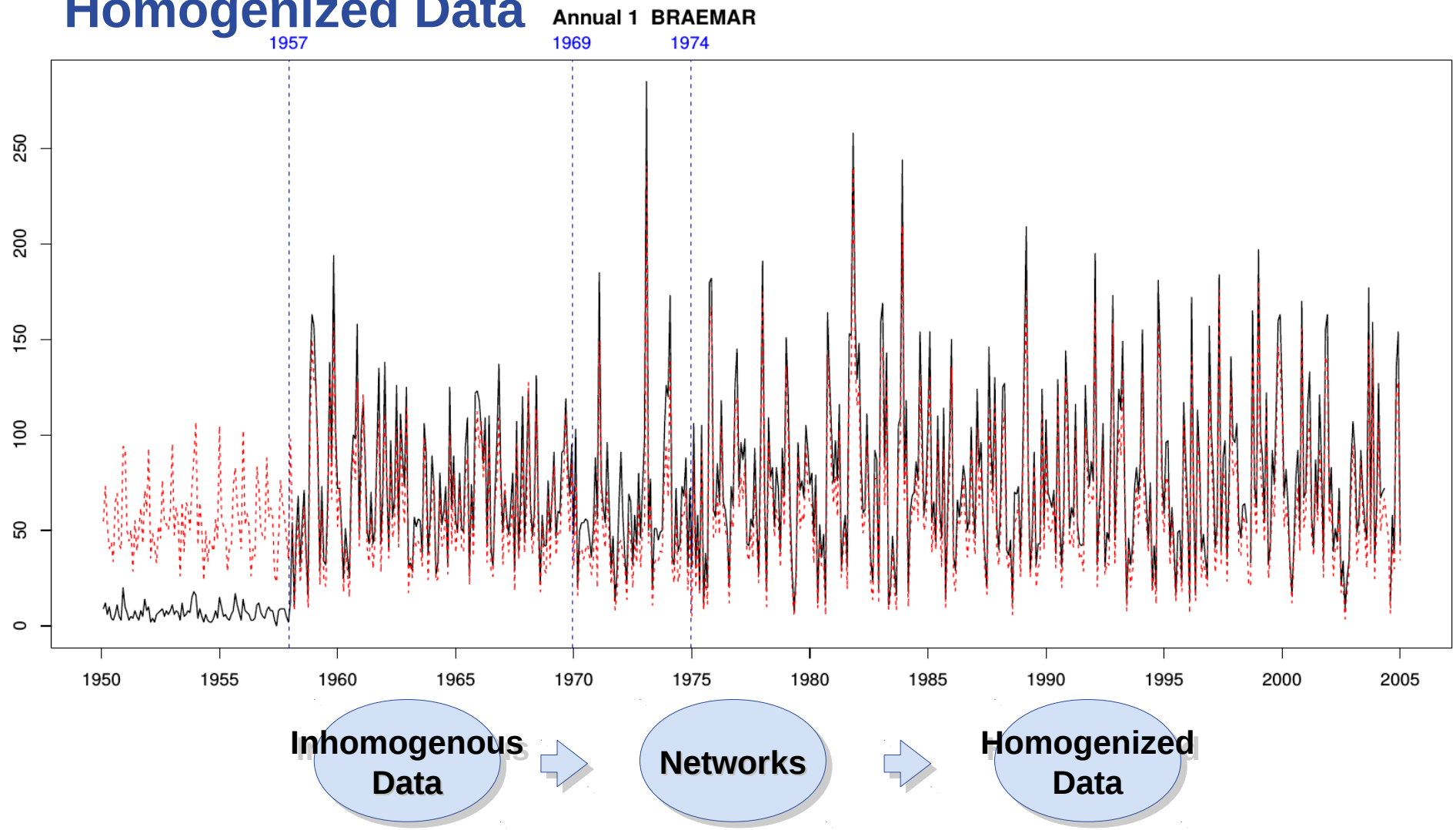
2) **Multiple linear regression** over homogeneous segments

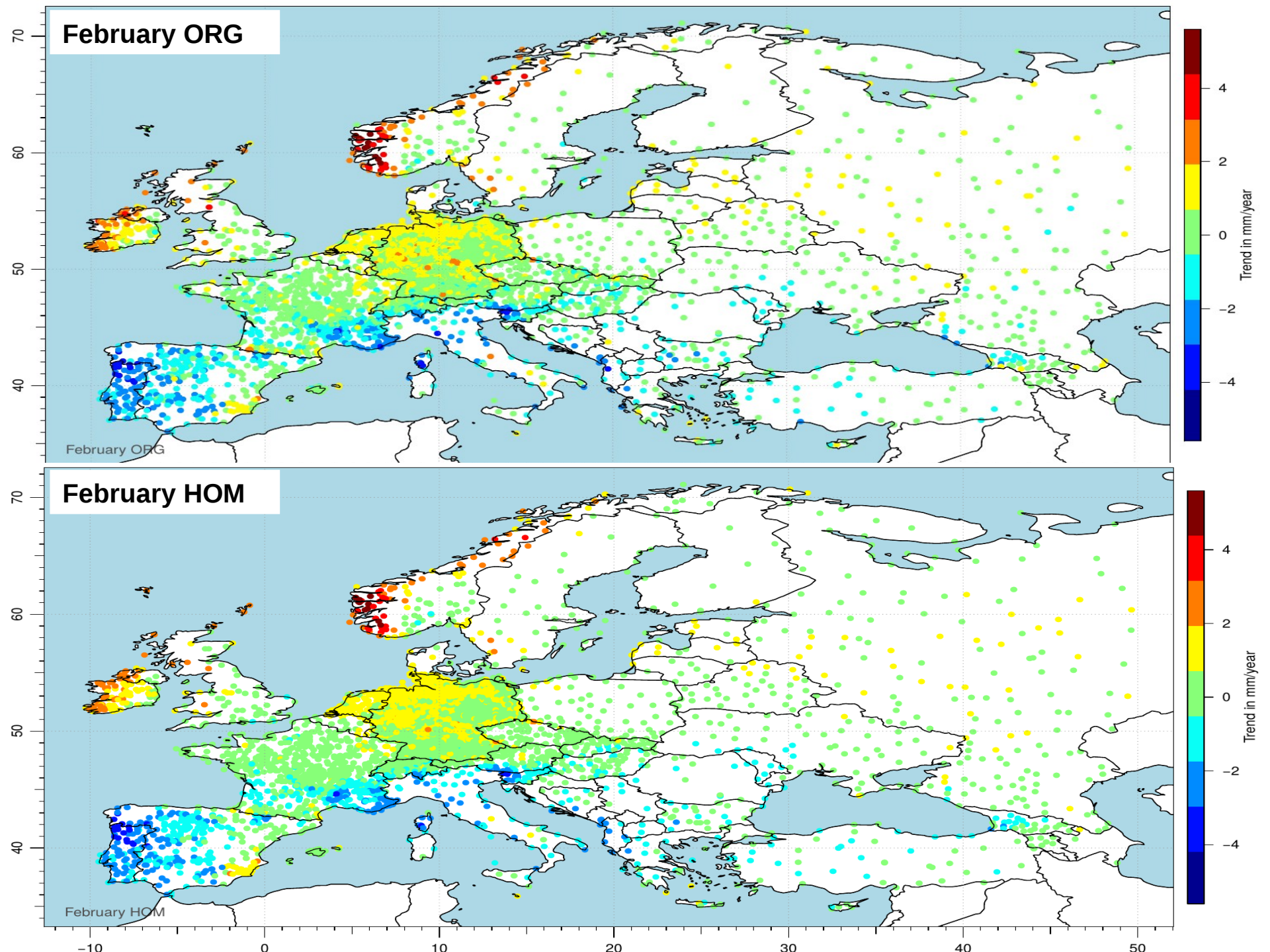
3) **Regression coefficients** indicate **break amplitude**

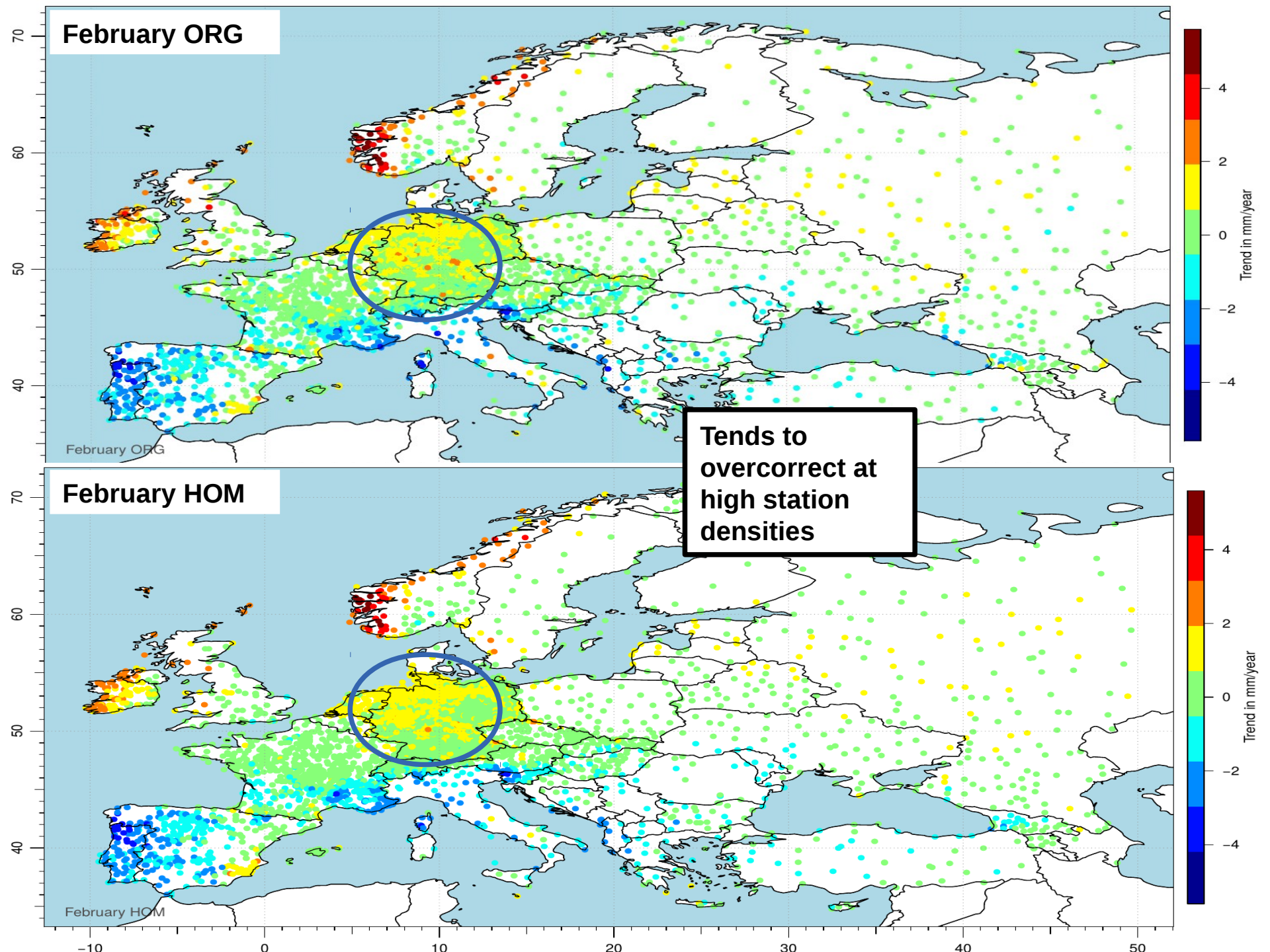
ANOVA

- Monthly regression parameter
- **Segment regression parameter**

Homogenized Data




February ORG
February HOM
Trend in mm/year
4
2
0
−2
−4
70
60
50
40
−10
0
10
20
30
40
50


February ORG
February HOM
Tends to overcorrect at high station densities
Trend in mm/year
4
2
0
−2
−4
70
60
50
40
−10
0
10
20
30
40
50

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

3.) Verification

➔ **Especially important due to automation**

➔ Usually testing on independent data

- **Artificial data**

➔ Sensitivity study

- **Variation of reference series**

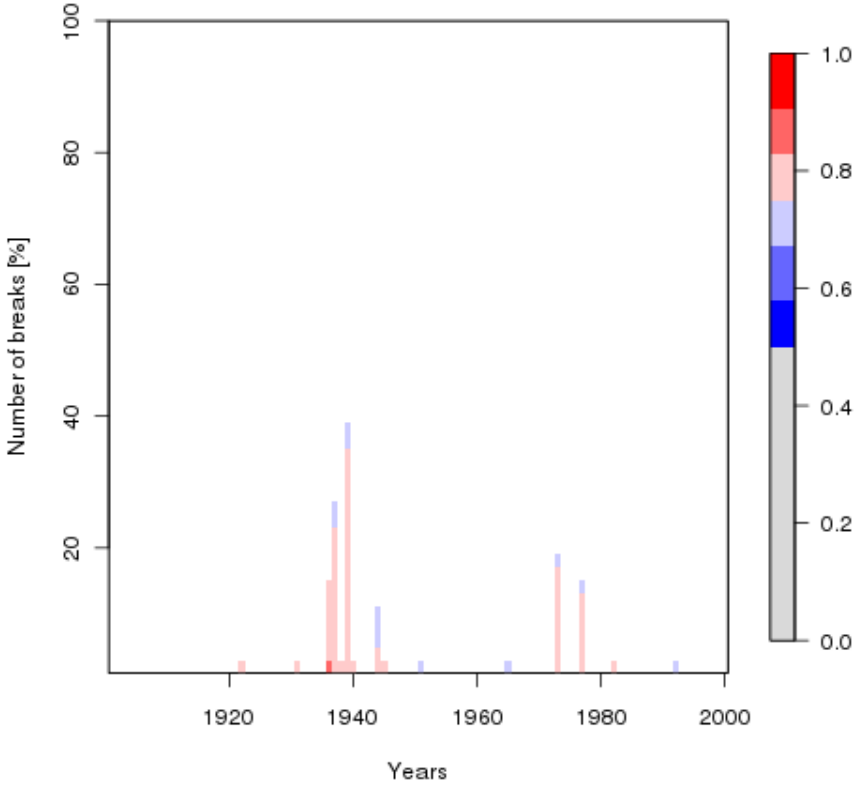
  (Has to be repeated for actual data set, because of changes in the software)

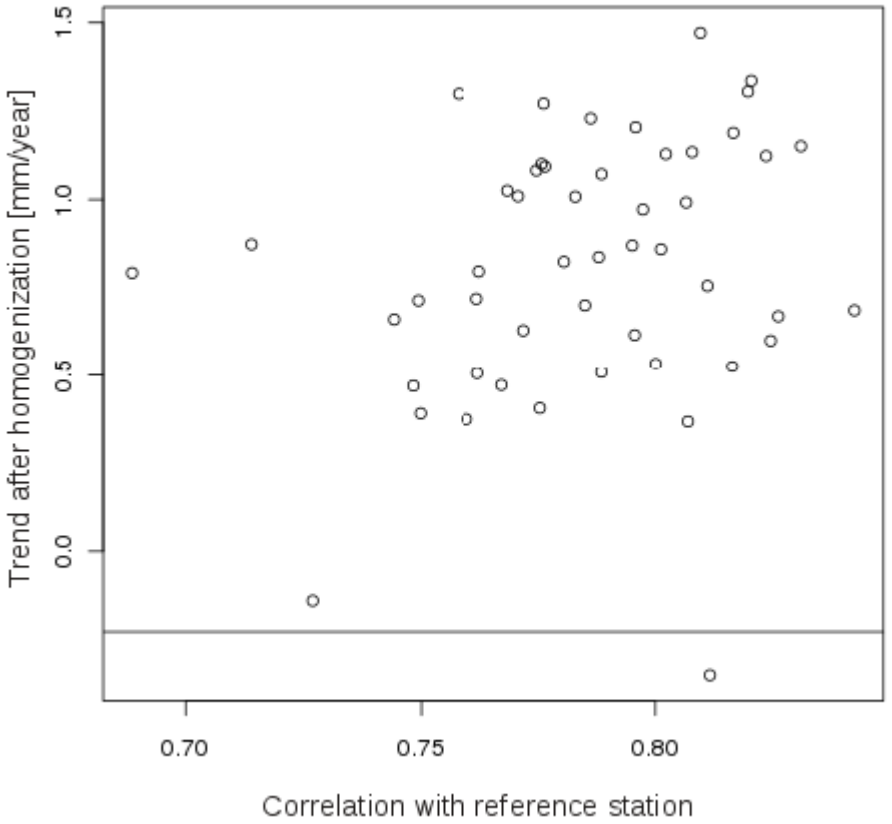
➔ **Comparison with digitized meta data**

(Has to be repeated for actual data set, because of changes in the software)

➔ Suspicious series are controlled **manually**

Sensitivity study (different data base)





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

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

➔ Kira Rehfeldt will present more information about the interpolation methods used at the GPCC in her talk

Summary and next steps

- Development of an **automatic algorithm**
  - Allows homogenization of large data sets
  - Over correction at dense station networks

Summary and next steps

- Development of an **automatic algorithm**
  - Allows homogenization of large data sets
  - Over correction at dense station networks
- Run **validation algorithm**
  - Meta data
  - Sensitivity study
- Interpolation
- Publish HOMPRA Europe (**gpcc.dwd.de**)
- Probably end of April

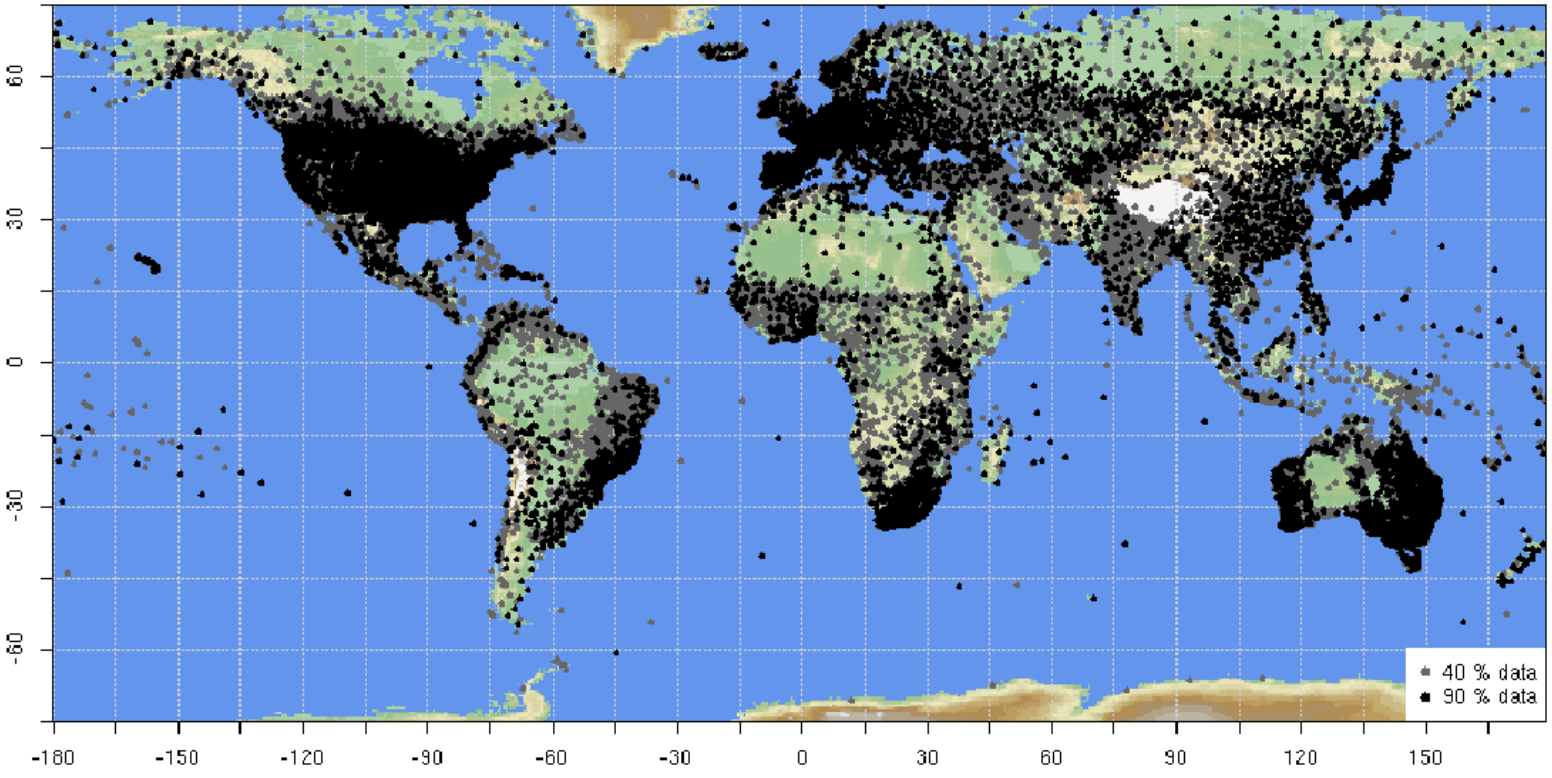
Summary and next steps

- Development of an **automatic algorithm**
  - Allows homogenization of large data sets
  - Over correction at dense station networks
- Run **validation algorithm**
  - Meta data
  - Sensitivity study
- Interpolation
- Publish HOMPRA Europe (**gpcc.dwd.de**)
- Probably end of April
- Comparison with **other homogenized data sets**

  (eg. Irland cooperation with John Coll, Mary Curley and Peter Domonkos)

Still remaining global issues (selection)

- **Include time series with many missing values as reference series**
- Correction (Needs validation)
- Detection (Not startet yet)



Still remaining global issues (selection)

- **Precipitation distribution is assumed to be constant over time (in comparison to neighbor series)**
- Changes in the distribution are detected as breaks

- In Australia the border between tropics and subtropics is not constant
- El Niño/ La Niña years in South America
- Hurricans in the United States
- ...

Thank you for your attention!