QUALITY CHECK AND HOMOGENIZATION OF ECA&D TEMPERATURE DATA-SET

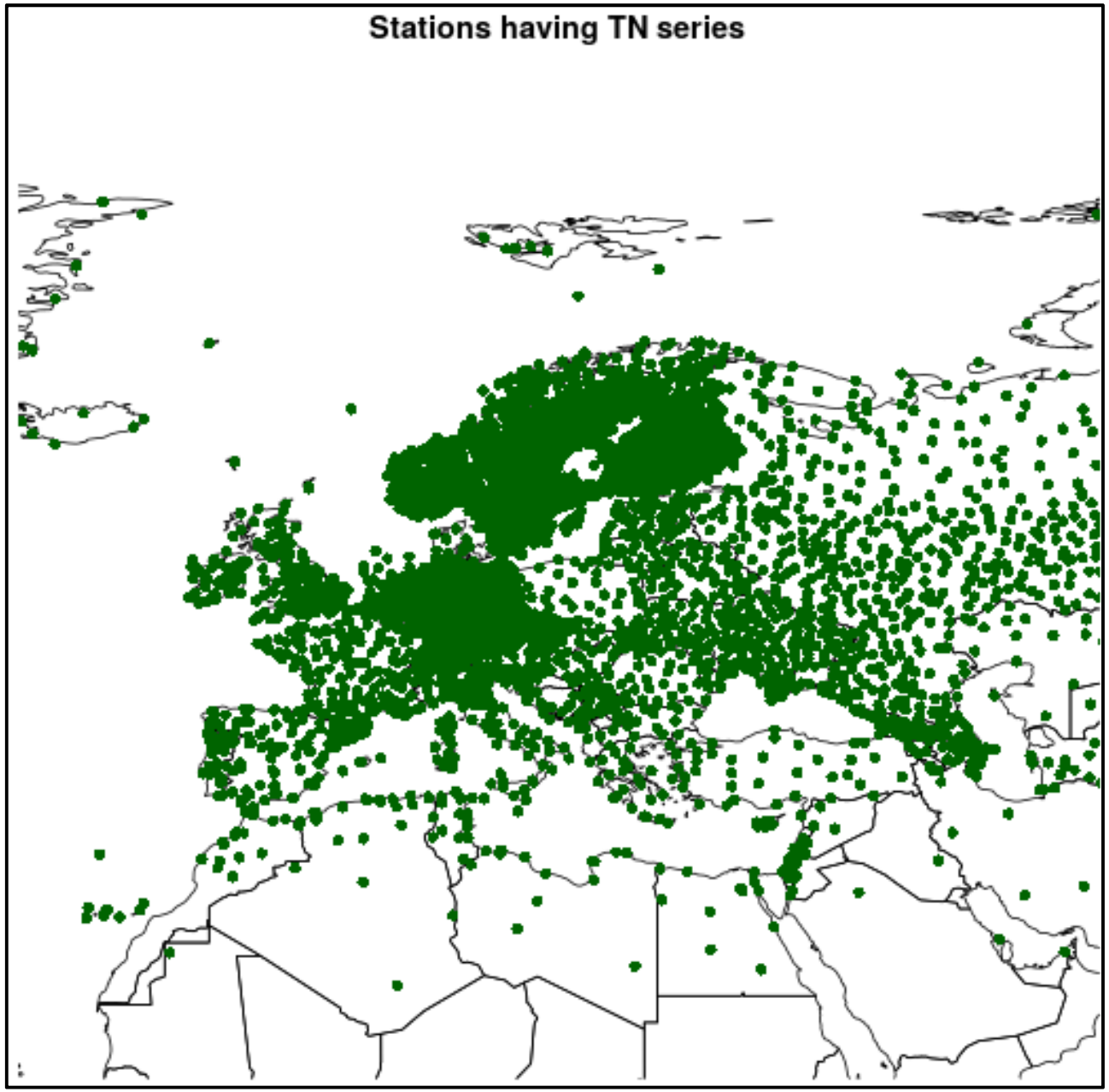
Antonello Squintu[1], Yuri Brugnara[2], **Gerard van der Schrier[1]**, Petr Stepanek[1], Pavel Zahraetr Stepan[1], Petr Stepanek[1]

[1]KNMI, [2]University of Bern

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

ECA&D DATASET



Collection of ground- based observation all over Europe

Temperature, radiation, precipitation, humidity, pressure, etc. etc.

http://www.ecad.eu/

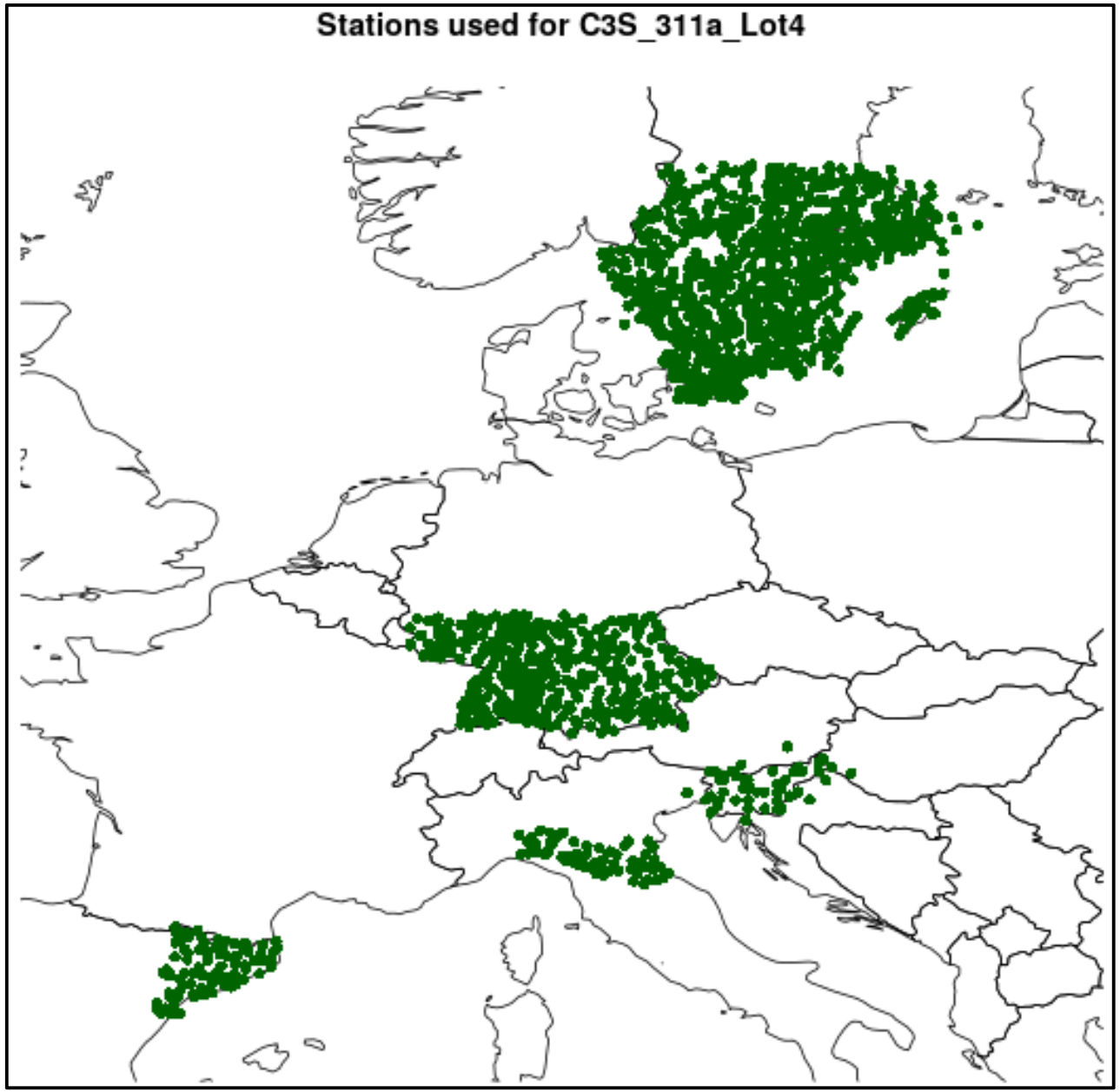
E-OBS gridded data-set (newly released vers. 16)

Number of stations and series constantly increases thanks to new participants and updates

EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

QUALITY CHECK: THE PLAN



Creation of a benchmark data-set with a preliminary quality check.

Evaluation of the methods:
-ProClimDB (CzechGlobe)
-MASH (OMSZ)
-GHCN (NCDC-NOAA)

Selection of the best QC method basing on performance on the benchmark data-sets.

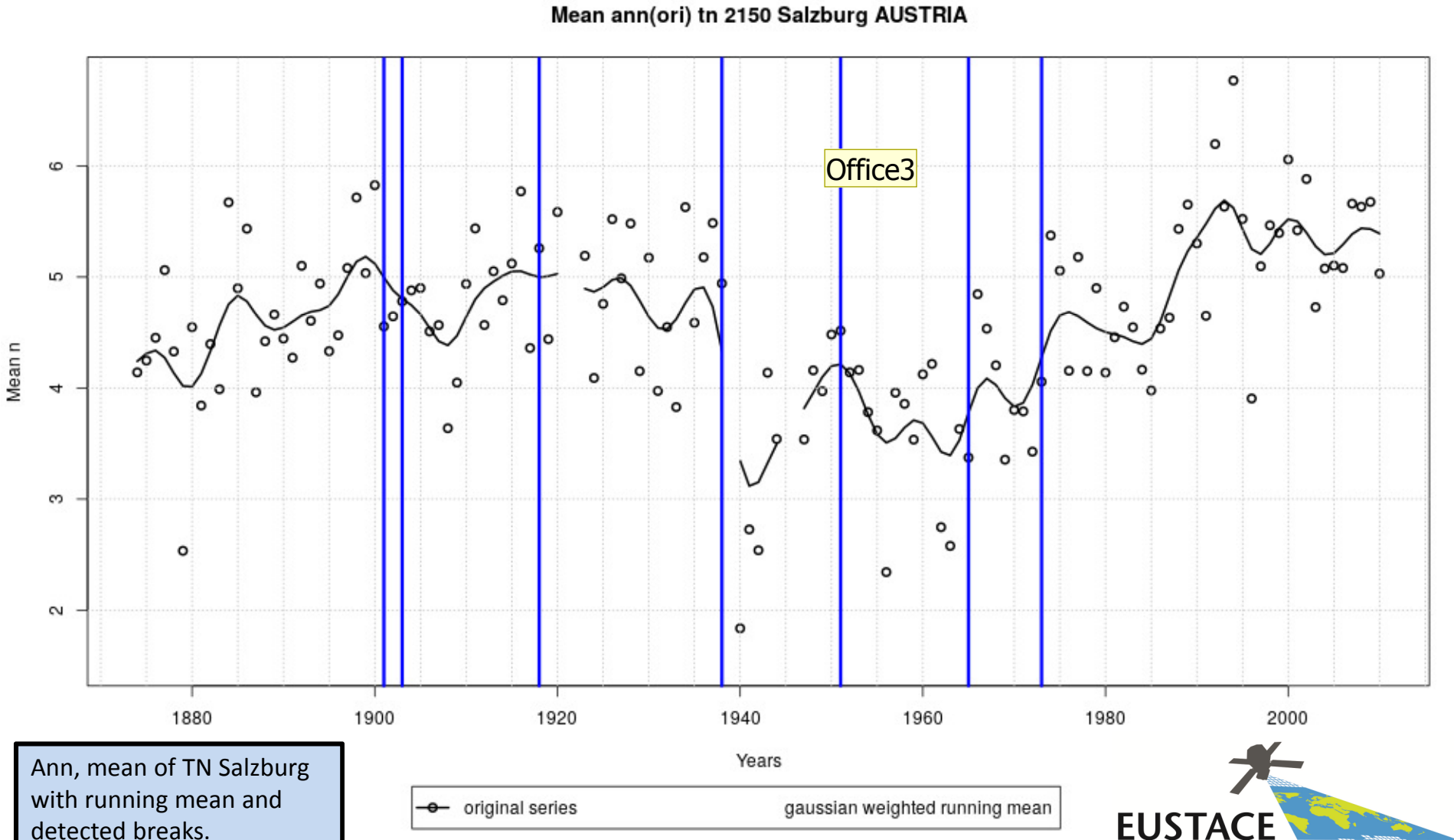
Improve it inserting performant features of the other two methods.

EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

BREAK DETECTION (BD)

Break detection as agreement between RHtest, CAUME, GAHMDI on a monthly base (Kuglitsch et al., 2012) performed by Yuri Brugnara (University of Bern)



Ann, mean of TN Salzburg with running mean and detected breaks.

EUSTACE

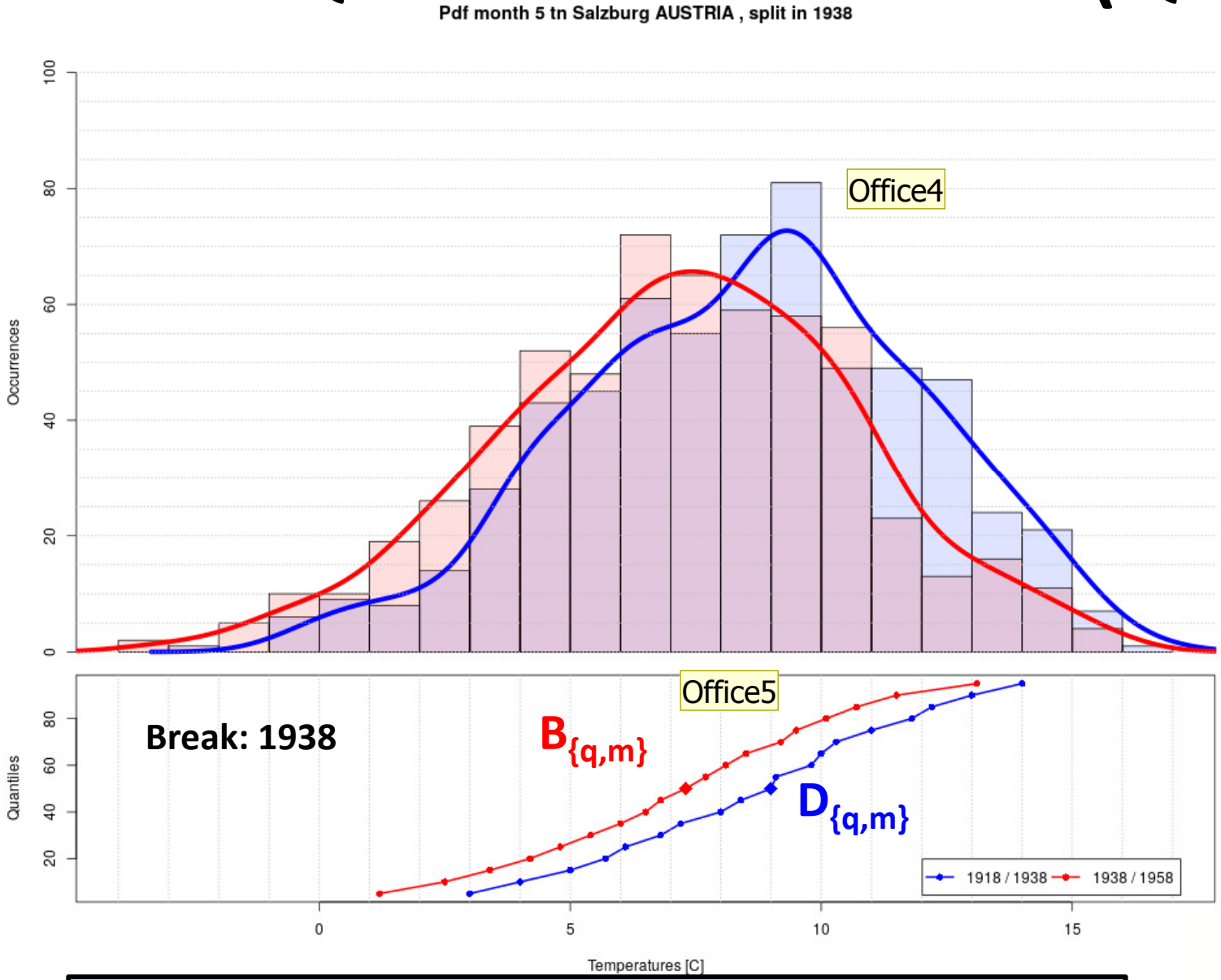
Slide 4

Office3 In this slide the annual mean of minimum temperature in Bamberg (Germany) is shown. The blue vertical lines indicate the breaks that have been found in the first iteration of the break detection (Kuglitsch et al. 2012, performed by Yuri Brugnata, University of Bern). The break in 1952, which shows a jump of at least 1 degree is chosen as sample for the study of the effects of the homogenization process on a series.

Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

QUANTILE MATCHING (QM)



Temp. pdf and quant. of TN Salzburg (May) 10 years before (blue) and after (red) 1938

Trewin (2012)

Statistical differences in the temperature pdf before and after the break.

If overlap period is present:

$A_{\{q,m\}} = B_{\{q,m\}} - D_{\{q,m\}}$

If not, it's necessary to use homogeneous reference series

EUSTACE

Slide 5

Office4 The 2 distributions represent the pdf of minimum temperatures in Bamberg before (blue) and after (red) 1952 (period of 10 years before and after the break are used) for the month of May. The distance between the two distribution is due to the combination of climatic signal (which has to be calculated thanks to the reference series) and artificial signal. (Don't forget that the corrections are always made such that the earliest parts become consistent with the latest one)
Utente di Microsoft Office; 24.8.2017.

Office5 The bottom panel displays the quantile functions, I.e. the values of the quantiles from 5 to 95 of the sub-sets before (blue) and after (red) the break in 1952. In case these two sequences corresponded to overlapping period, the difference between them would have been sufficient for the calculation of the adjustments. But this is not the case...
Utente di Microsoft Office; 24.8.2017.

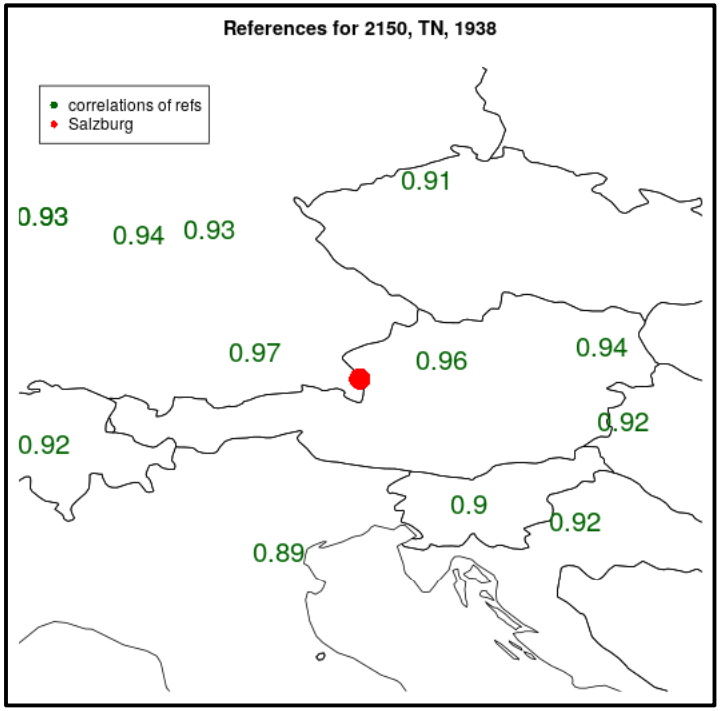
Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

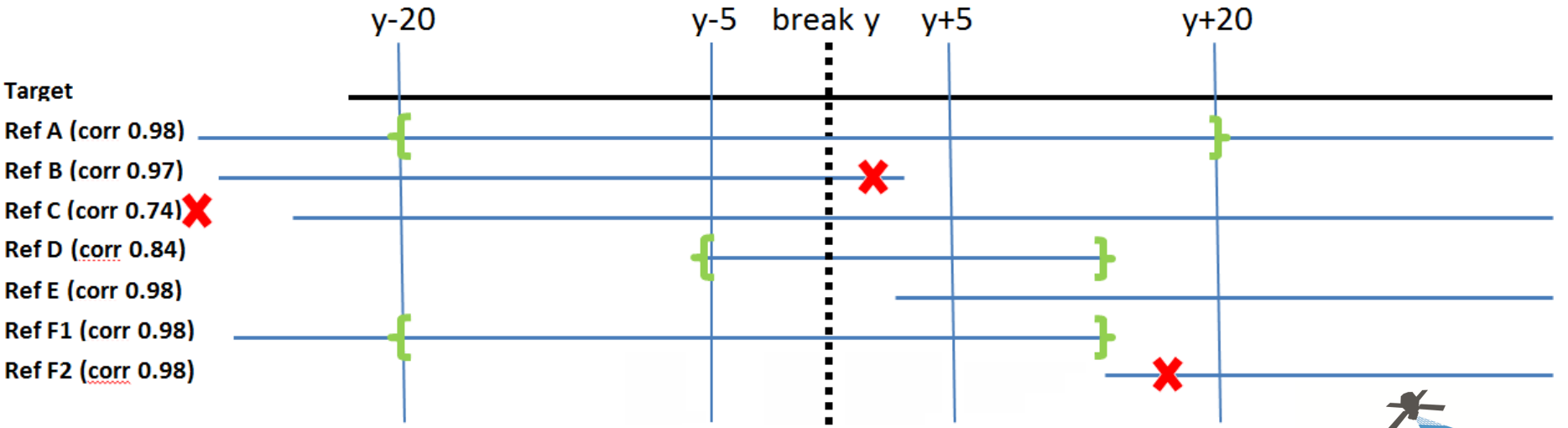
REFERENCE SELECTION

References are selected from the dataset in a 6°x6° box centred on the target s., with no more than 500 m of elevation difference (height/2 for mountain stations).

In very dense areas the **longest** and the **earliest** series are selected. These are splitted in homogeneous sub-series according to the results of the break detection.

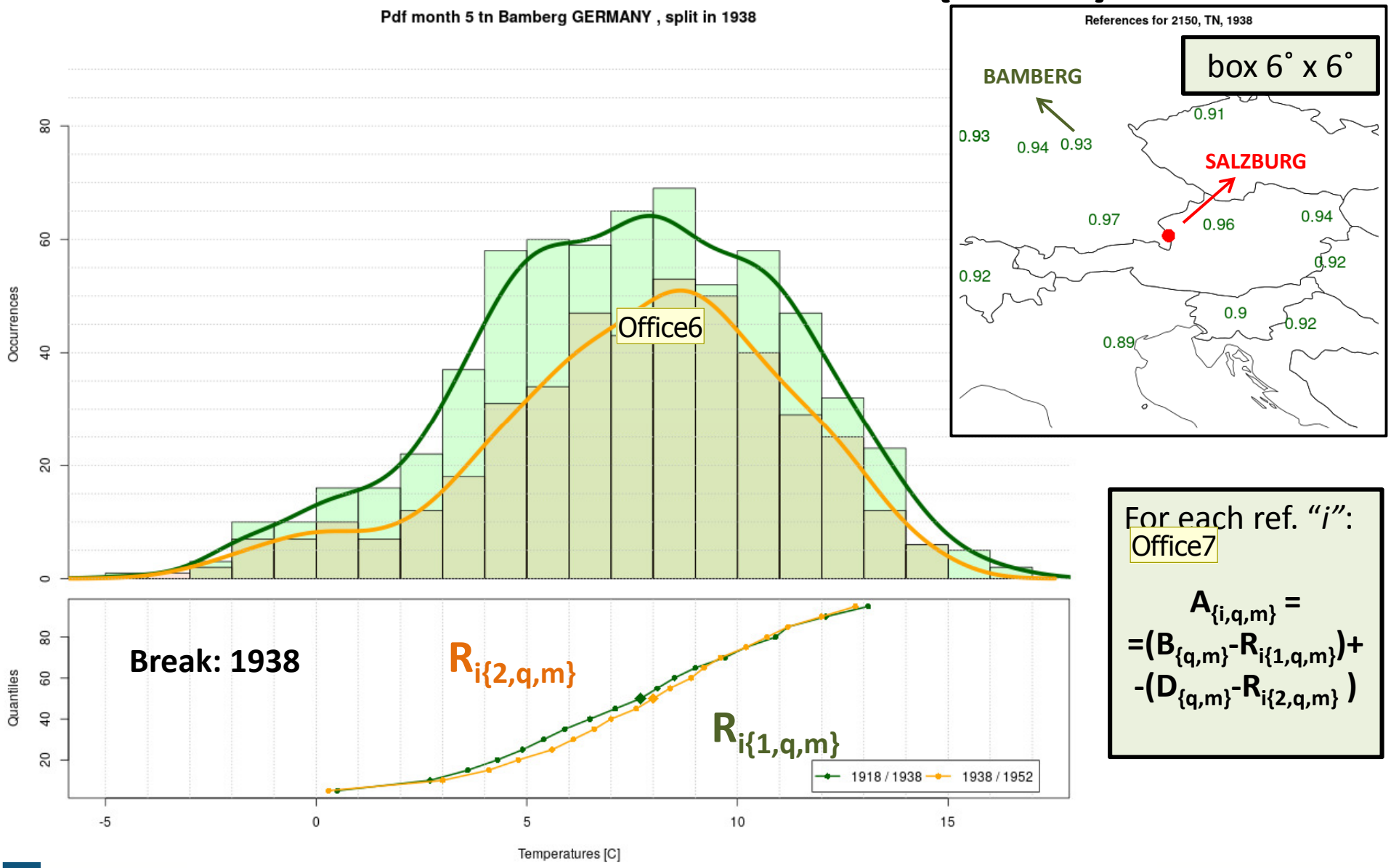
Criteria: overlap >(**5+5) years**, maximum (20+20) years of length, choice of highest correlated, in low density areas correlation **never below 0.75**.





EUSTACE

HOMOGENEOUS REFERENCE (SUB-)SERIES



For each ref. "*i*": Office7

$$A_{\{i,q,m\}} = =(B_{\{q,m\}}-R_{i\{1,q,m\}})+ -(D_{\{q,m\}}-R_{i\{2,q,m\}})$$

Temp. pdf and quant. of TN Bamberg (May) 20 years before (green) and after (orange) 1938

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Environment

EUSTACE

Slide 7

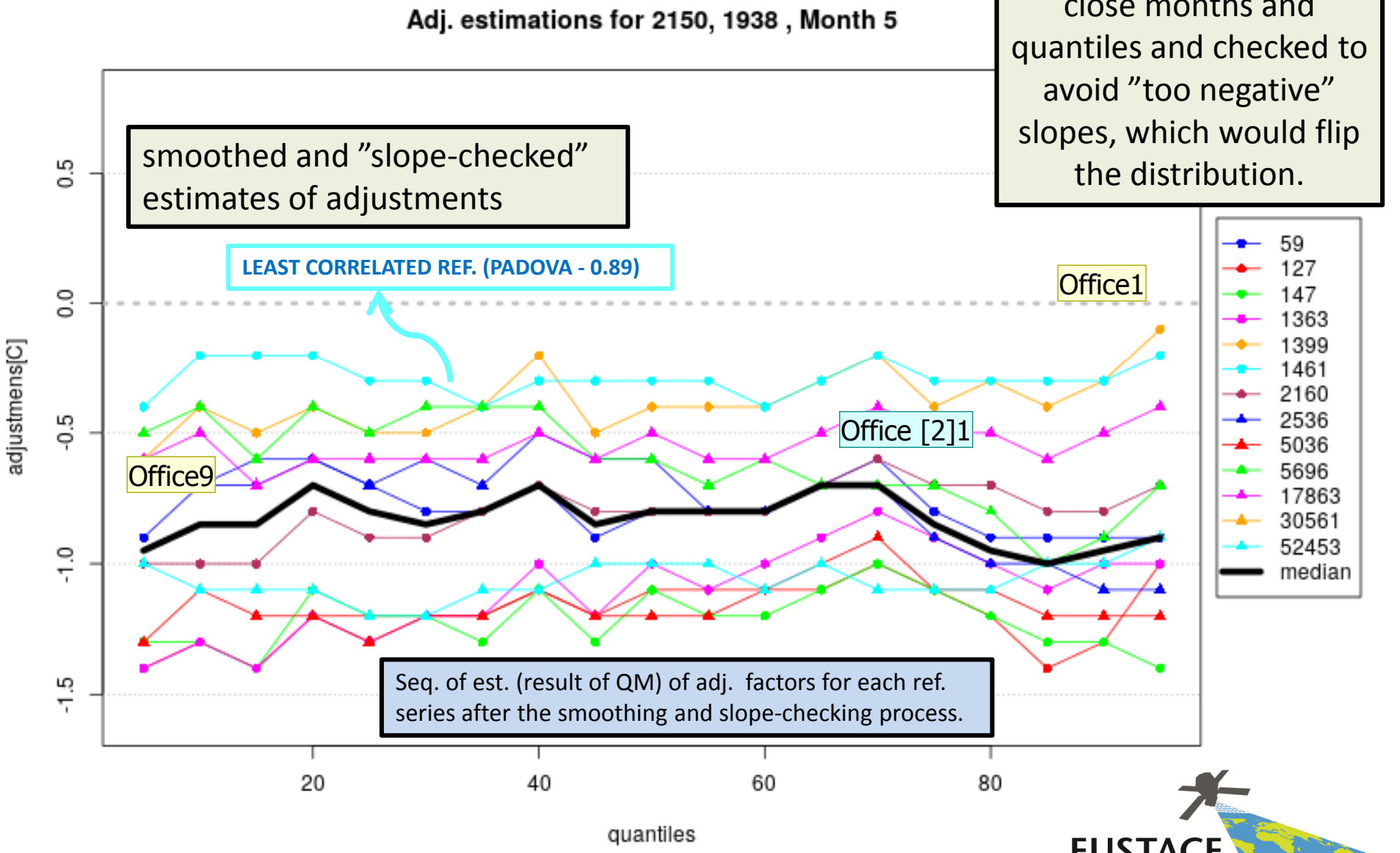
Office6 A set of references in a 3°x3° box is considered. These are split into homogeneous reference sub-series taking into consideration the break detection that has been performed. Pdfs and quantile functions for the same period in the station of Hell are shown here. These indicate the estimation of the climatic signal according to this station. Every reference series provides an estimate of it.
Utente di Microsoft Office; 24.8.2017.

Office7 This equation is "The Difference of The Difference". Quantile functions of same period are compared and the two results are finally compared, since the subtraction of the quantile functions of the reference series to the qunatile functions of the candidate series are supposed to remove the climatic signal. This means that what remains are two quantile functions whose discrepancy is due only to the artificial intervention.
Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

THE ADJUSTMENTS

Estimations related to each reference are smoothed looking at close months and quantiles and checked to avoid ”too negative” slopes, which would flip the distribution.



EUSTACE

Slide 8

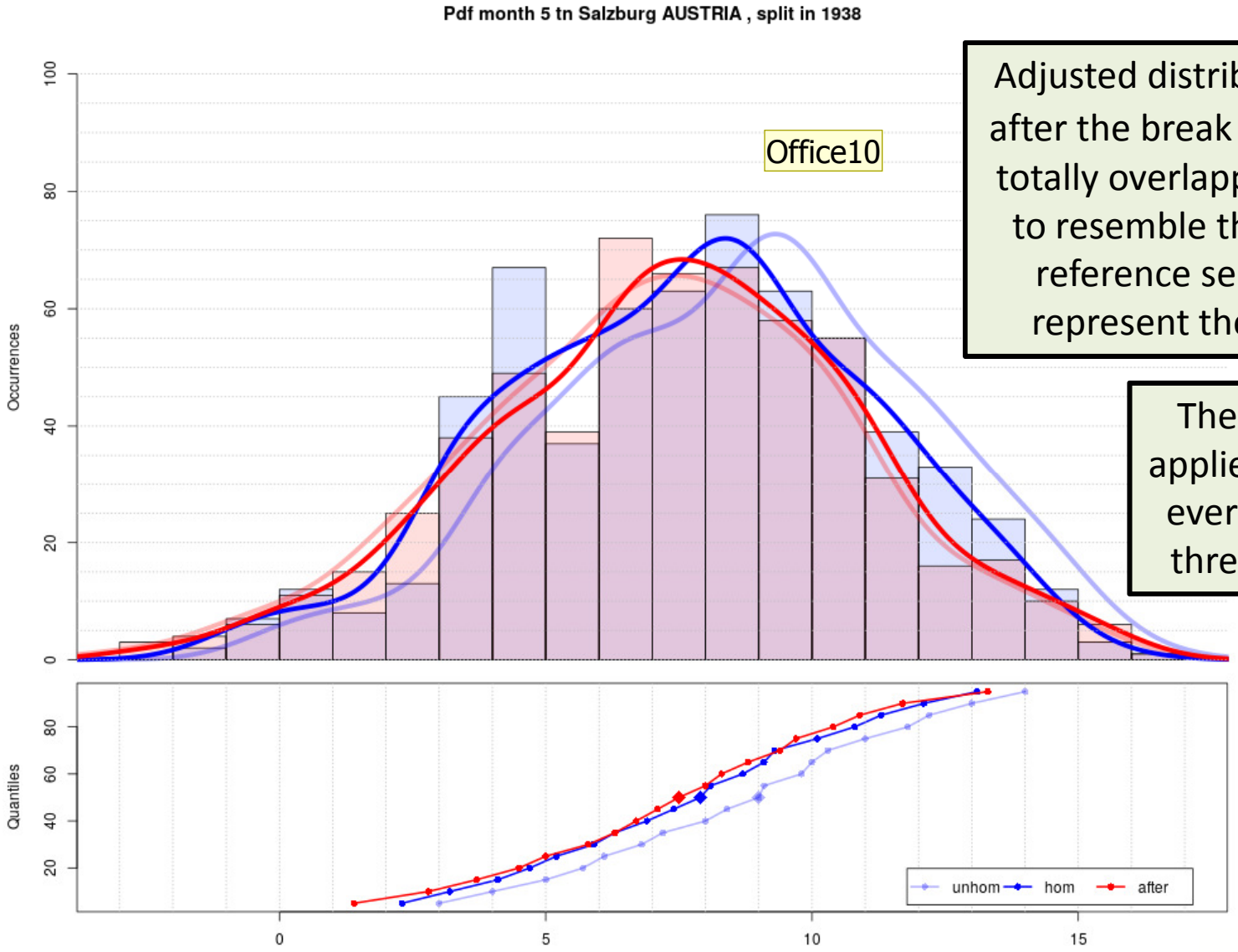
Office1 Negative slope check corrects those estimates whose position cause "too" negative slopes in the sequence. Threshold to identify these values is the difference between the correspondent quantile values, in order to avoid the distribution to flip. In this case the negative slope check hasn't found any critical value.
Utente di Microsoft Office; 23.8.2017.

Office [2]1 The median in this plot represents those very rare cases in which the data to be corrected belongs to the same quantile in every overlapping period with every reference series. In most of the cases data may belong to different (but close) quantiles with the different reference series.
Utente di Microsoft Office; 23.8.2017.

Office9 The smoothing process takes the arithmetic mean of the nearby month and nearby quantile estimation. In this case the first quantile may look as noisier than before, the large correction introduced by the smoothing is due to the fact that estimation for quantile 5 of April and June are way higher than the one in May.
Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

ADJUSTED DISTRIBUTIONS



Adjusted distributions before and after the break **DON'T** have to be totally overlapped, but they have to resemble the features of the reference series, since these represent the climatic signal.

The QM procedure is applied to all months for every breaks if at least three ref. are present.

Hom. temp. pdf and quant. of TN Salzburg (May) 10 years before (blue) and after (red) 1952.
Light blue: unhom. version

EUSTACE

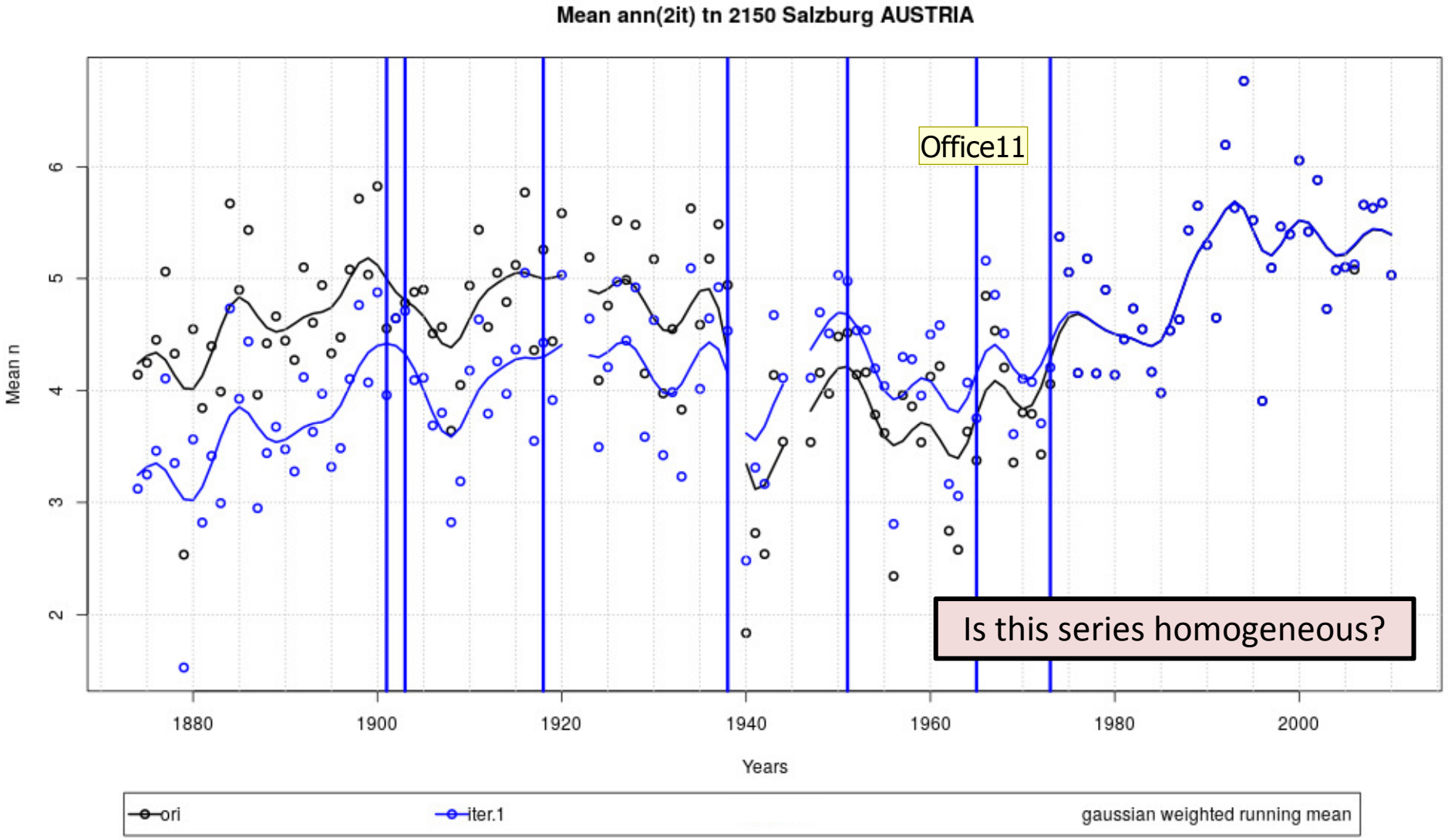
Slide 9

Office10 In this figure, the blue items represent the features of the homogenized part of the series, which is before the break. The old version can be seen in light blue. As described in the captions, the red and the blue distribution are not overlapped because part of the discrepancy that was present before was due to the climatic signal (looks like there was a negative trend in the late 40's and early 50's, which explains the fact that the after-pdf has lower values than the before-pdf)

Utente di Microsoft Office; 24.8.2017.

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

ADJUSTED SERIES



Annual means of TN Salzburg before and after the homogenization with running mean and change points identified during break detection.

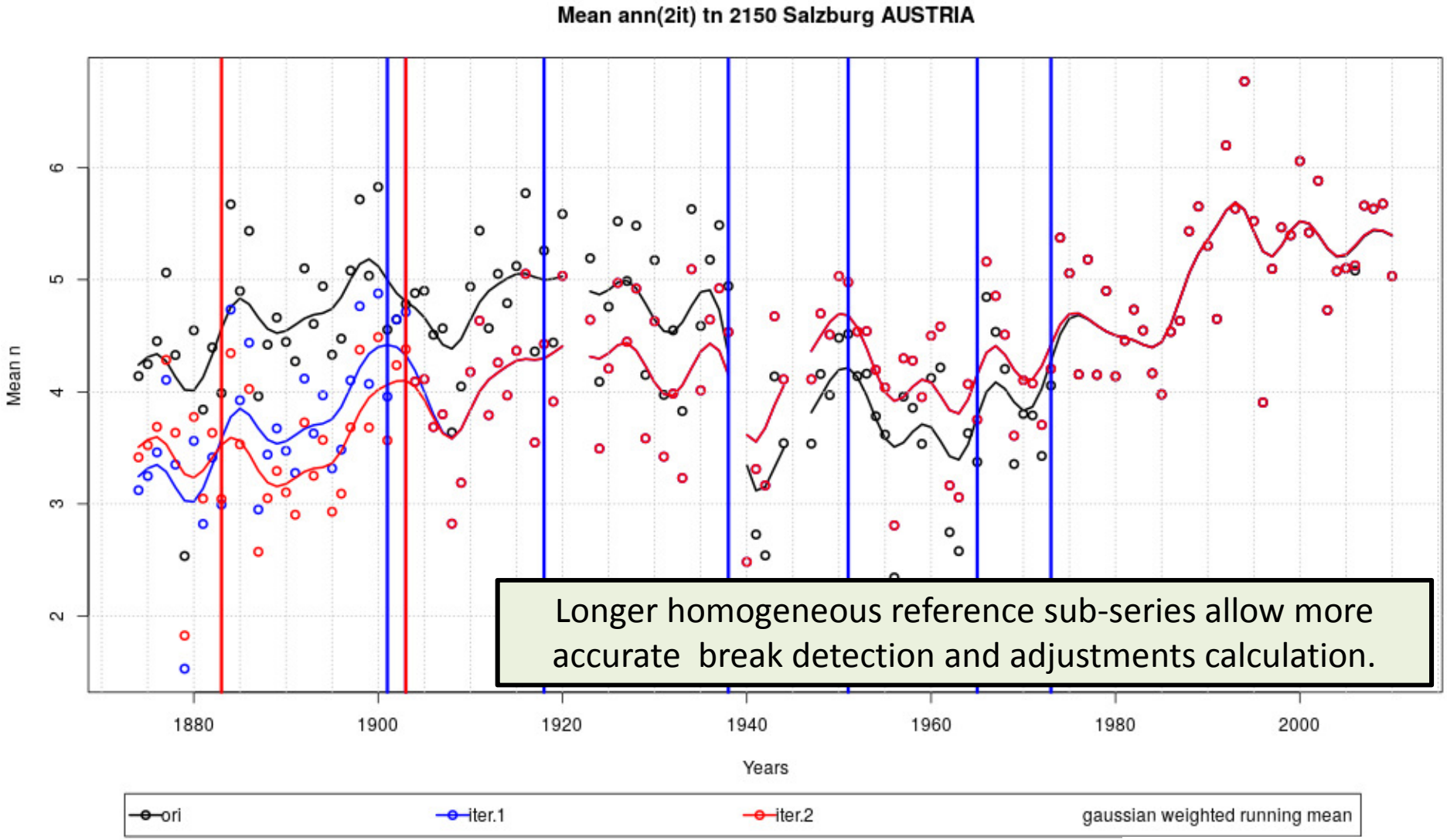
EUSTACE

Office11 The results of adjustment of the series show that the earliest period haven't been corrected (at least in this case, but also a lot of other series have the same problem). Furthermore the splitting of the reference series into homogeneous sub-series causes the loss of data that would have been useful during the adjustment calculation. Hence a second iteration of the homogenization process is performed.
Utente di Microsoft Office; 24.8.2017.

Office12

Meteorological Institute
Ministry of Infrastructure and the Environment

SECOND ITERATION



Longer homogeneous reference sub-series allow more accurate break detection and adjustments calculation.

Ann. means of SalzburgTN through the 2 phases of hom. with running mean and change points identified during the 2 it. of break detection.

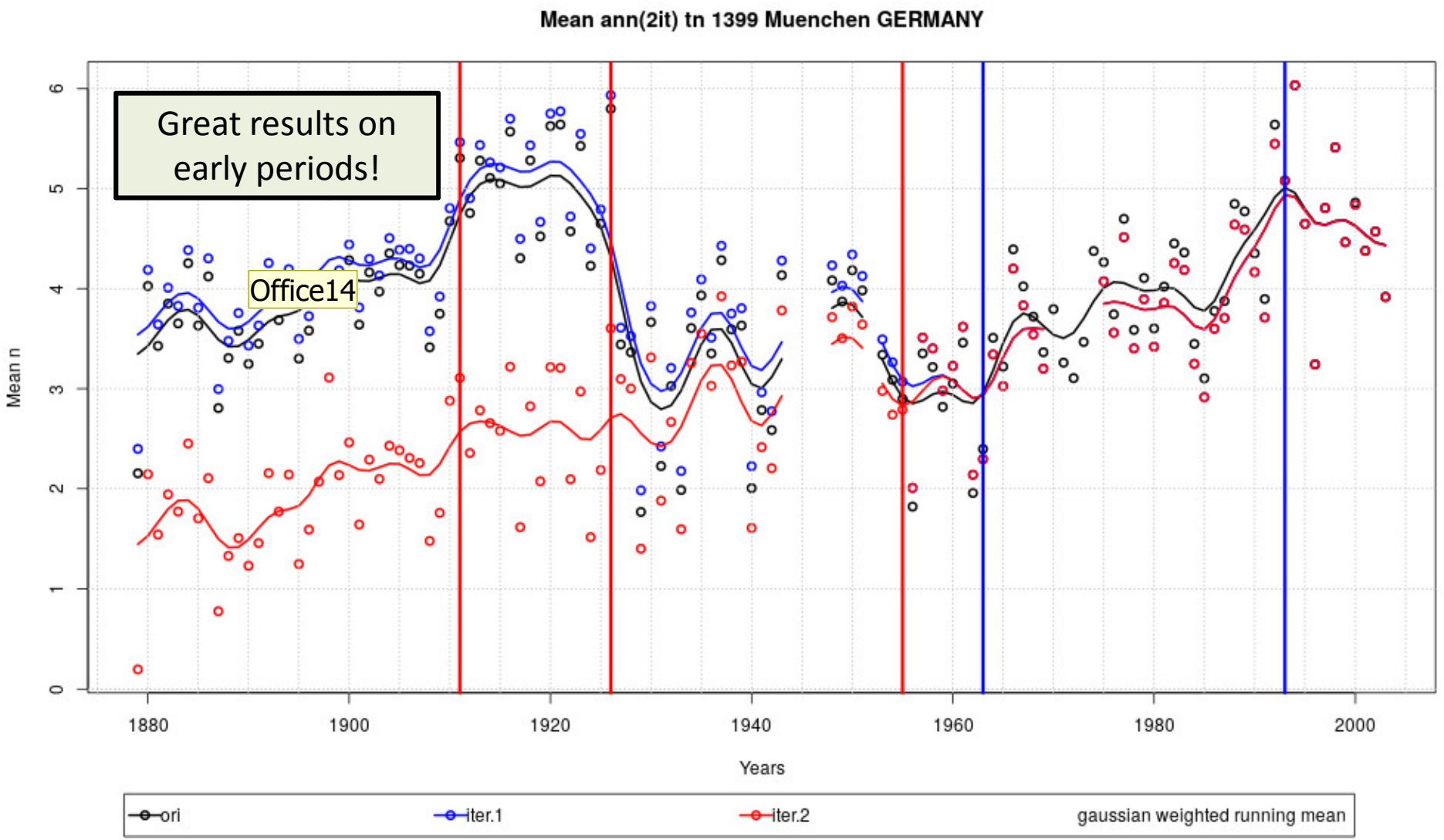
EUSTACE

Slide 11

Office12 In this case the second iteration of homogenization has worked as a perfectioning of the results of the first iteration and unfortunately has not been able to correct the earlier periods. An example of correction of homogeneous series is shown in slide 12 (Munich).
Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

MORE ON SECOND ITERATION



Ann. means of Munich TN through the 2 it . of homogenization with running mean and breaks.

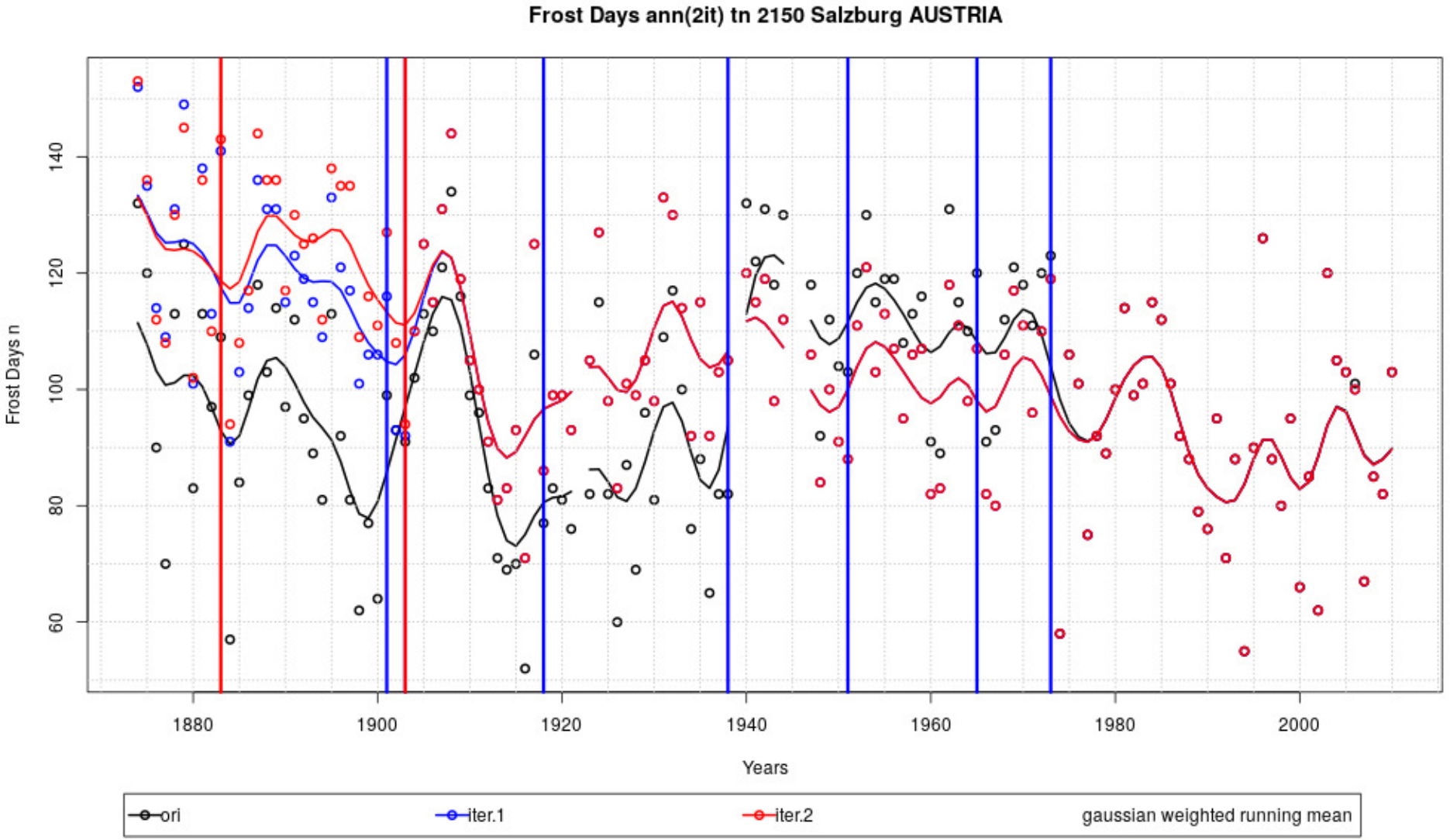
EUSTACE

Office14 This series shows how effective the second iteration has been. The larger availability of long reference series, especially in the earliest periods, allow to correct the oldest parts of the series.
Utente di Microsoft Office; 24.8.2017.

Office13

Meteorological Institute
Ministry of Infrastructure and the Environment

FROST DAYS



Number of **Frost Days** of Salzburg TN through the 2 it. of hom. with running mean and breaks.

EUSTACE

Slide 13

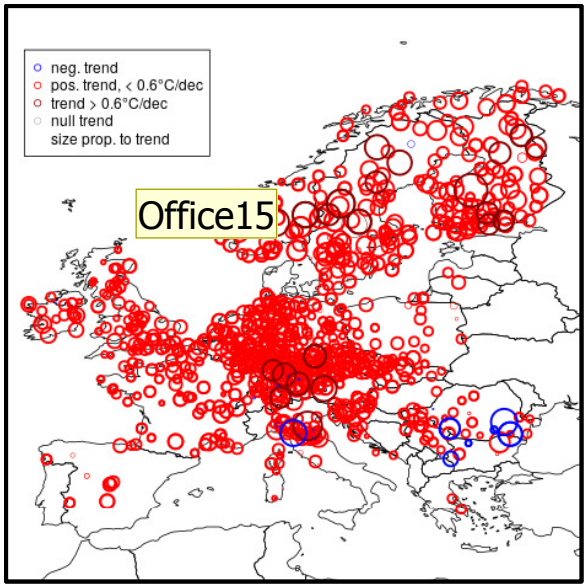
Office13 Homogenization process has been very effective also on extreme values, in this example the series of Frost Days in Bamberg is displayed.
Utente di Microsoft Office; 24.8.2017.

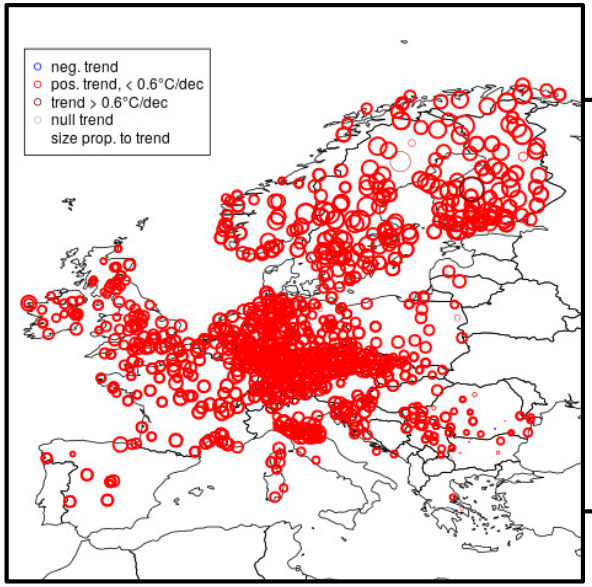
Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

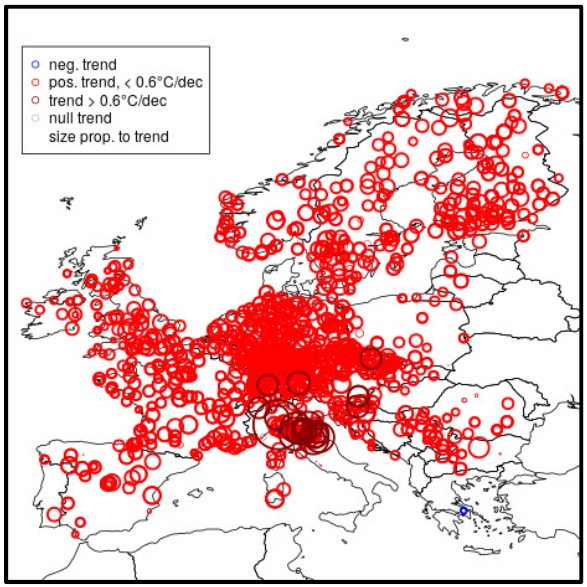
EFFECTS ON TRENDS: ANNUAL MEAN

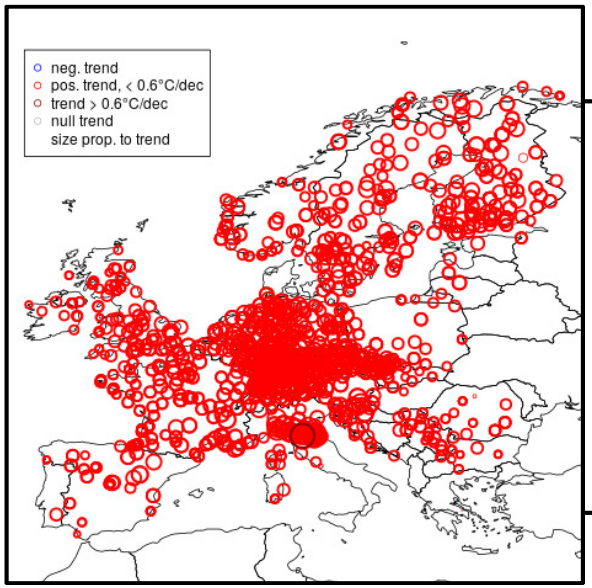
BEFORE HOM

AFTER HOM



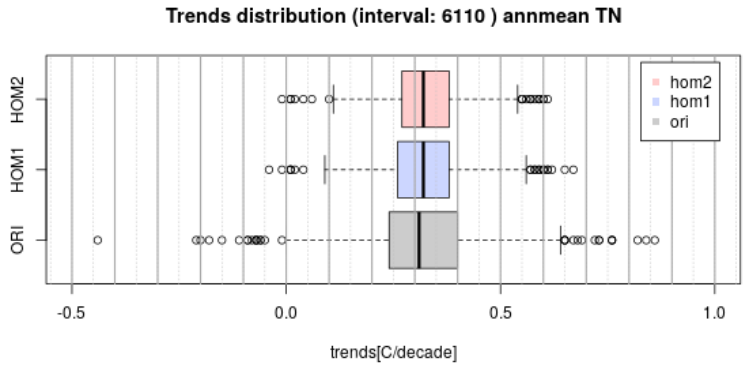




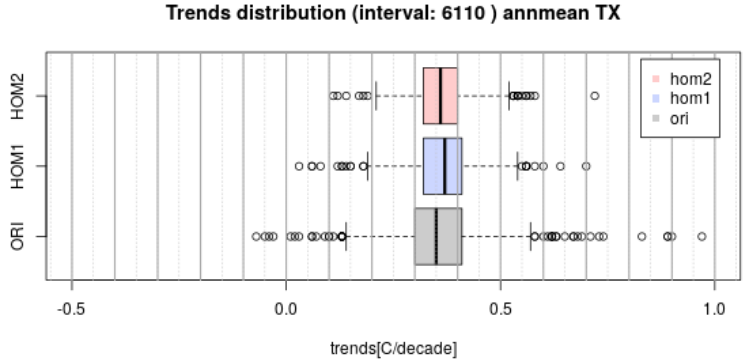


TN

Office16



TX



EUSTACE

Slide 14

Office15 The maps represent the trends on annual mean before (left) and after (right) the homogenization. Blue circles are negative trends, red circles are positive trends, brown circles are very large trends (above 0.6 °C/dec). The threshold for very large trends is chosen as the 95th percentile of the distribution of trends in the non-homogenized dataset. This value (indicated by the "wings" of the boxplots) is almost the same for TN and TX (almost).

Utente di Microsoft Office; 24.8.2017.

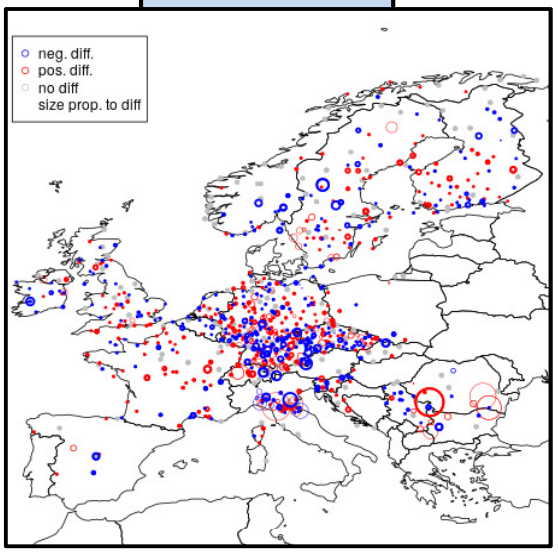
Office16 Removal of blue and brown circles is evident, this effect might be appreciated also in the box plots where the red boxes (second iteration) are clearly narrower than the blue (first iteration) and the black ones (original)

Utente di Microsoft Office; 24.8.2017.

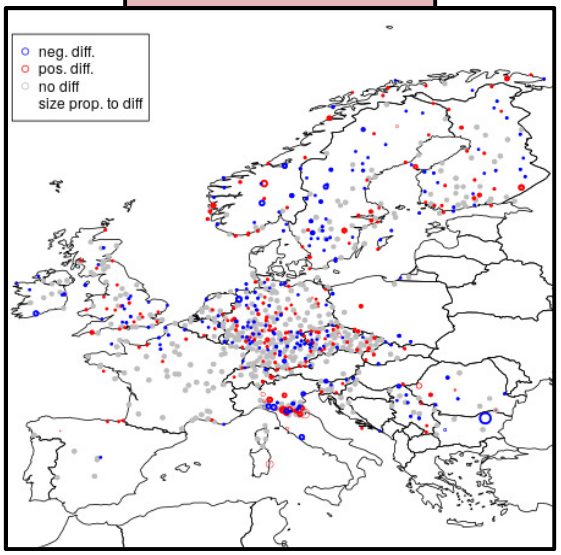
Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

DIFFERENCES IN TRENDS

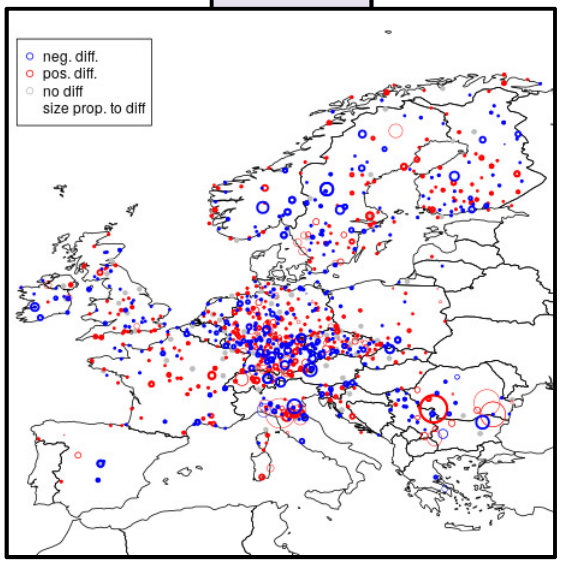
HOM1-ORI

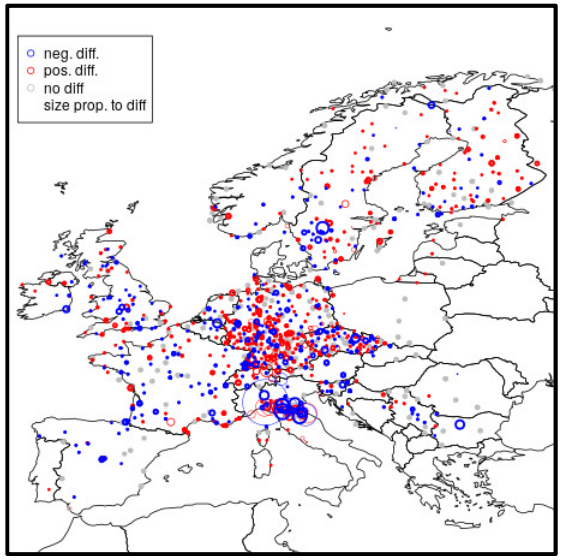


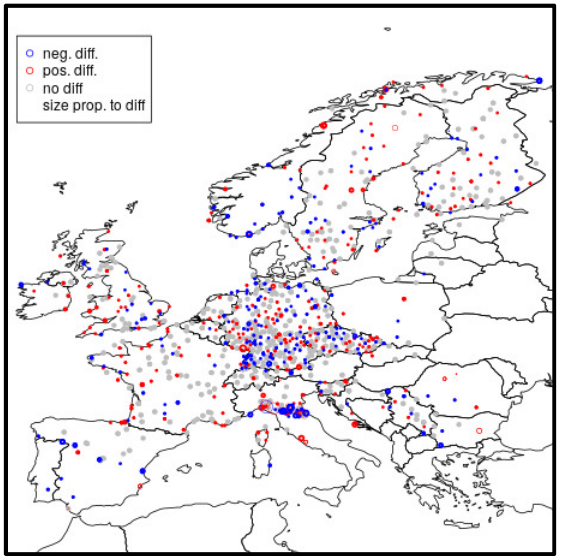
HOM2-HOM1

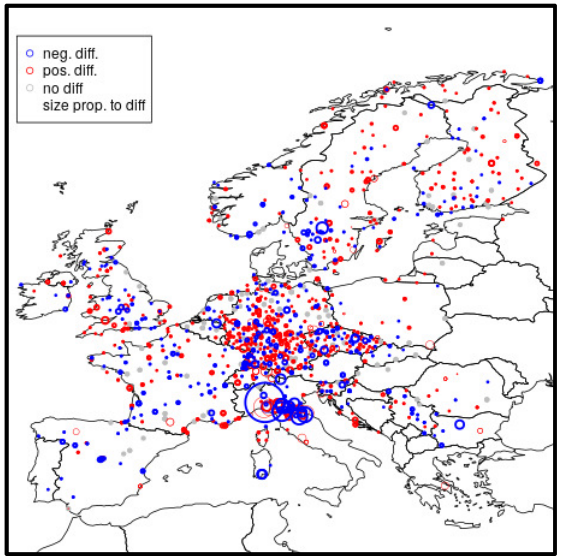


TOTAL









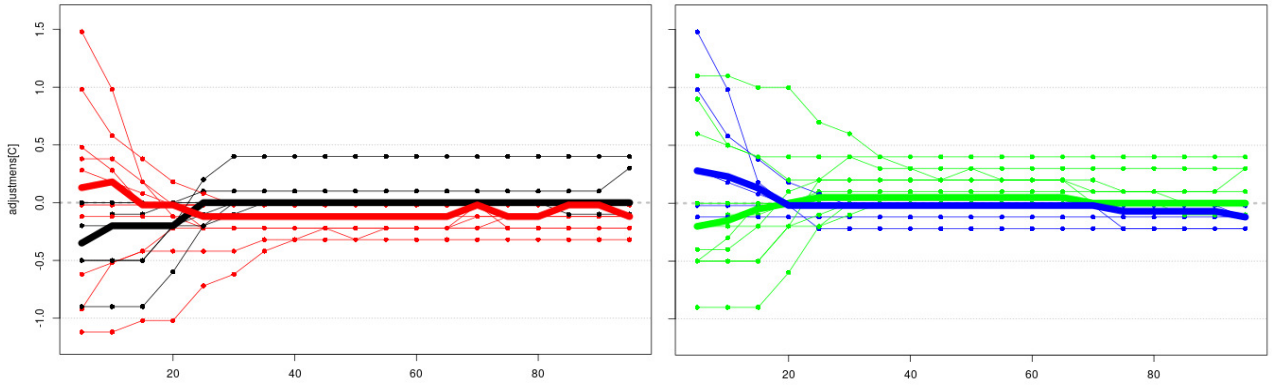
EUSTACE

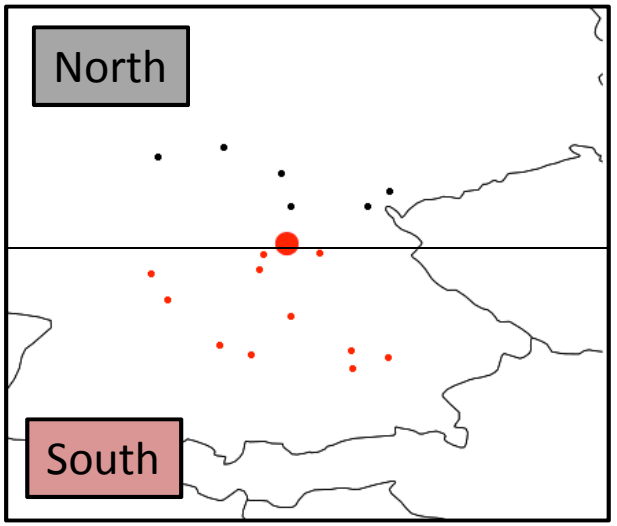
Office21

Meteorological Institute
Ministry of Infrastructure and the Environment

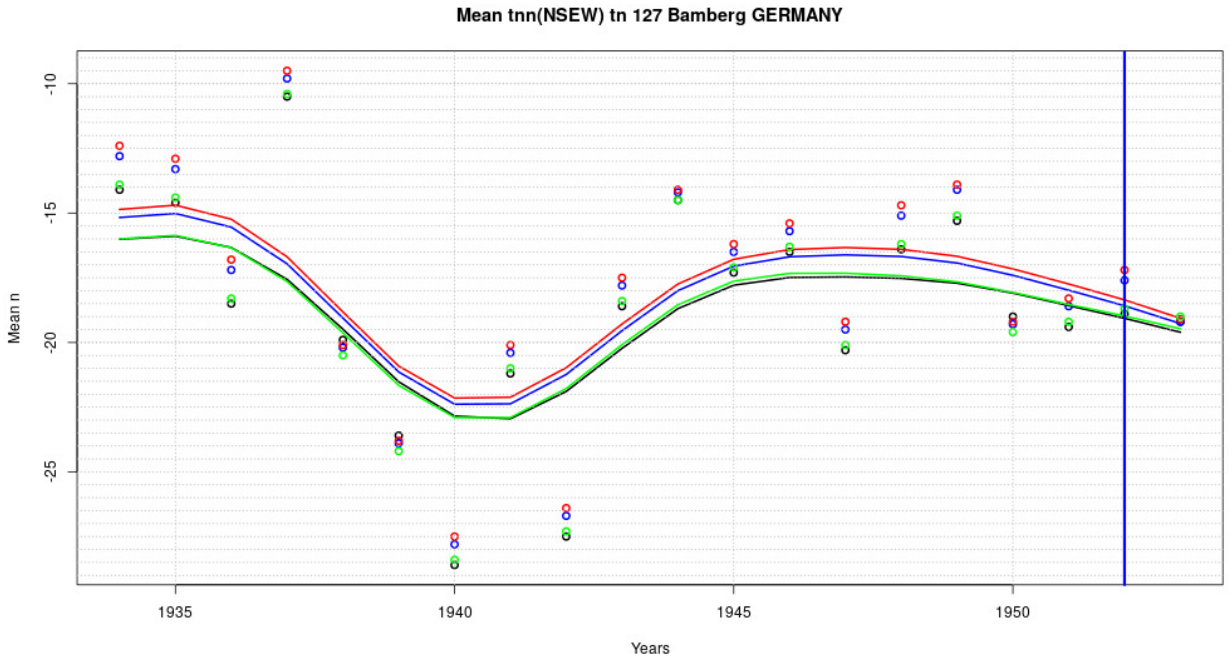
TOWARDS VALIDATION TESTS

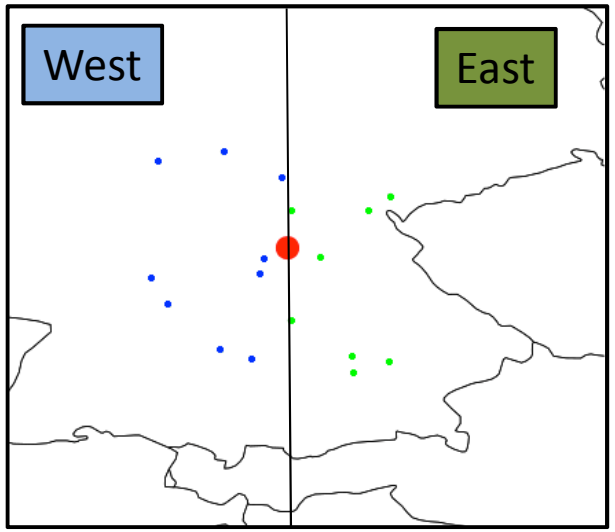
How does the distribution of reference series around the target station affect the results of homogenization?





Est.s of adj.s for the 4 sub-data sets: N (black), S (red), E (green), W (blue).





EUSTACE

Slide 16

Office21 This slide wants to describe how the location of the reference series may affect the results of the adjustment calculation. In particular for the well-known break in 1952 of Bamberg is studied. The references series used above are split first into Northern and Southern data-set, then into Eastern and Western data-sets. Homogenization is run again using only these sub-sets. The plots show how the median of the adjustment changes. For the values in the centre of the sequences the changes are not relevant (around 0.1 ˚C). For the lowest quantiles the difference reaches 0.5 ˚C. These differences have effects especially on the series of TNn (while on the annual mean no big changes are observed). In this case this is maybe due to the orographic features of the region, especially to the presence of the Alps in the south.

Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

CONCLUSIONS

For a reliable calculation of trends, gridded data-sets and more products homogeneous series are indispensible.

Thanks to a process of Break Detection and Quantile Matching Homogenization a homogenized data-set is produced.

High number of portions of the series are not adjusted due to lack of long enough reference homogeneous sub-series.

A 2nd iteration of BD and QM is done, using the outcomes of 1st it. as input.

Clear improvement is obtained on homogeneity of the series in their average and extreme indices.

High spacial consistency is found when looking at the spatial consistency of the trends of the indices.

Dismission and replacement of stations causes series to be short. For the best gridded data and trend calculation, long series are required.

Series from neighbour stations are blended, caring of induced inhomogeneities that are adjusted with a QM procedure

NEXT STEPS TO UNDERTAKE:
- develop a validation system based on manually homogenized series

EUSTACE

Science & Technology Facilities Council

University of Leicester

Met Office Hadley Centre

University of Reading

University of Bath

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

The University of Edinburgh

b UNIVERSITÄT BERN
OESCHGER CENTRE CLIMATE CHANGE RESEARCH

DMI

THANKS FOR YOUR ATTENTION

QUESTIONS?
VRAGEN?
DOMANDE?

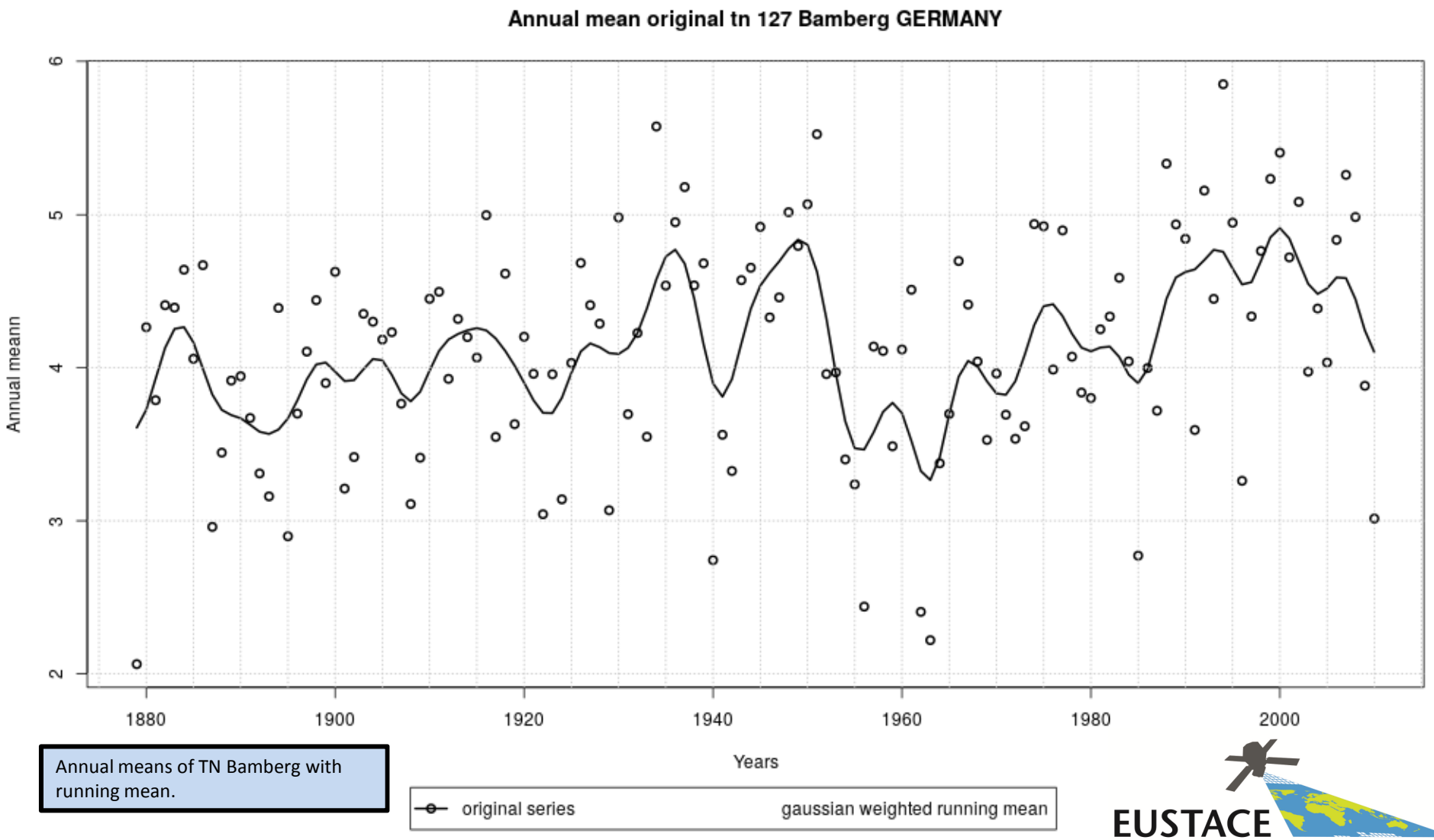
EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

INHOMOGENEOUS SERIES

Undocumented change points affect calculation of trend indices and gridded data



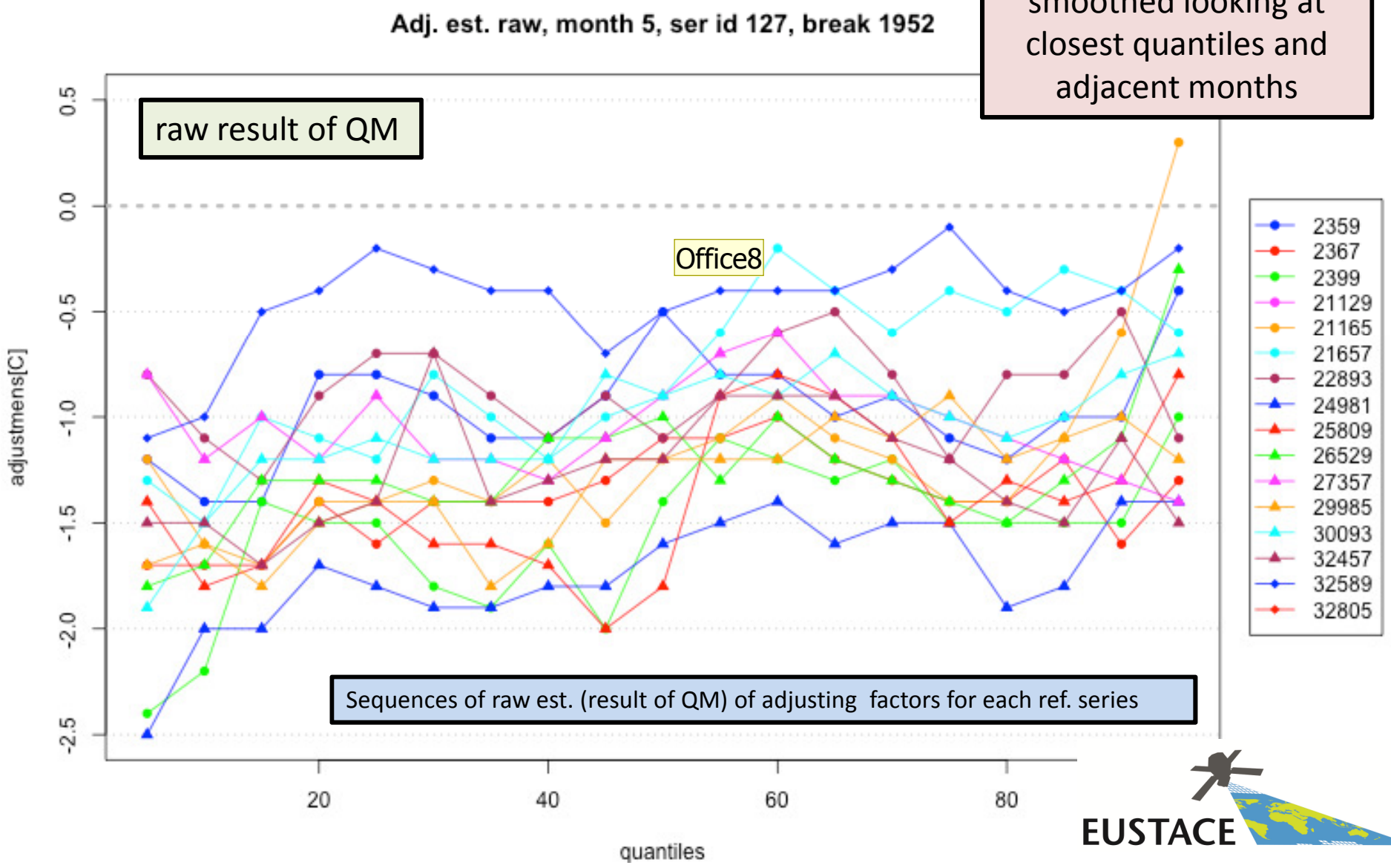
Annual means of TN Bamberg with running mean.

EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

THE ADJUSTMENTS

Estimations related to each reference are very noisy and need to be smoothed looking at closest quantiles and adjacent months



EUSTACE

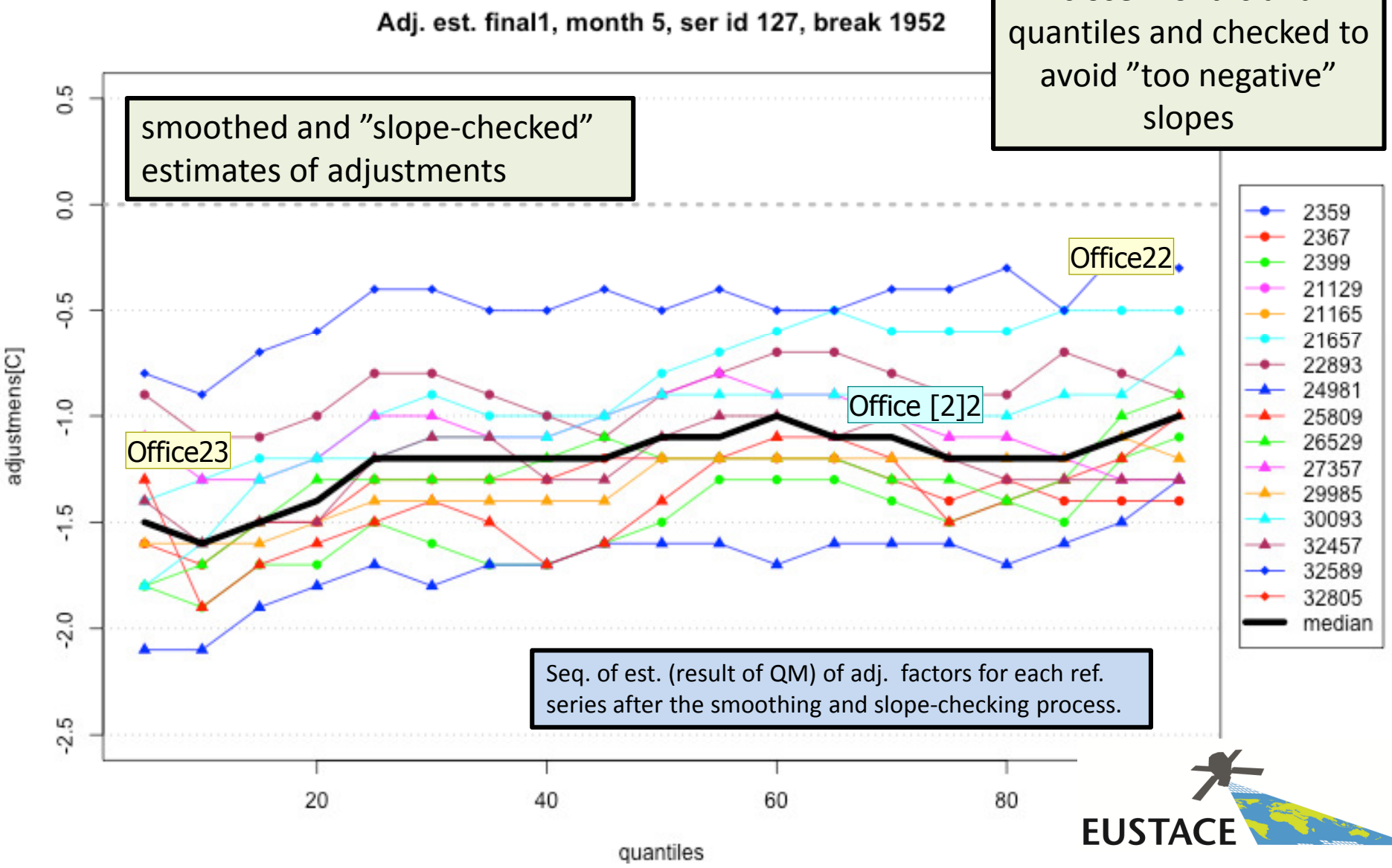
Slide 20

Office8 Each reference series produces a sequence of estimates. These sequences are very noisy and require a process of smoothing.
Utente di Microsoft Office; 24.8.2017.

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

THE ADJUSTMENTS

Estimations related to each reference are smoothed looking at close months and quantiles and checked to avoid ”too negative” slopes



EUSTACE

Slide 21

Office22 Negative slope check corrects those estimates whose position cause "too" negative slopes in the sequence. Threshold to identify these values is the difference between the correspondent quantile values, in order to avoid the distribution to flip. In this case the negative slope check hasn't found any critical value.
Utente di Microsoft Office; 23.8.2017.

Office [2]2 The median in this plot represents those very rare cases in which the data to be corrected belongs to the same quantile in every overlapping period with every reference series. In most of the cases data may belong to different (but close) quantiles with the different reference series.
Utente di Microsoft Office; 23.8.2017.

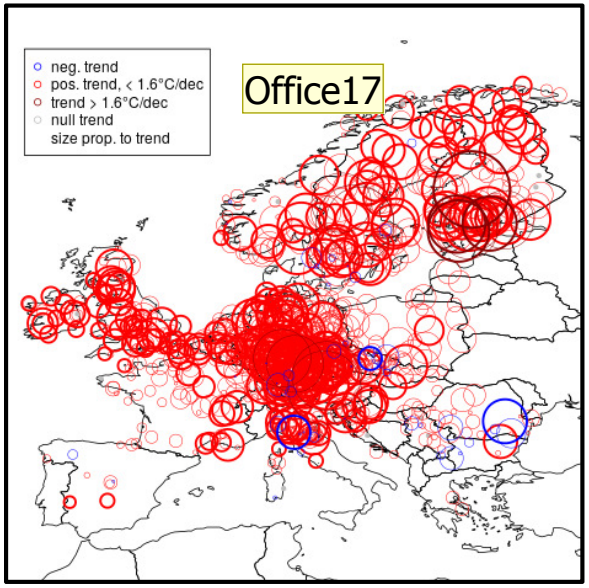
Office23 The smoothing process takes the arithmetic mean of the nearby month and nearby quantile estimation. In this case the first quantile may look as noisier than before, the large correction introduced by the smoothing is due to the fact that estimation for quantile 5 of April and June are way higher than the one in May.
Utente di Microsoft Office; 24.8.2017.

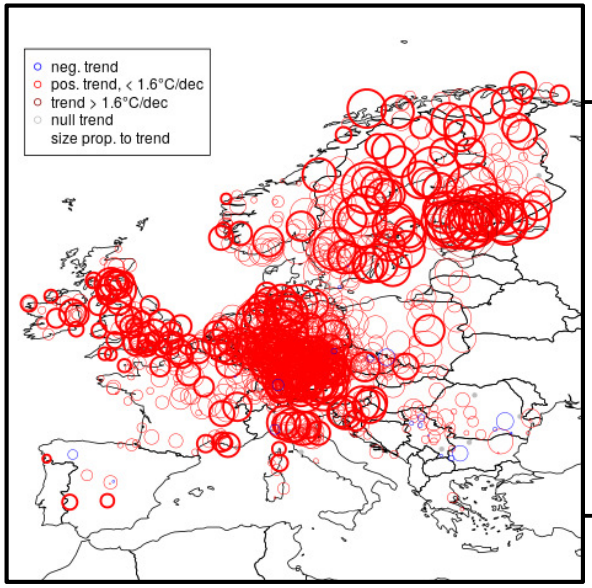
Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

EFFECTS ON TRENDS: TNN AND TXX

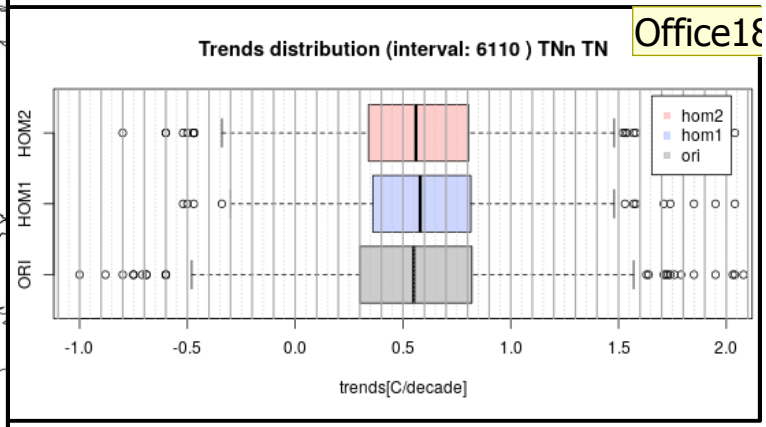
BEFORE HOM

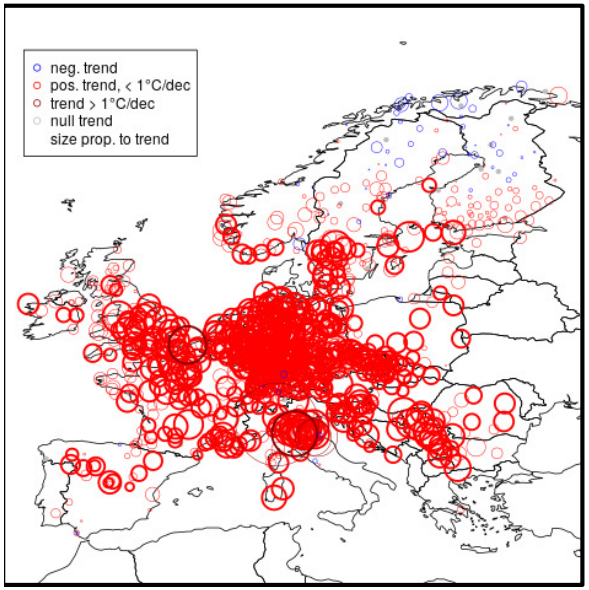
AFTER HOM

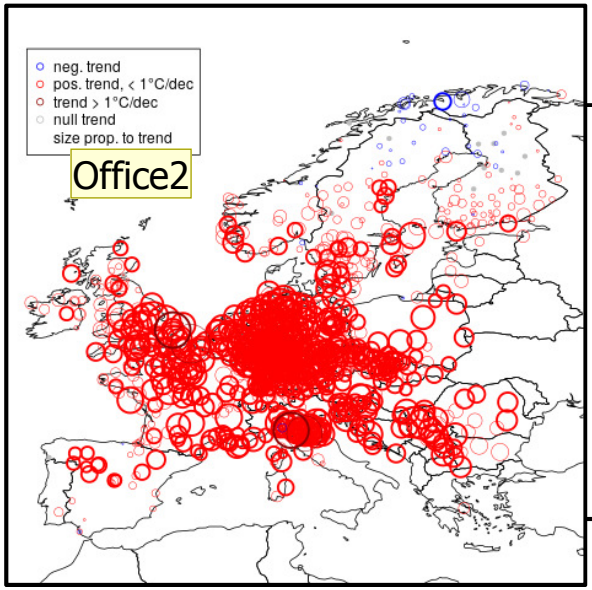




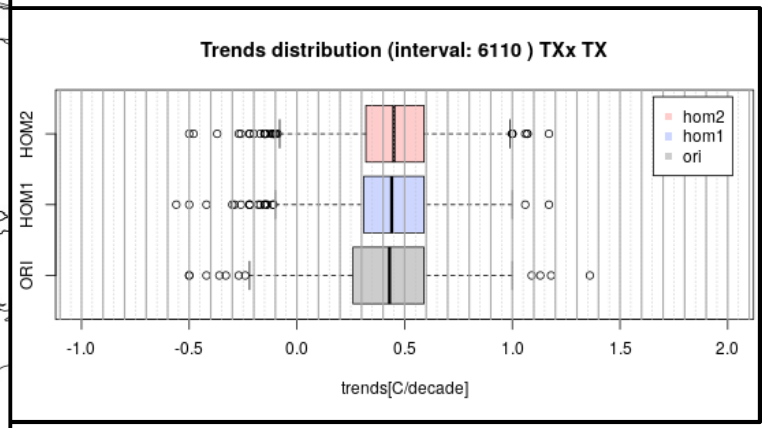
TNn







TXx



EUSTACE

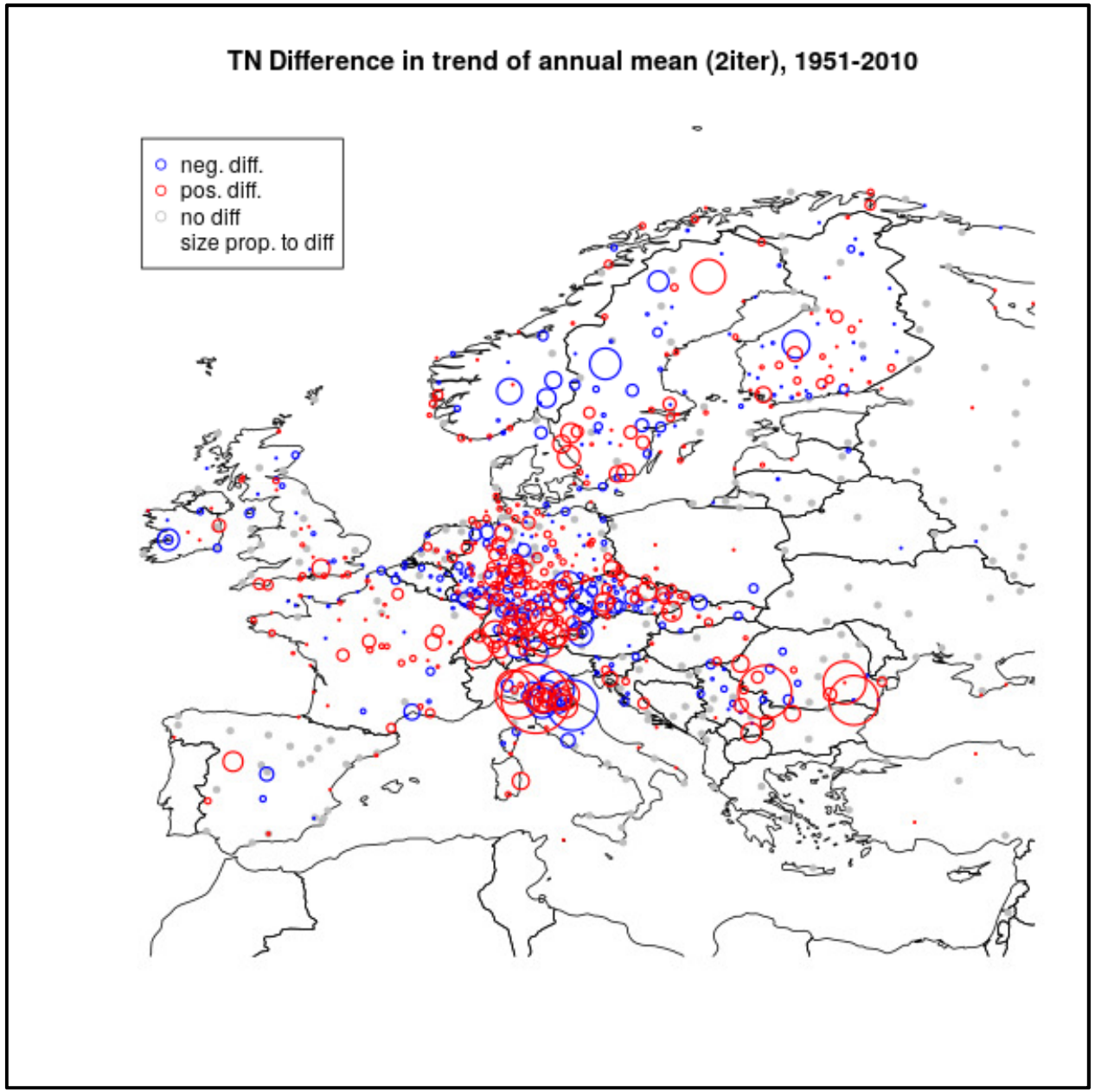
Slide 22

Office2 Utente di Microsoft Office; 24.8.2017.

Office17 Same as previous slide but for TXx and TNn, some circles are shaded since the trends they display are not significant. Some particular behaviors are noted in Romania/Bulgaria (low trends) and Scandinavia (large trends for TN and low trends for TX), this case will be inspected further.
Utente di Microsoft Office; 24.8.2017.

Office18 Boxplots in this case are way larger than for annual mean, especially TNn is affected by a high variance. Also the general trend for TNn looks to be generally larger, indicating a narrowing and skewing effect on the pdfs.
Utente di Microsoft Office; 24.8.2017.

DIFFERENCE IN TRENDS



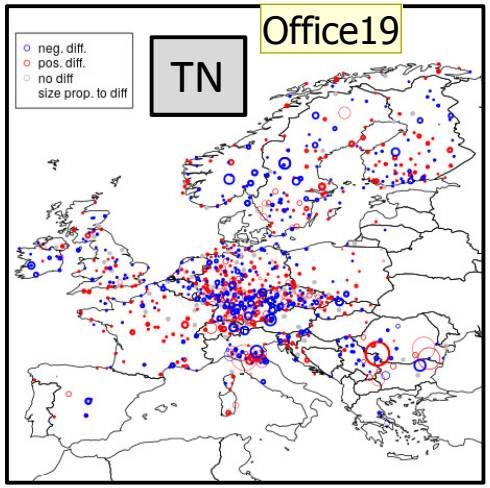
Difference signs in the corrections.
The aim is not to have warmer trends, therefore corrections on trends can also be negative

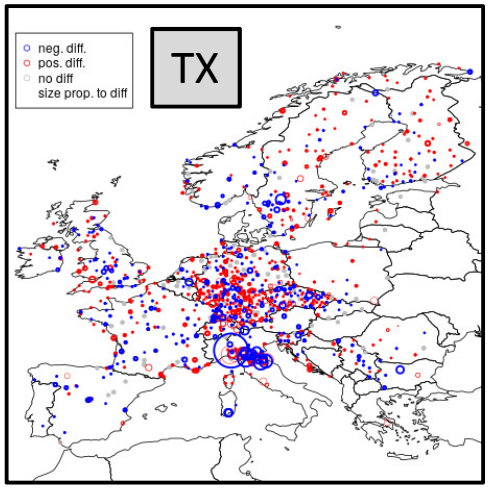
EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

SOMETHING MORE ON TRENDS

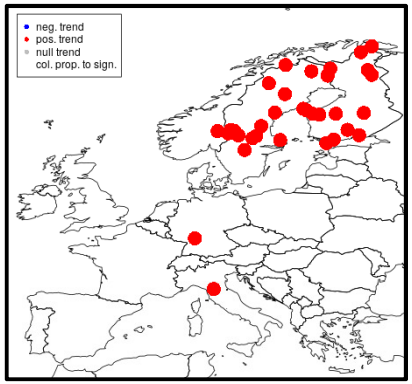
DIFFERENCE ANN MEAN

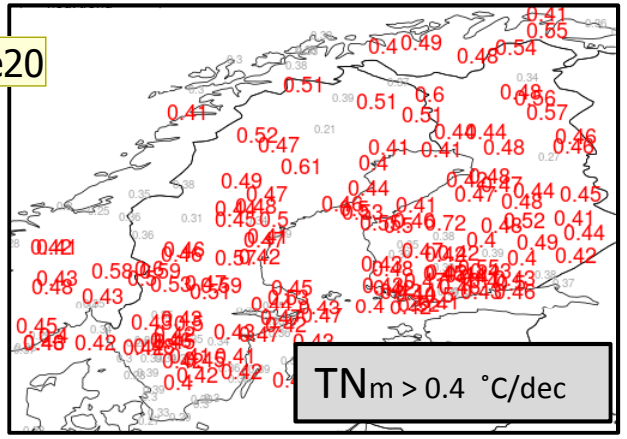




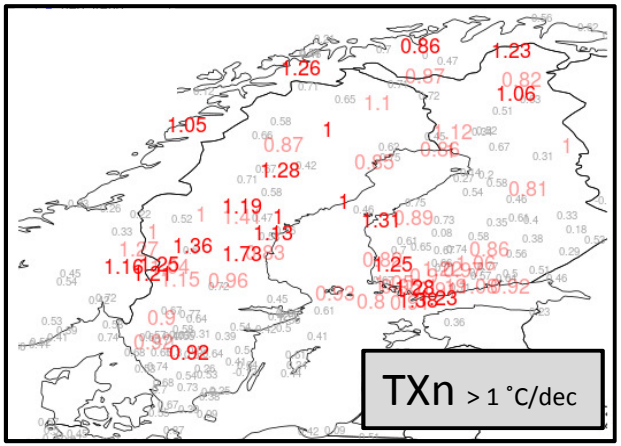
SCANDINAVIAN STATIONS

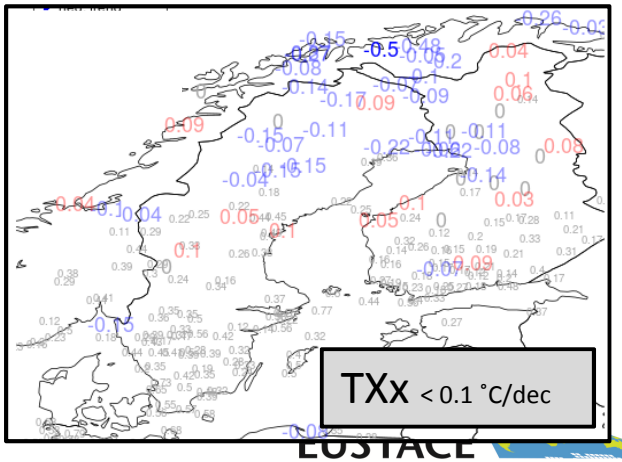
Series with trend on ann. mean of TN > 0.5 ˚C/dec are mostly located in Scandinavia, same behaviour for TNn





But for TX extremes:



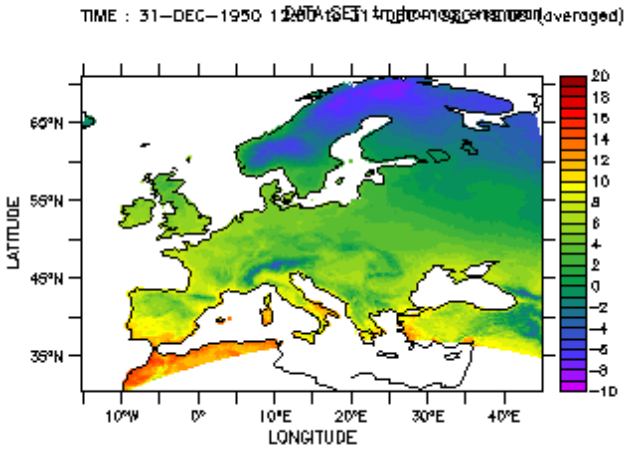


Slide 24

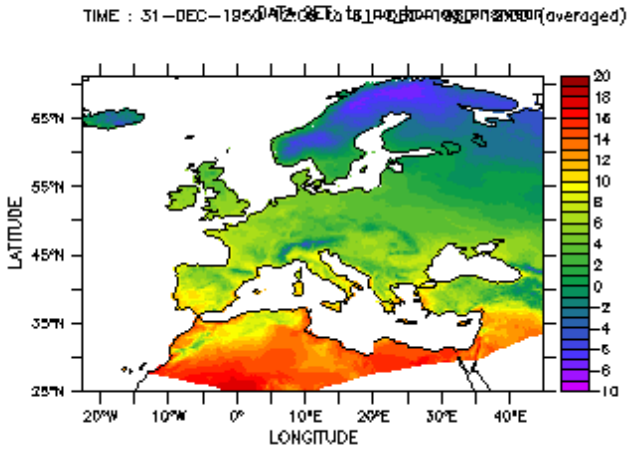
Office19 Differences in the trends between second iteration and original data set show how the corrections are not uniquely aimed at the remotion of negative trends. Indeed they correction are in all the directions, depending on the kind of inhomogeneity and of reference series.
Utente di Microsoft Office; 24.8.2017.

Office20 Trends on Scandinavian stations have peculiar behaviors: mean of TN, TNn and TXn show very large trends, the largest of whole Europe (see top left map). At the same time, the trends on TXx appear to be very low, even negative, but almost always not significant. This means that in this area of Europe the narrowing of the pdf is even stronger.
Utente di Microsoft Office; 24.8.2017.

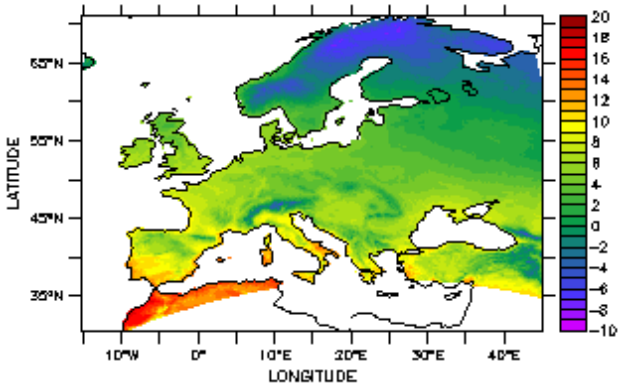
GRIDDED DATASET - TN



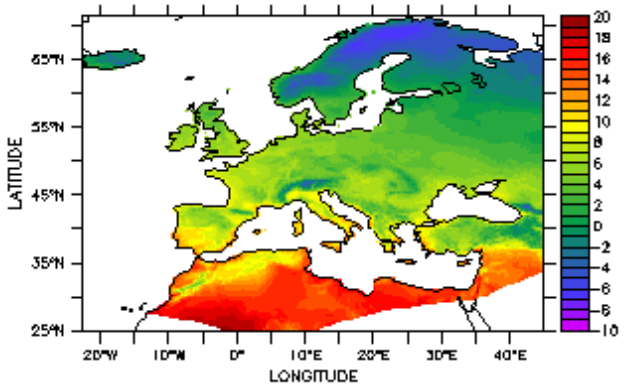
minimum temperature (Celsius)



minimum temperature (Celsius)



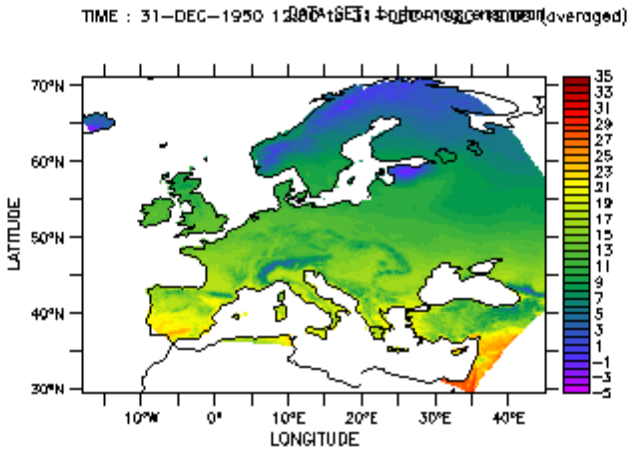
minimum temperature (Celsius)



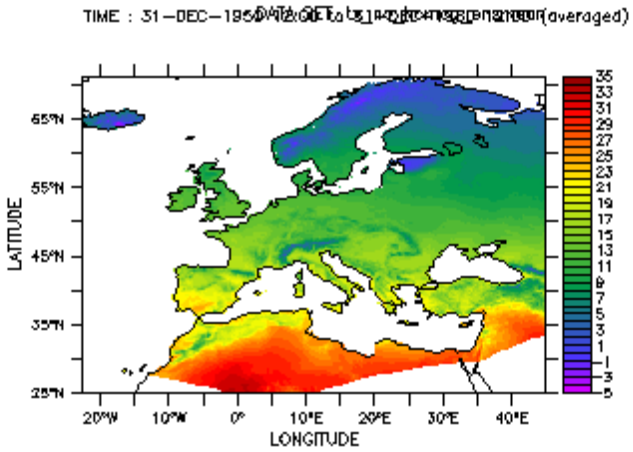
minimum temperature (Celsius)

EUSTACE

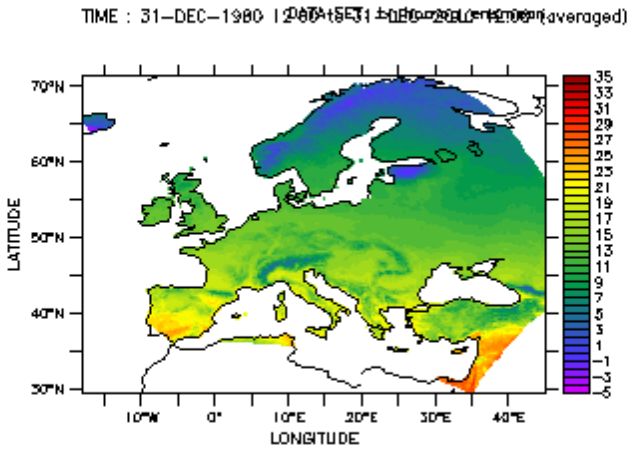
GRIDDED DATASET - TX



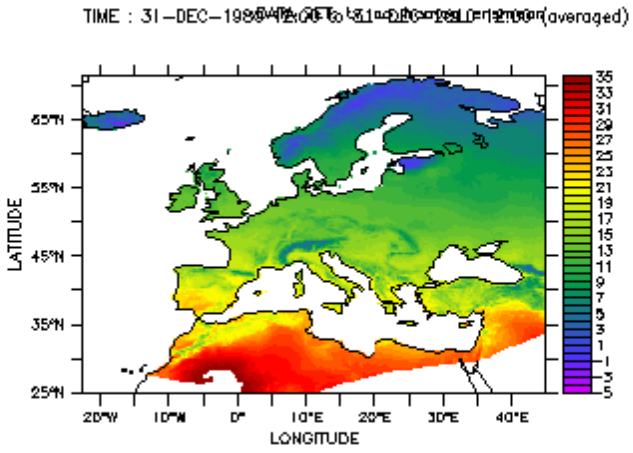
maximum temperature (Celsius)



maximum temperature (Celsius)



maximum temperature (Celsius)



maximum temperature (Celsius)

EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

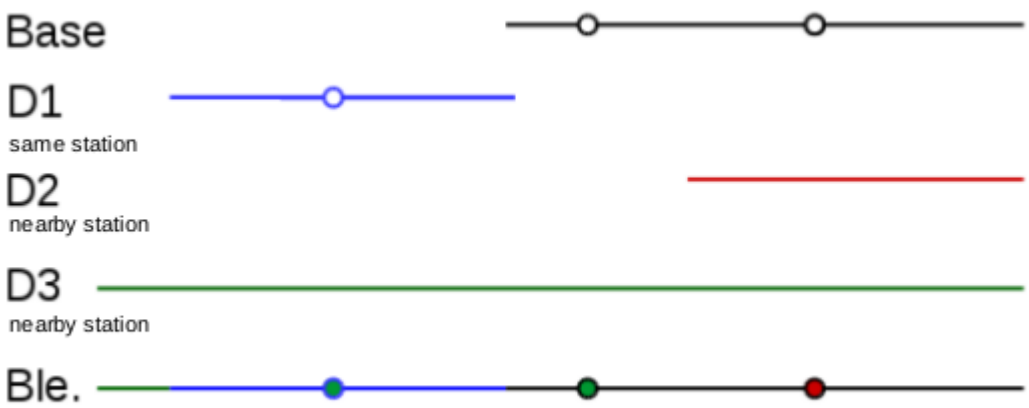
BLENDING OF SERIES

Hom. series from same station and from stations at a maximal distance of 12.5 km and height difference of no more than 100 meters are blended.

Very useful in case of dismissed or moved stations (for example from the centre to the airport).

Latest ending series from target station is selected as “base”, the other ones as “donating”

Blended series are longer and eventual gaps are filled with data from donating series.

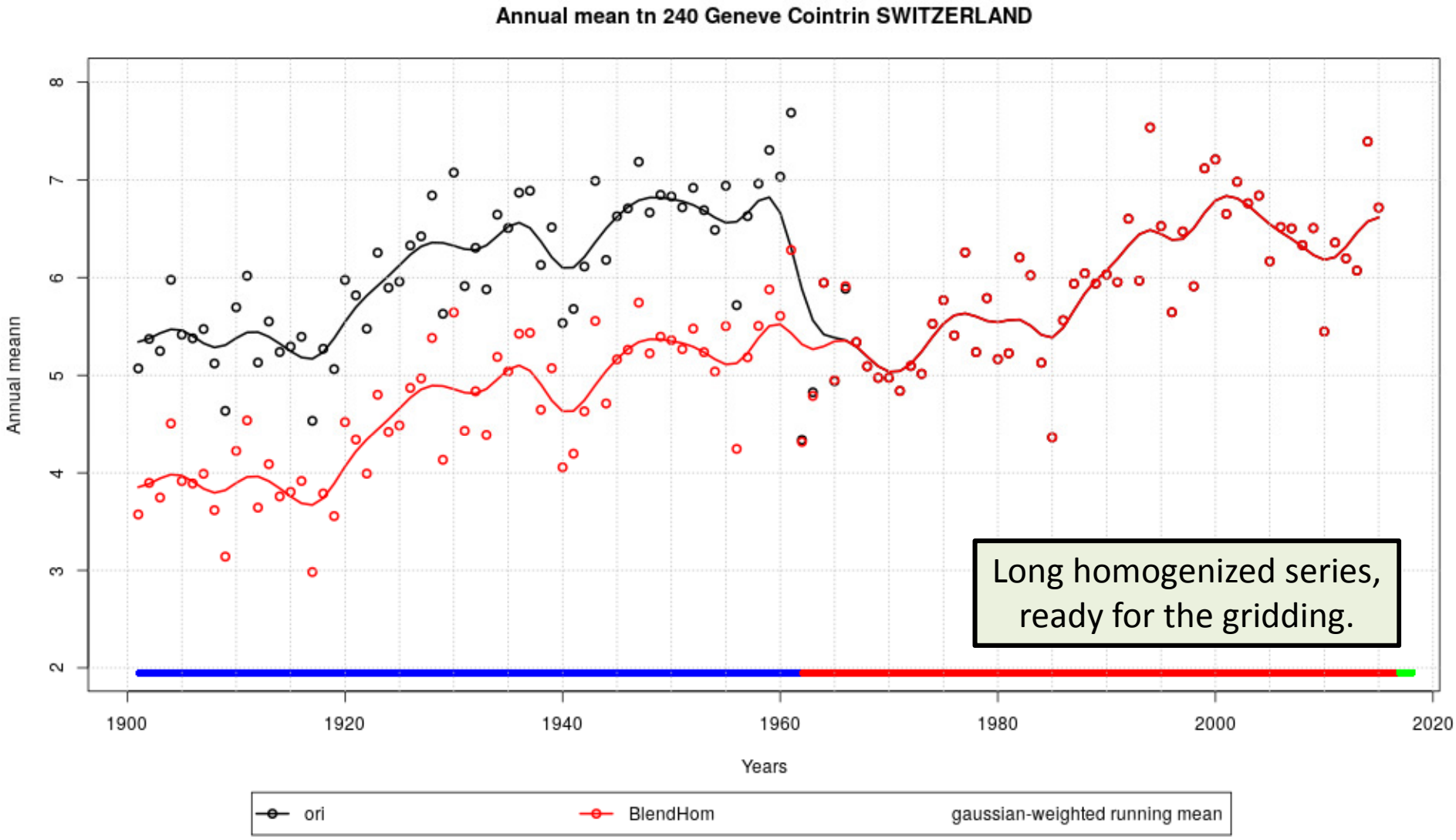


Though, blending procedure introduces new inhomogeneities that need to be adjusted with an “ad-hoc” homogenization process

EUSTACE

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

HOMBLENDED SERIES



Ann. means of TN Geneva before and after the blend. hom. with running mean. Don. series are identified by the colors: Geneva(blue), Airport (red), synoptical extension to now (green)

EUSTACE