On homogenization of Portuguese meteorological and geophysical data series

A.L. Morozova[1], P. Pribeiro[1], M.A.Valente[2], M.A. Pais[1]

[1] **Centro de Geofísica da Universidade de Coimbra**

[2] **Instituto Dom Luiz, University of Lisbon**

7ª Asamblea Hispano Portuguesa de Geodesia y Geofísica

7ª Assembléia Hispano Portuguesa de Geodesia e Geofísica

Donostia - San Sebastián, del 25 al 28 de junio de 2012

CGUC CENTRO DE GEOFÍSICA UNIVERSIDADE DE COIMBRA

Data description

- Geophysical Institute of the University of Coimbra (IGUC): $\varphi = 40^{o}12'$ N, $\lambda = 8^{o}25'$ W, h = 141 m
- digitized monthly series of temperature parameters (*Tmin*, *Tmax* and *mean daily T*): from 1865 to 2005
- digitized daily series of temperature parameters (*Tmin*, *Tmax* and *mean daily T*): from 1941 to 2005
- digitized monthly series of geomagnetic *K-index* computed in IGUC: from 1952 to 2005

Main ideas

- *There are*: artificial shifts due to the changes in the instruments, instruments positions, measurement procedures and calculation methods.
- *We need*: <u>a study of homogeneity level</u> of the meteorological and geomagnetic data series and <u>a correction</u> for artificial shifts
- *We can use*: a number of absolute and relative homogeneity tests; metadata
- *We will get*: a homogenized temperature and K-index series with predefined significance level (e.g. 95%)

Steps

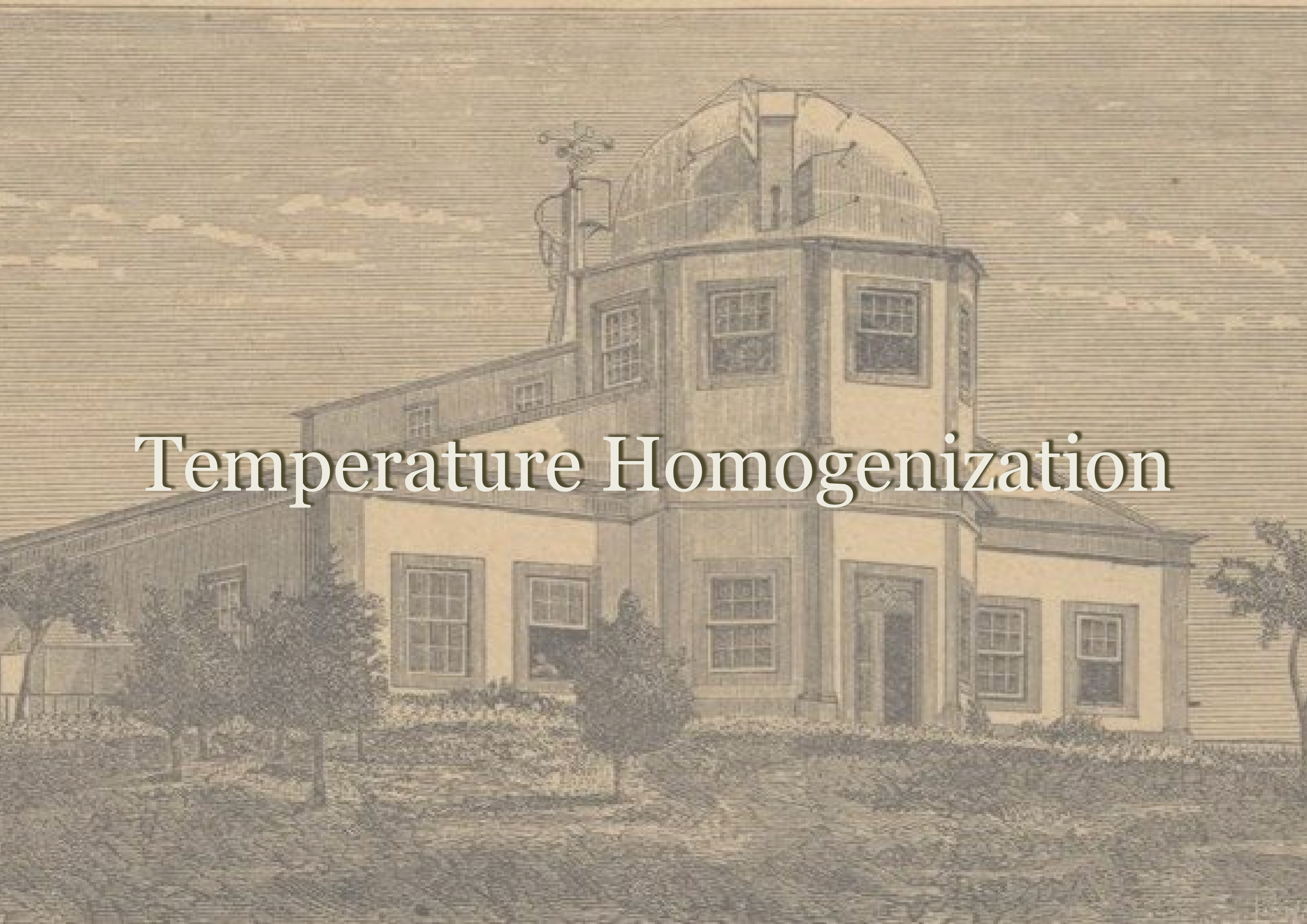
- Visual analysis
- Statistical homogeneity tests – absolute and relative
- Study of the station's records and logbooks – *metadata*

- Correction of artificial homogeneity breaks: maximization of the number of the monthly series that
  - do not have significant peaks in absolute and relative homogeneity tests statistics around non-climatic breaks
  - do not increase the centered root mean square errors (CRMSE) of the corrected series comparing to the CRMSEs of the original data (*CRMSEs are calculated using reference series*)

- Visual analysis & homogeneity tests of corrected series

Homogeneity tests

- likelihood ratio standard normal homogeneity test (SNHT)
- parametric Buishand cumulative deviation test
- non-parametric rank Pettitt test (*only for temperature data*)



Temperature Homogenization

Metadata for meteorological station

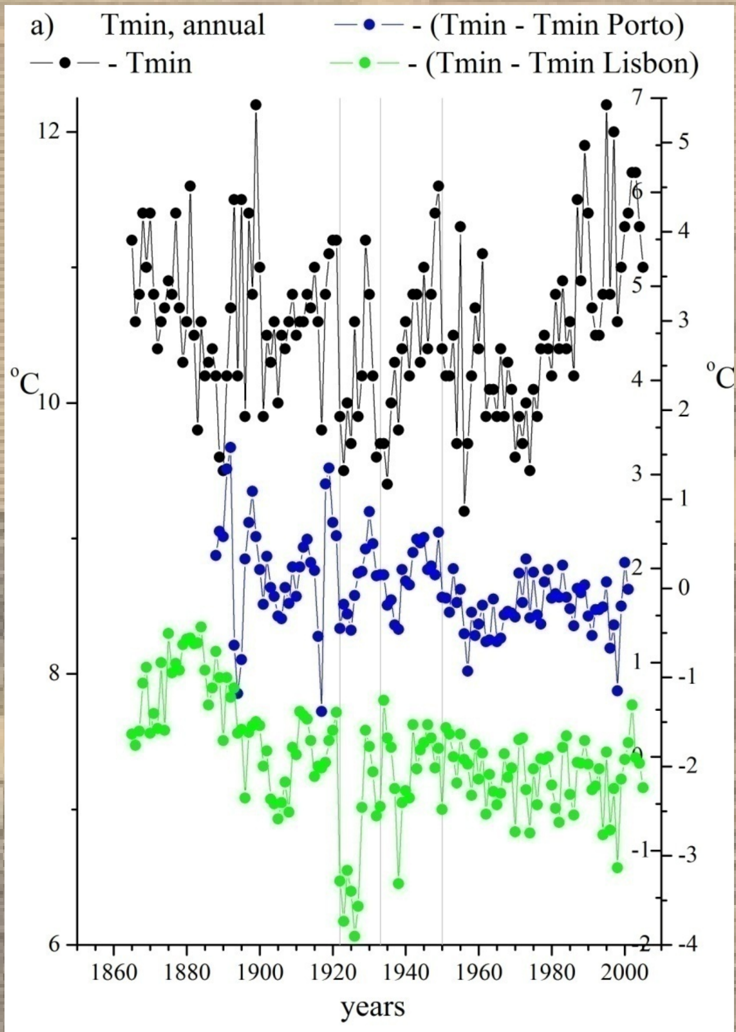
- 1922 - relocation of the instruments set and installation of the standard shelter
- 1933 - small relocation
- 1950 - change of the thermometer height (from 1.15 m to 1.45 m)

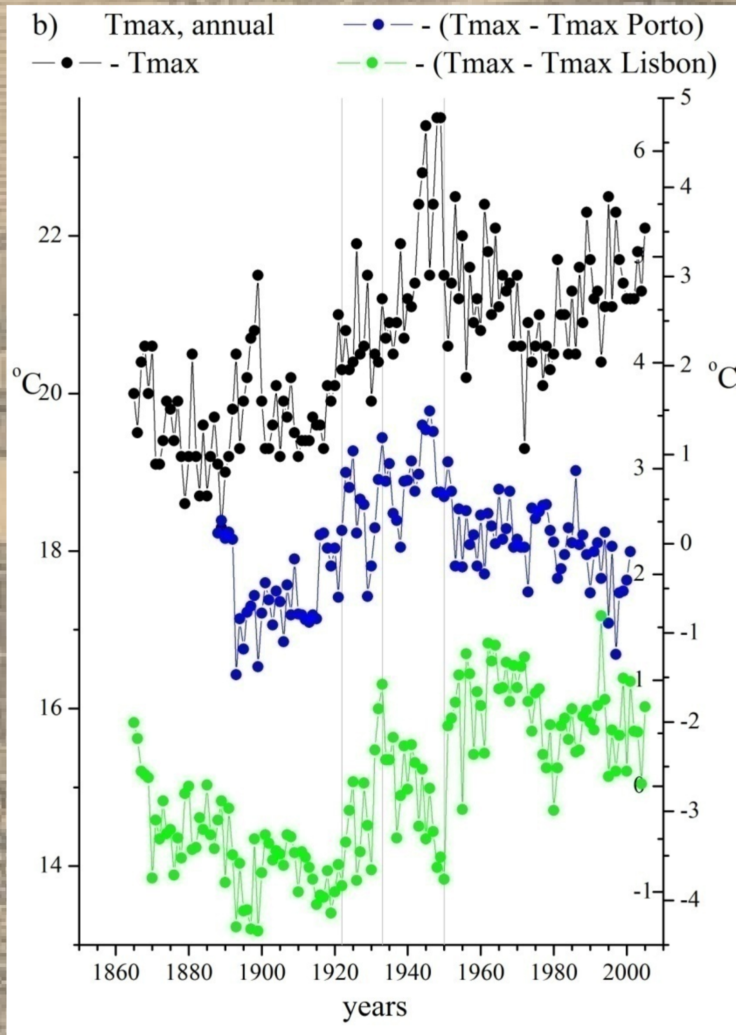
Temperature Homogenization: parameters

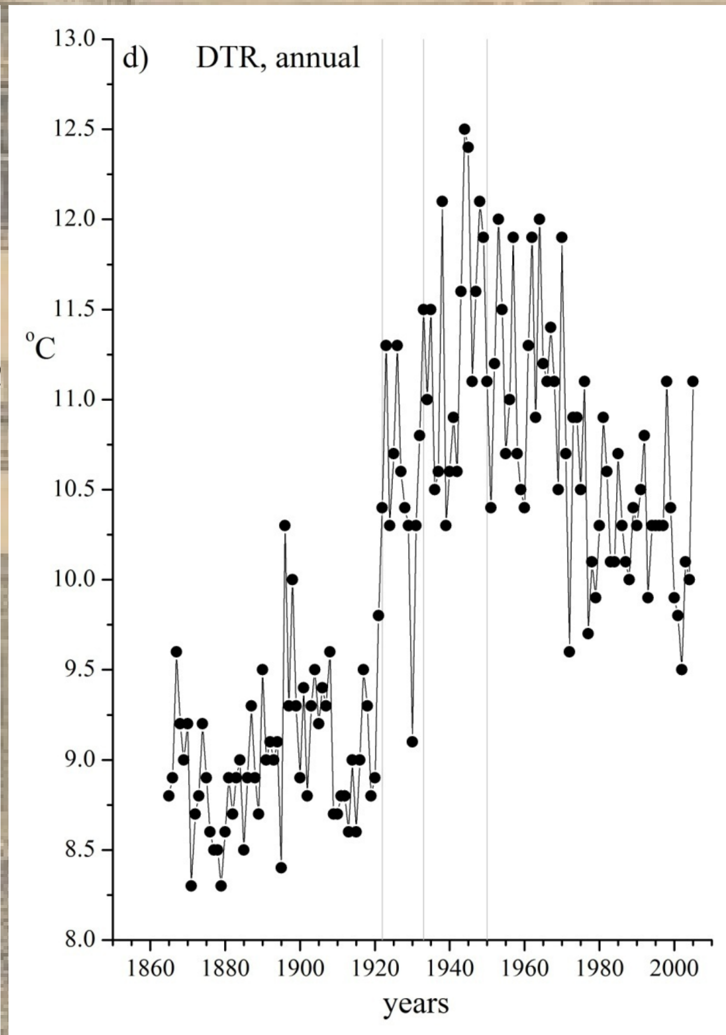
- 12 monthly series of
  - minimum temperature: *Tmin*
  - maximum temperature: *Tmax*
  - daily temperature range (*DTR*): *DTR* = *Tmax* – *Tmin*

- *Reference series*: temperature series from Lisbon (1856-2008) and Porto (1888-2001)

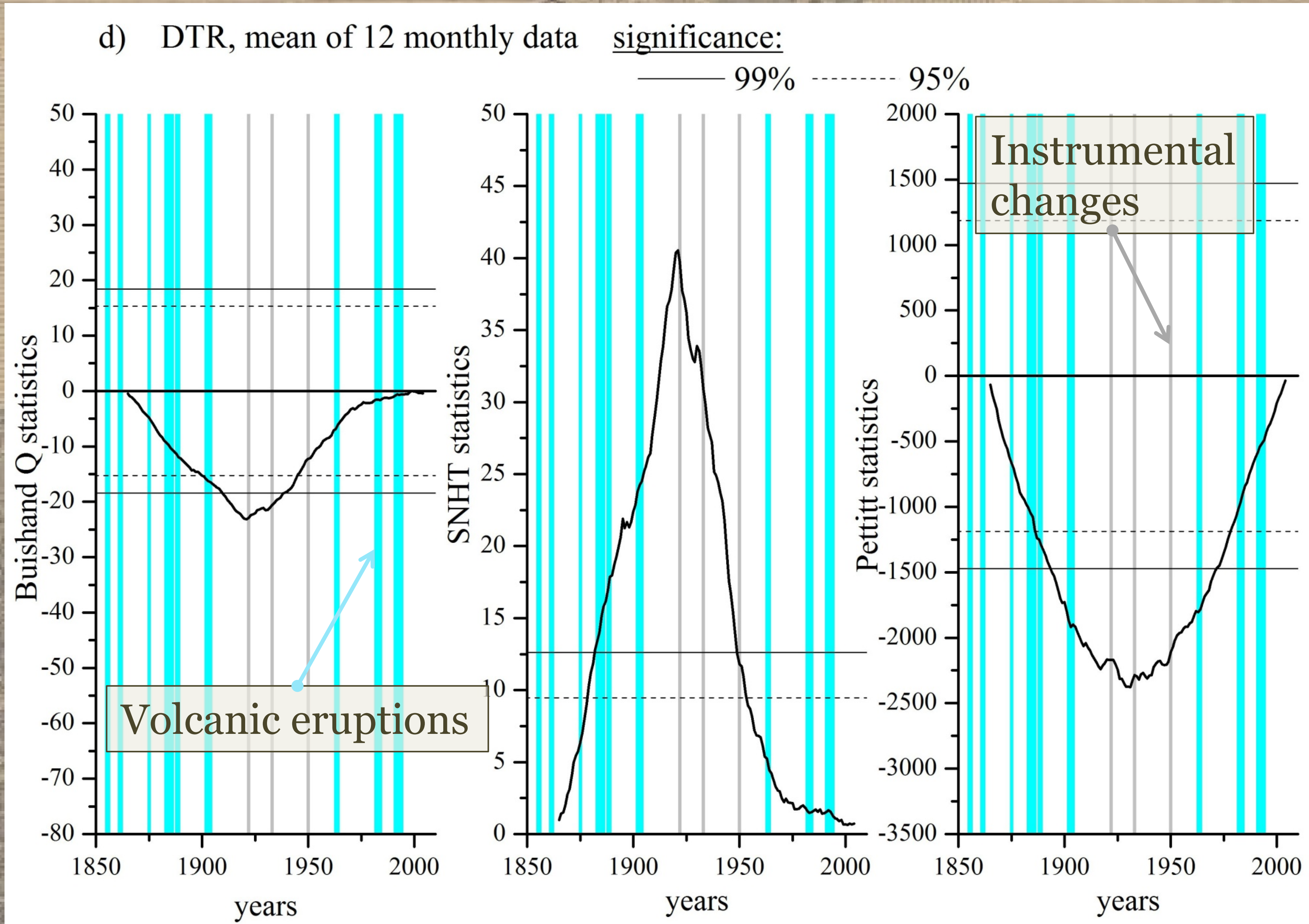
Temperature Homogenization: Visual Analysis







Temperature Homogenization: Homogeneity tests



Temperature Homogenization: breaks for correction

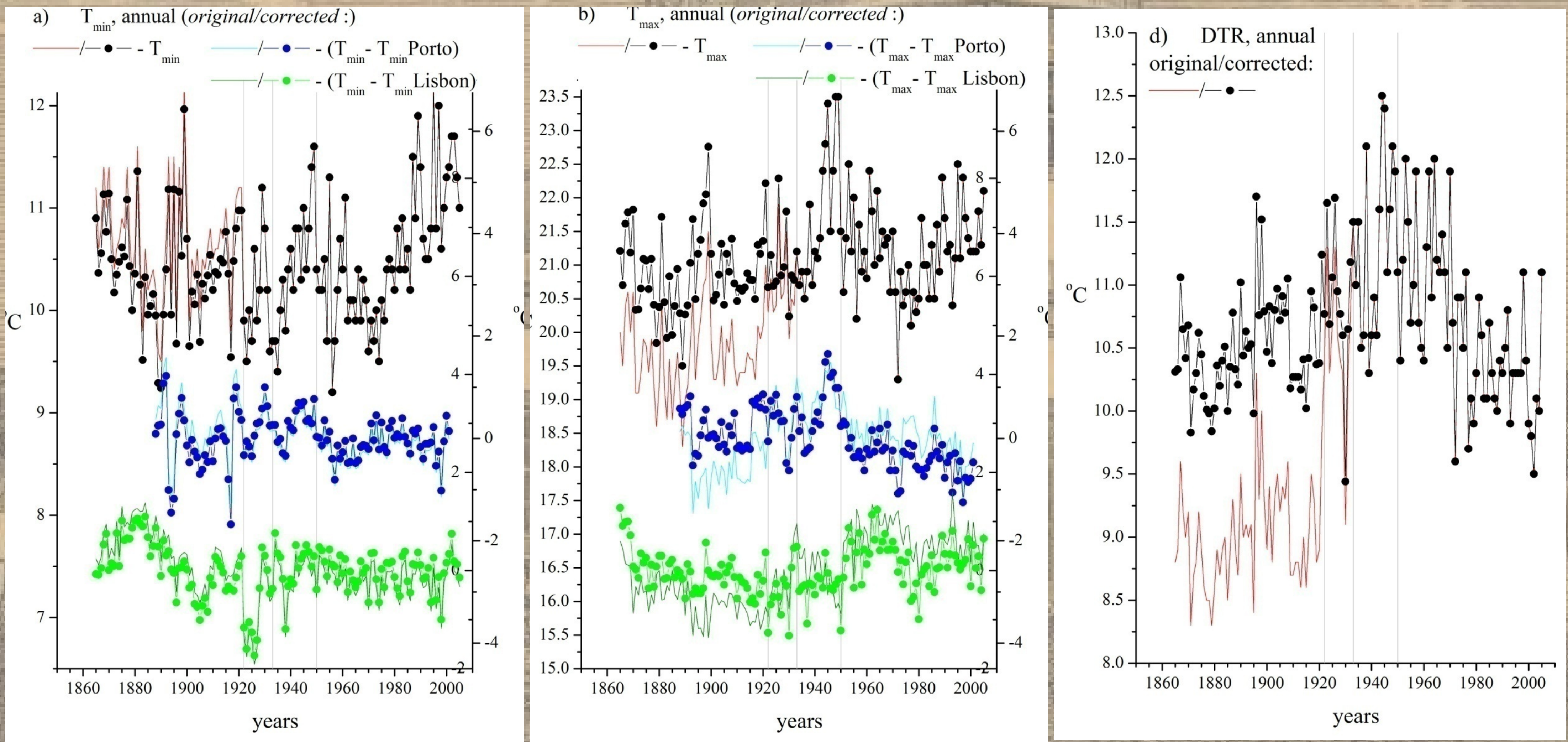
- *Tmin*: 1922

  (correction values are calculated using periods of 40 years before/after the break)

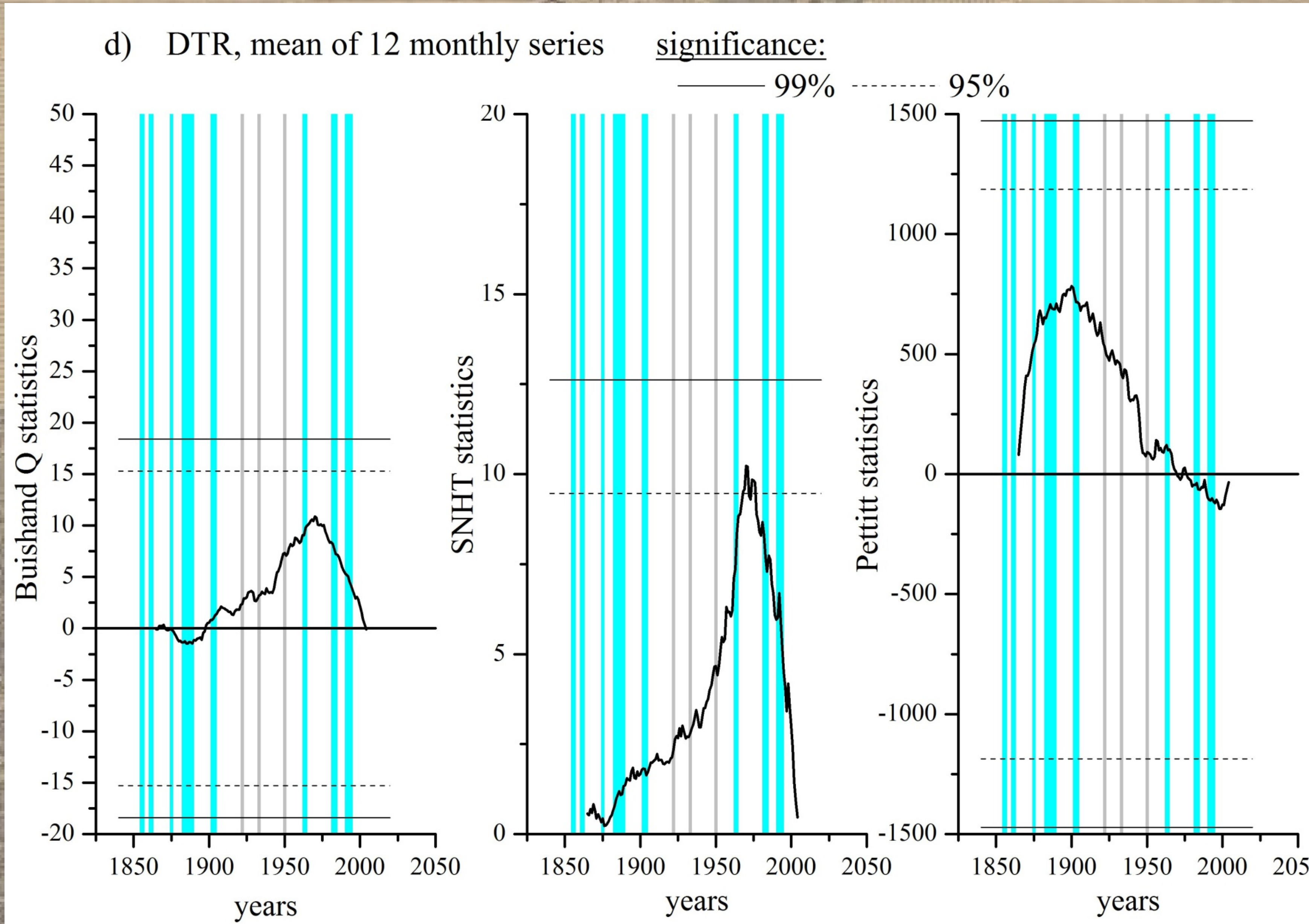
- *Tmax: 1922 & 1933*

  (correction values are calculated using periods of 10 years before/after the breaks)

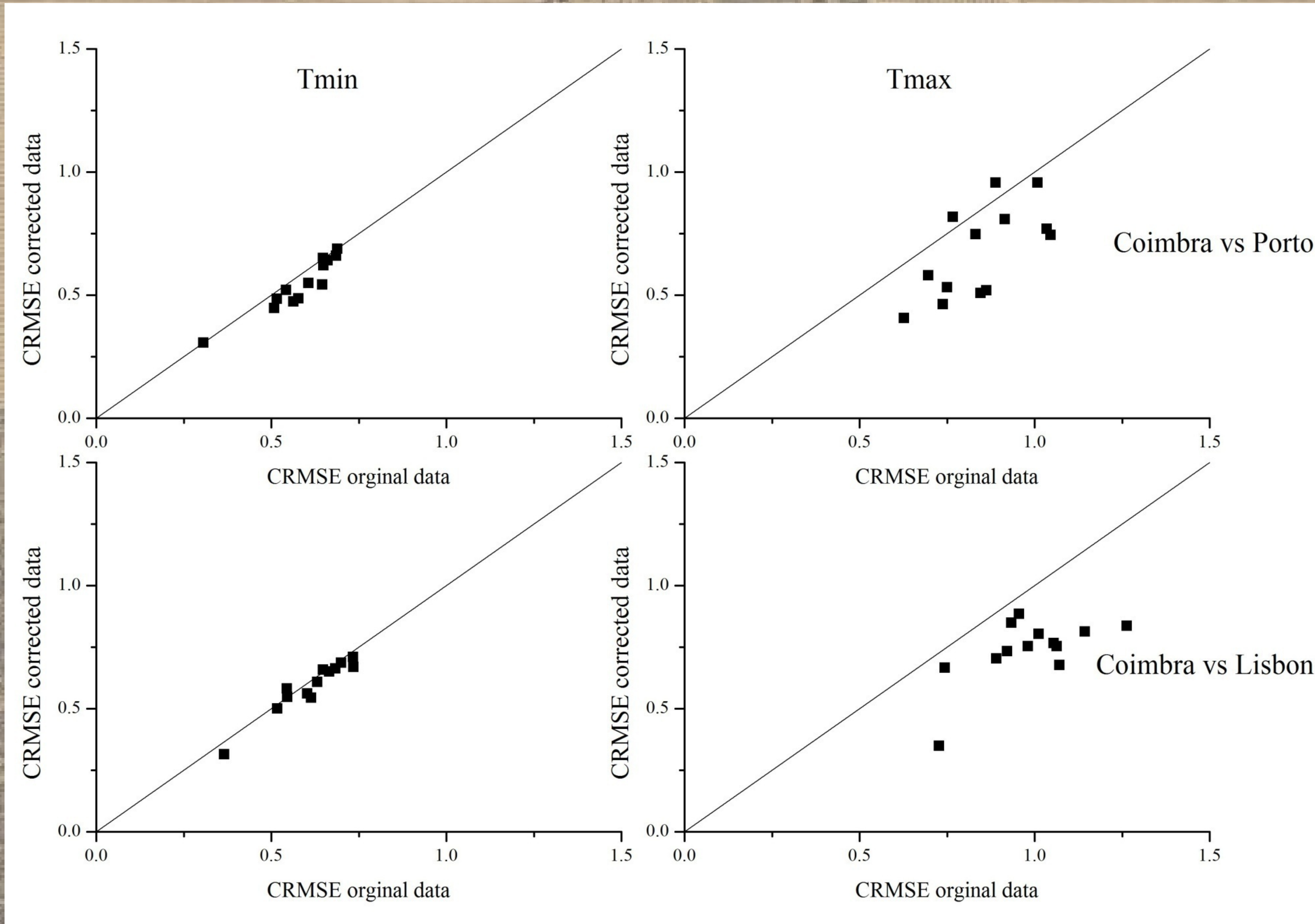
Temperature Homogenization: Correction



Temperature Homogenization: Homogeneity tests of corrected data

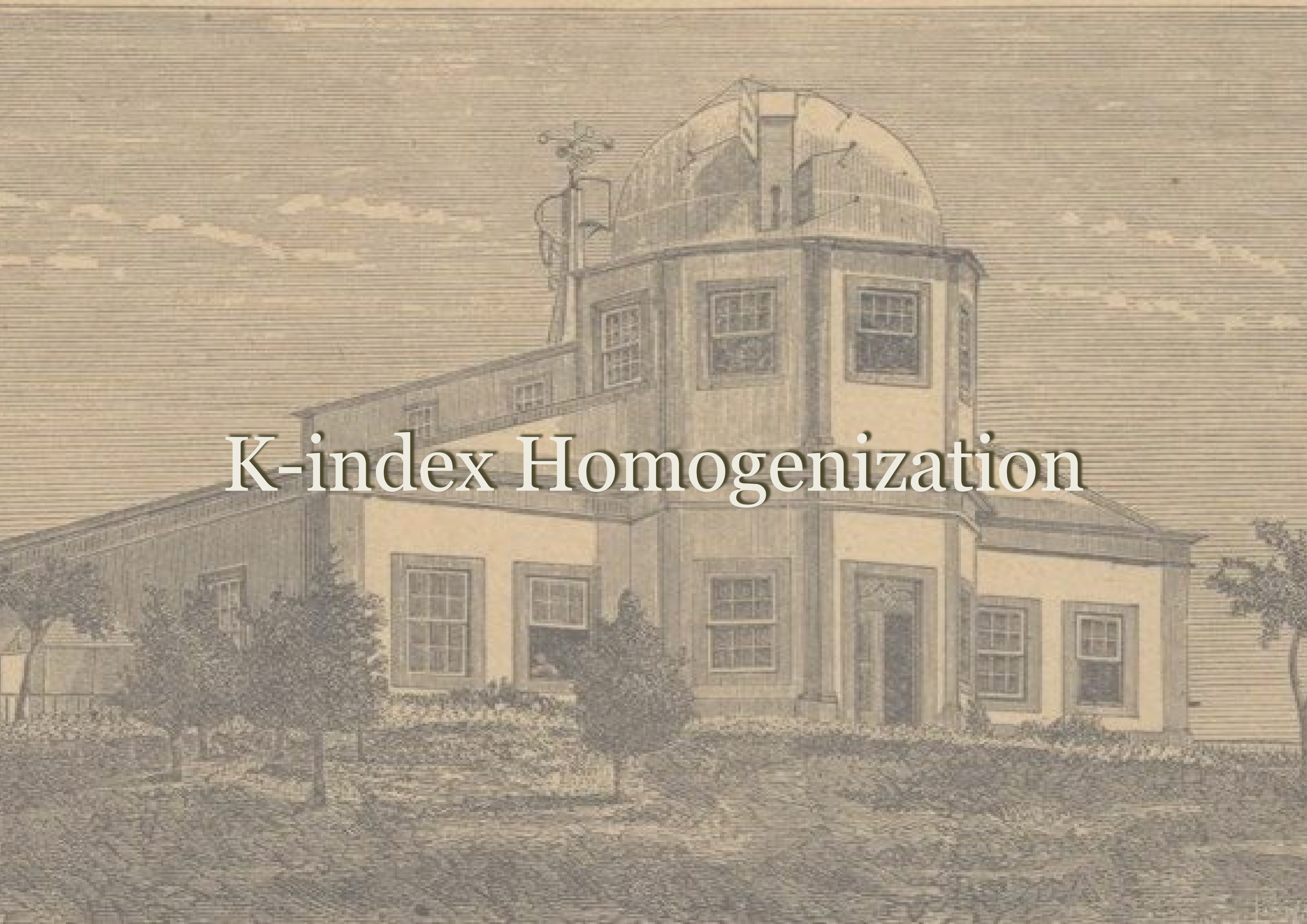


Temperature Homogenization: CRMSE changes



Temperature Homogenization: Results

- Corrected IGUC temperature series are considered to be free of artificial shifts (significance at least of 95%)
- There are still inhomogeneities caused by natural climate variability, e.g. volcanic influence



K-index Homogenization

K-indices & Kp

- K-indices show solar particle effects on the earth's magnetic field.
- They are calculated on the base of the geomagnetic field disturbance level during every 3-hour period.
- K-indices range from 0 (quiet) to 9 (greatly disturbed).
- Each activity level relates almost logarithmically to its corresponding disturbance amplitude.
- The arithmetic mean of the K-indices scaled at the 13 worldwide distributed observatories gives Kp “planetary K-index”.

AA & Ap indices

- AA is a global index of magnetic activity.
- AA is produced from the K-indices of two nearly antipodal magnetic observatories in England and Australia.

- Ap is the arithmetic mean of the day's eight ap values.
- *where ap represents K-indices converted to a linear scale in nT*

Metadata for geomagnetic station

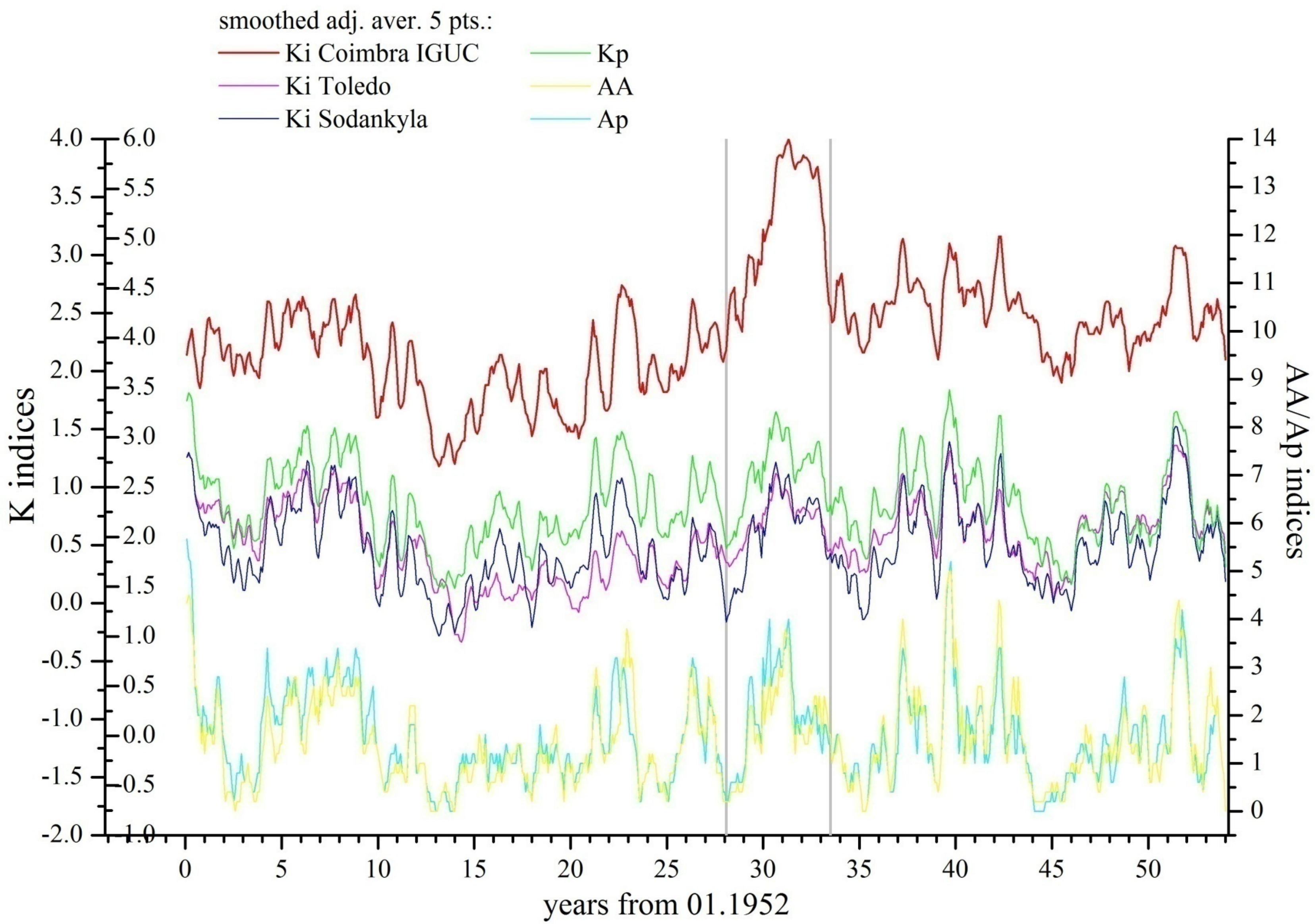
- 1980, January – replacement of the original suspension wire (made of quartz) of the H variometer by a Tungsten wire
- 1985, June – new method to calculate the K index scales

K-index Homogenization: Parameters

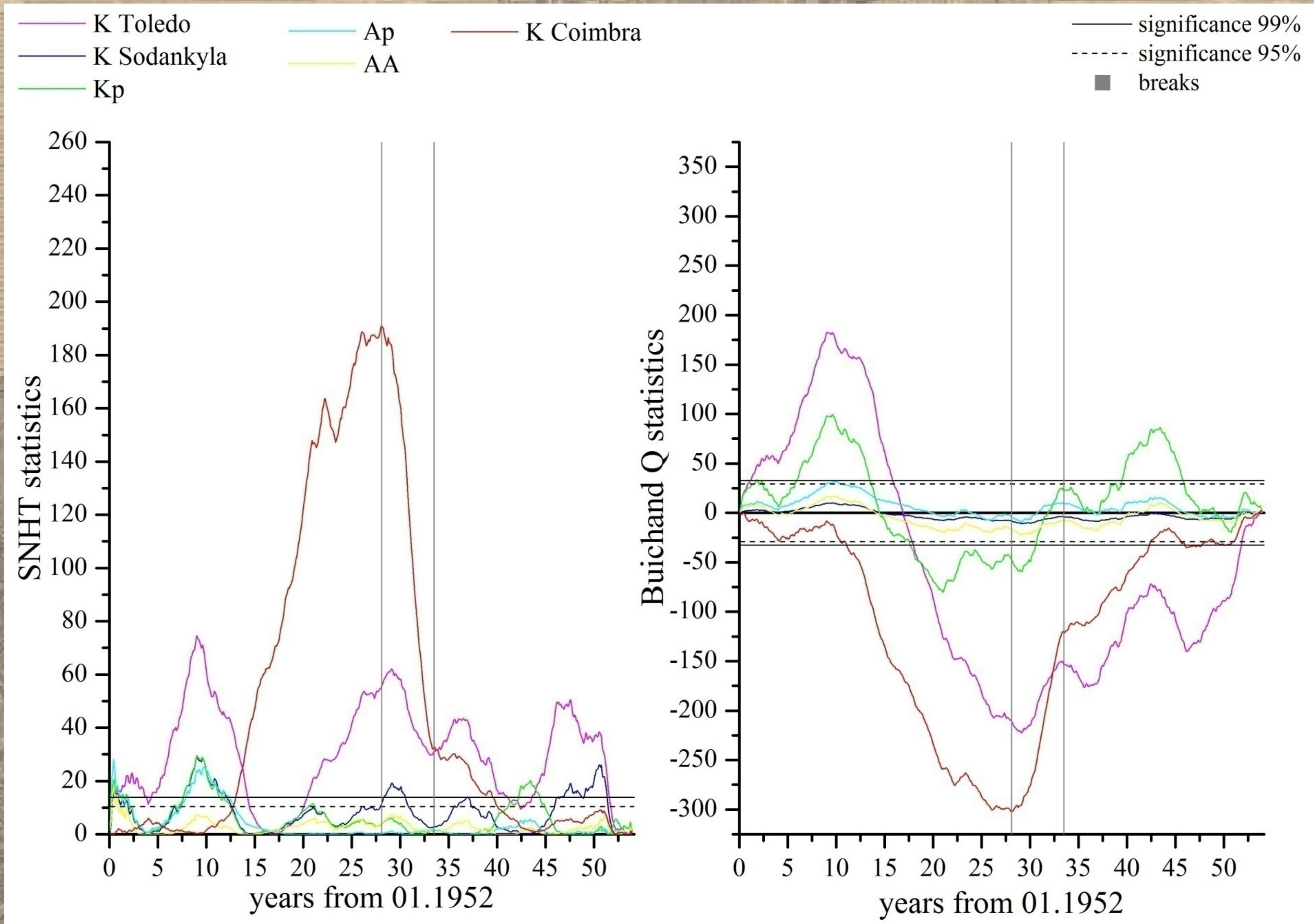
- 1 series of monthly K-index data
  (*not 12 monthly series!*)

- 4 reference series:
  - Monthly series of K-indices of two geomagnetic stations
    Toledo, Spain (φ=39º33’N, λ=4º21’W)
    Sodankylä, Finland (φ=67º22’N, λ=26º38’E)
  - Kp index
  - AA & Ap indices

K-index Homogenization: Visual analysis



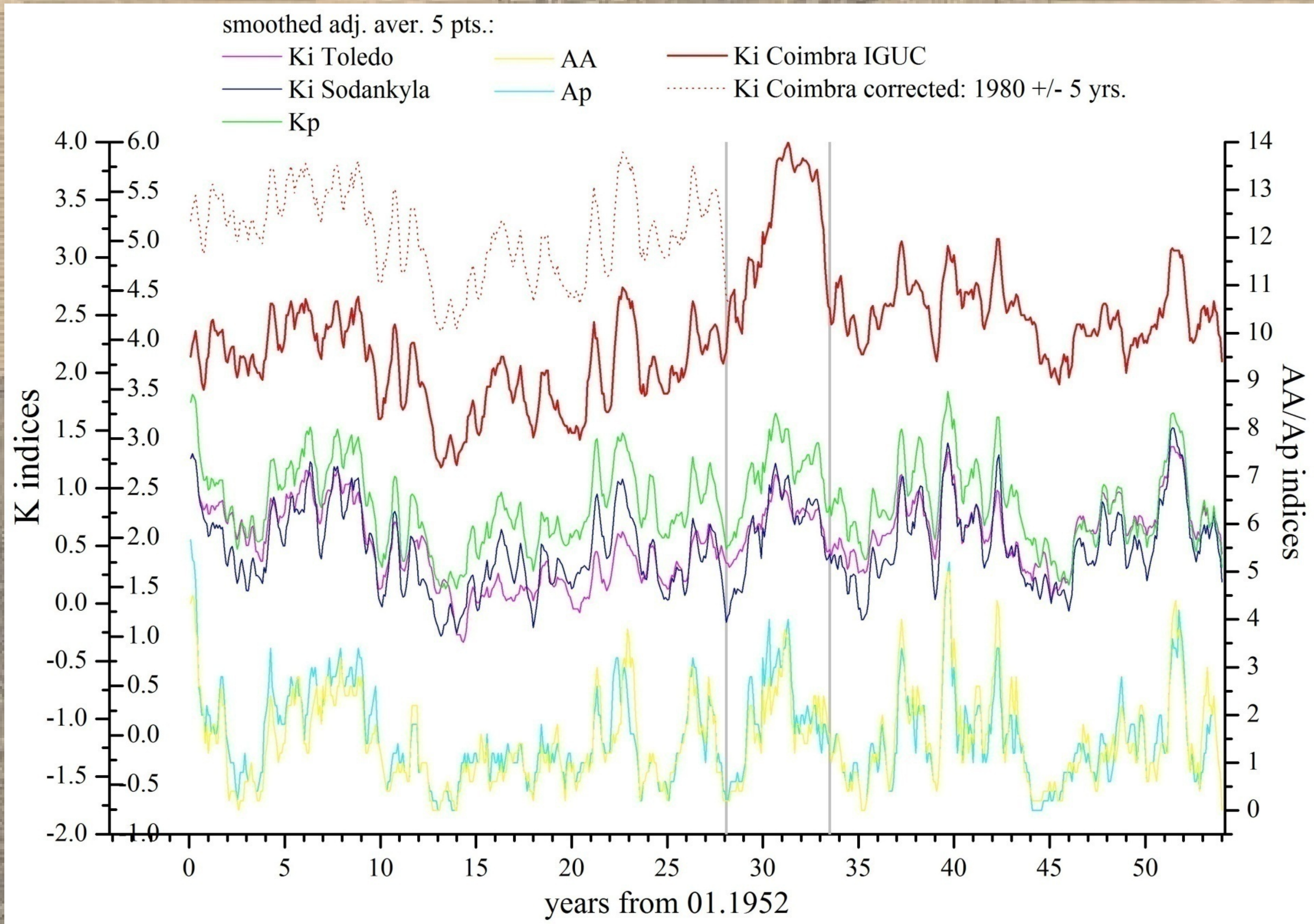
K-index Homogenization: Homogeneity tests



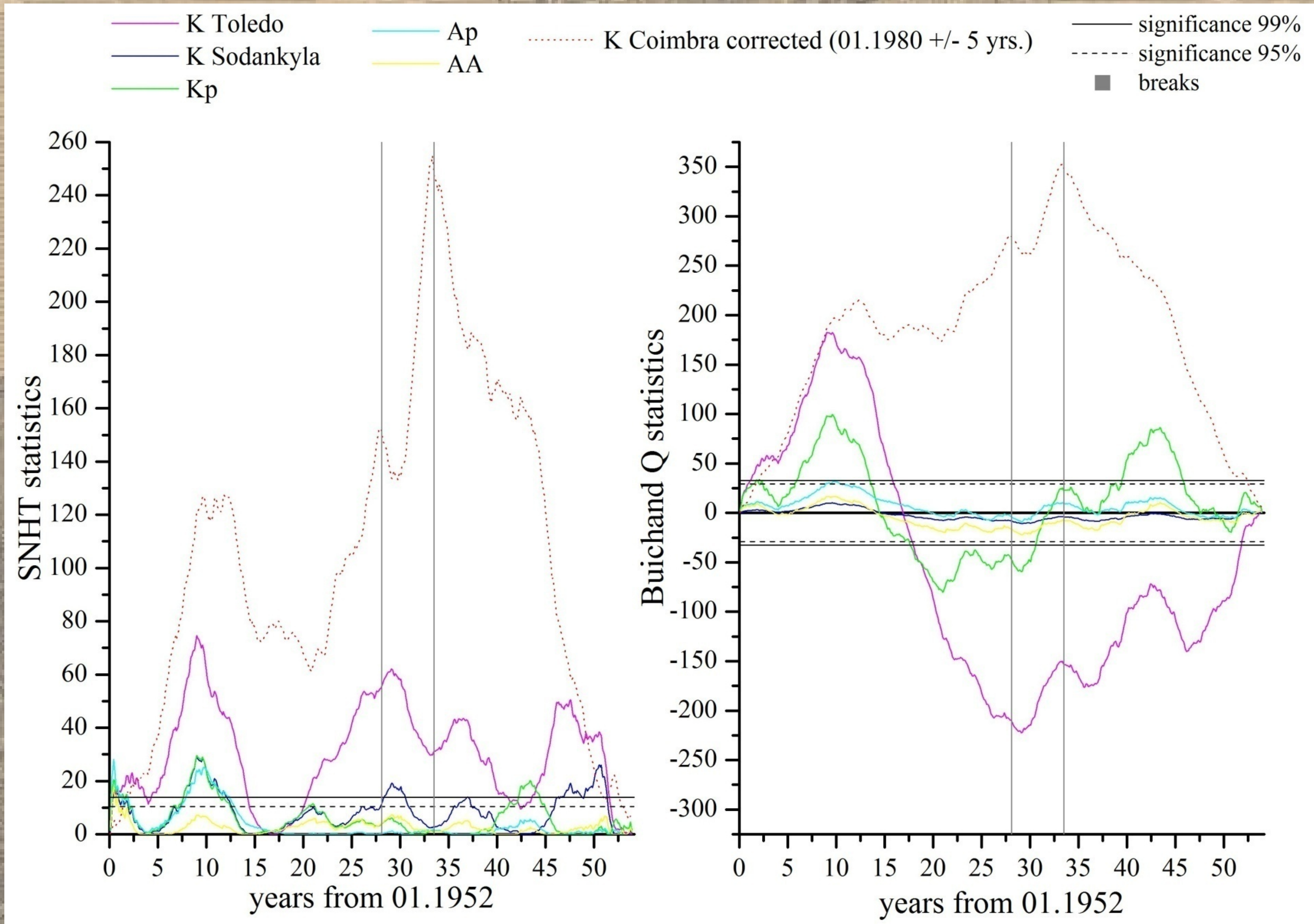
K-index Homogenization: 1st break for correction

- 01.1980
  (correction values are calculated using periods of 5 years before/after the breaks)

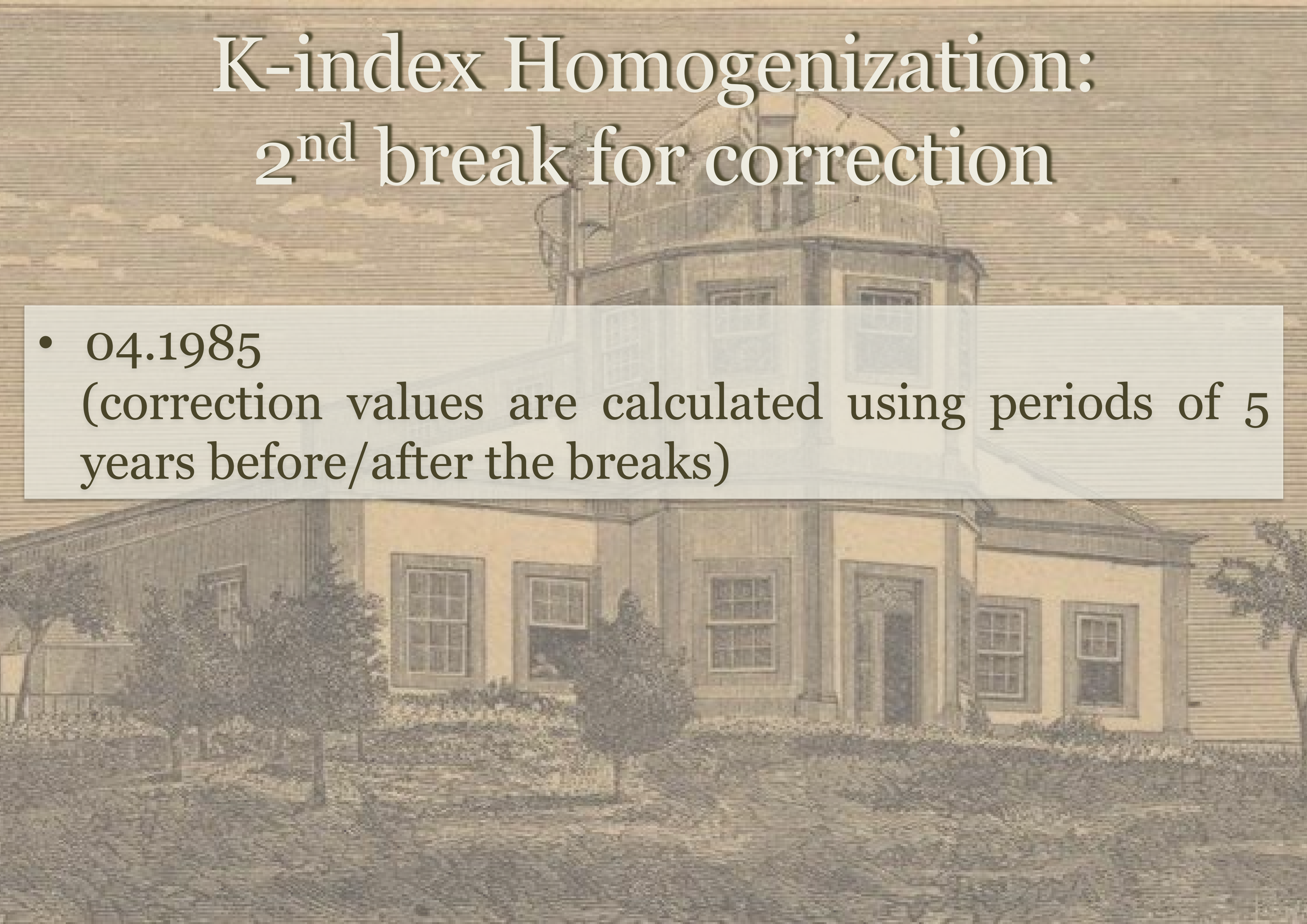
K-index Homogenization: 1st correction



K-index Homogenization: Homogeneity tests 2

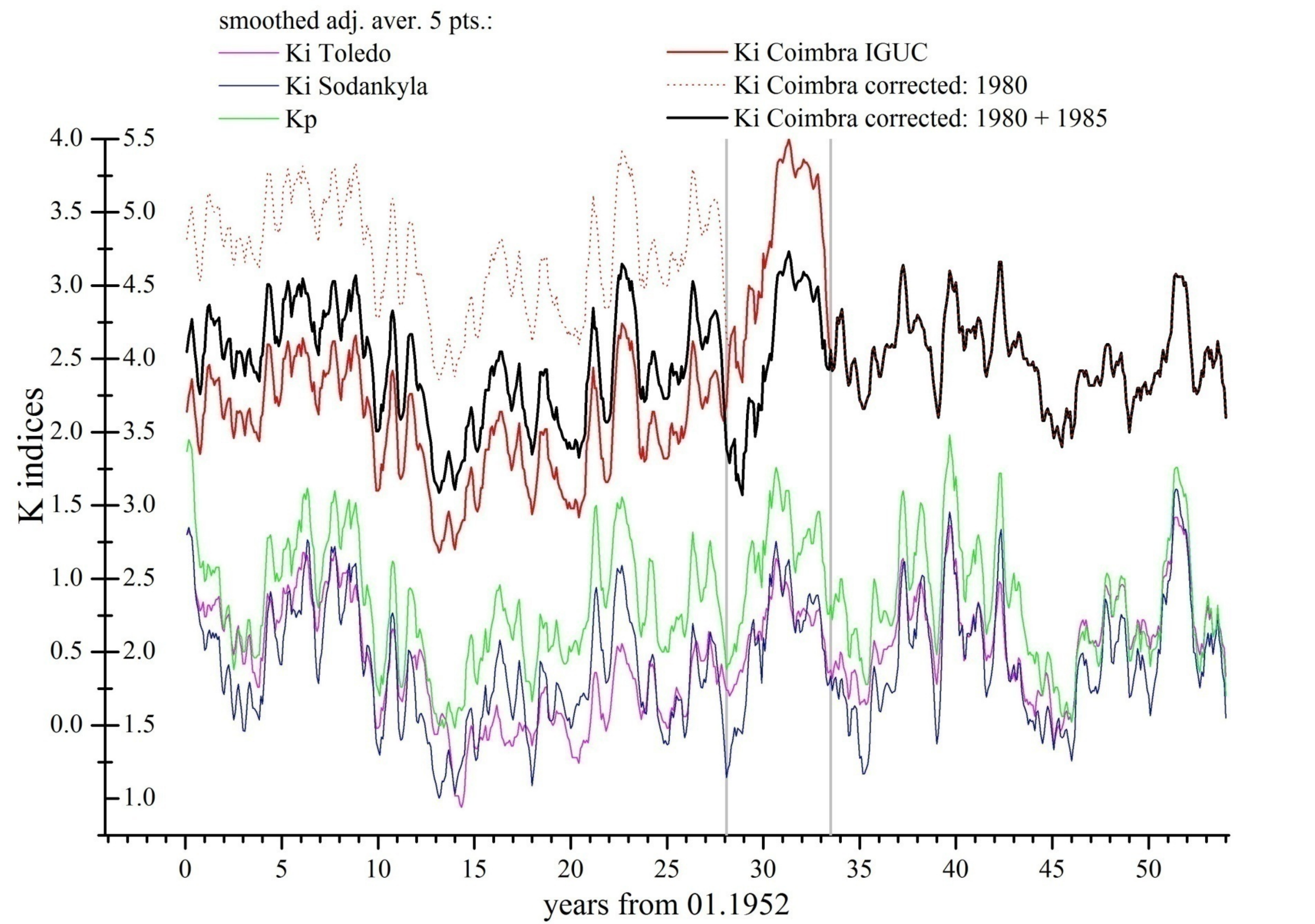


K-index Homogenization: 2nd break for correction

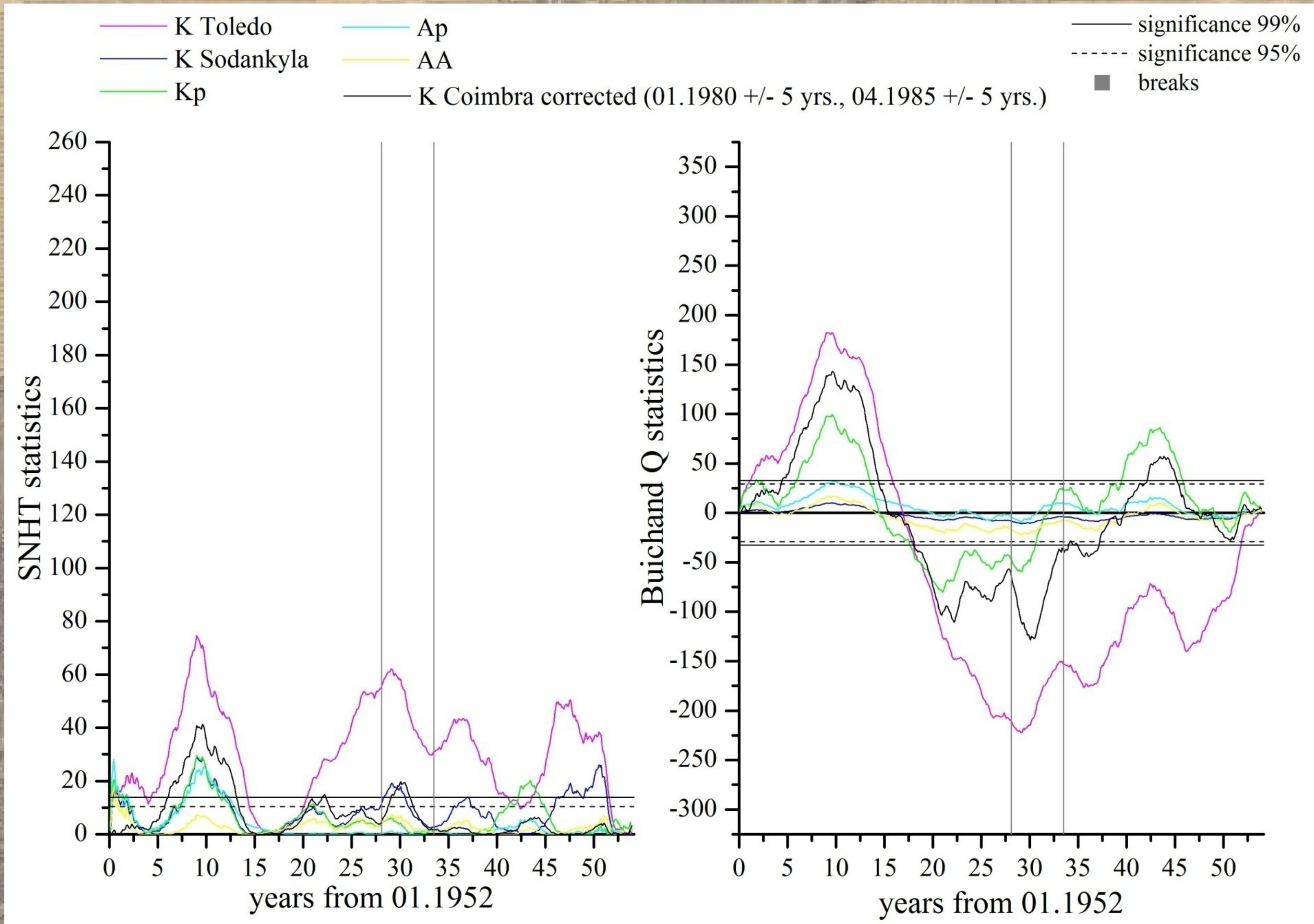


- 04.1985
  (correction values are calculated using periods of 5 years before/after the breaks)

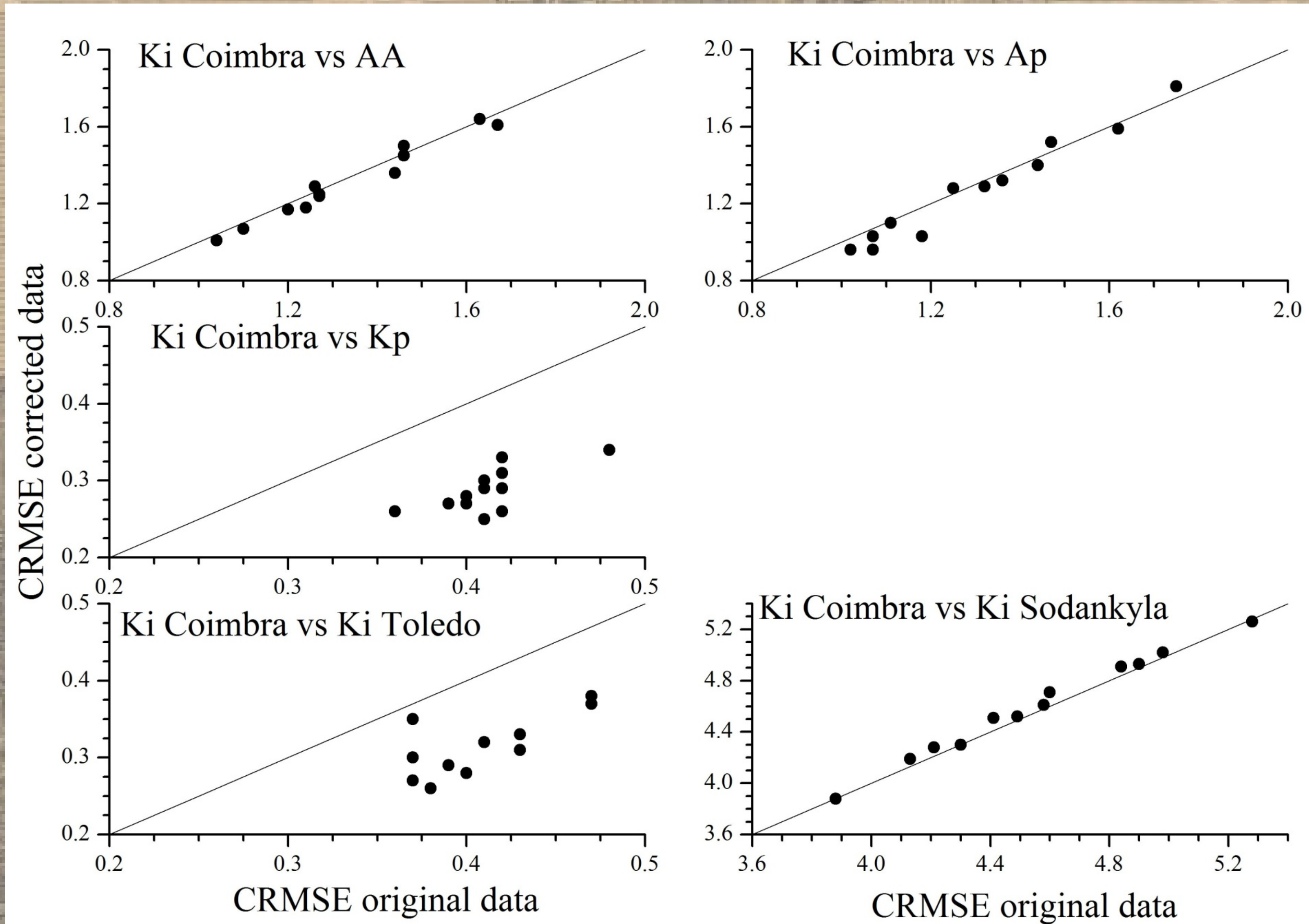
K-index Homogenization: Final correction



K-index Homogenization: Homogeneity tests 2

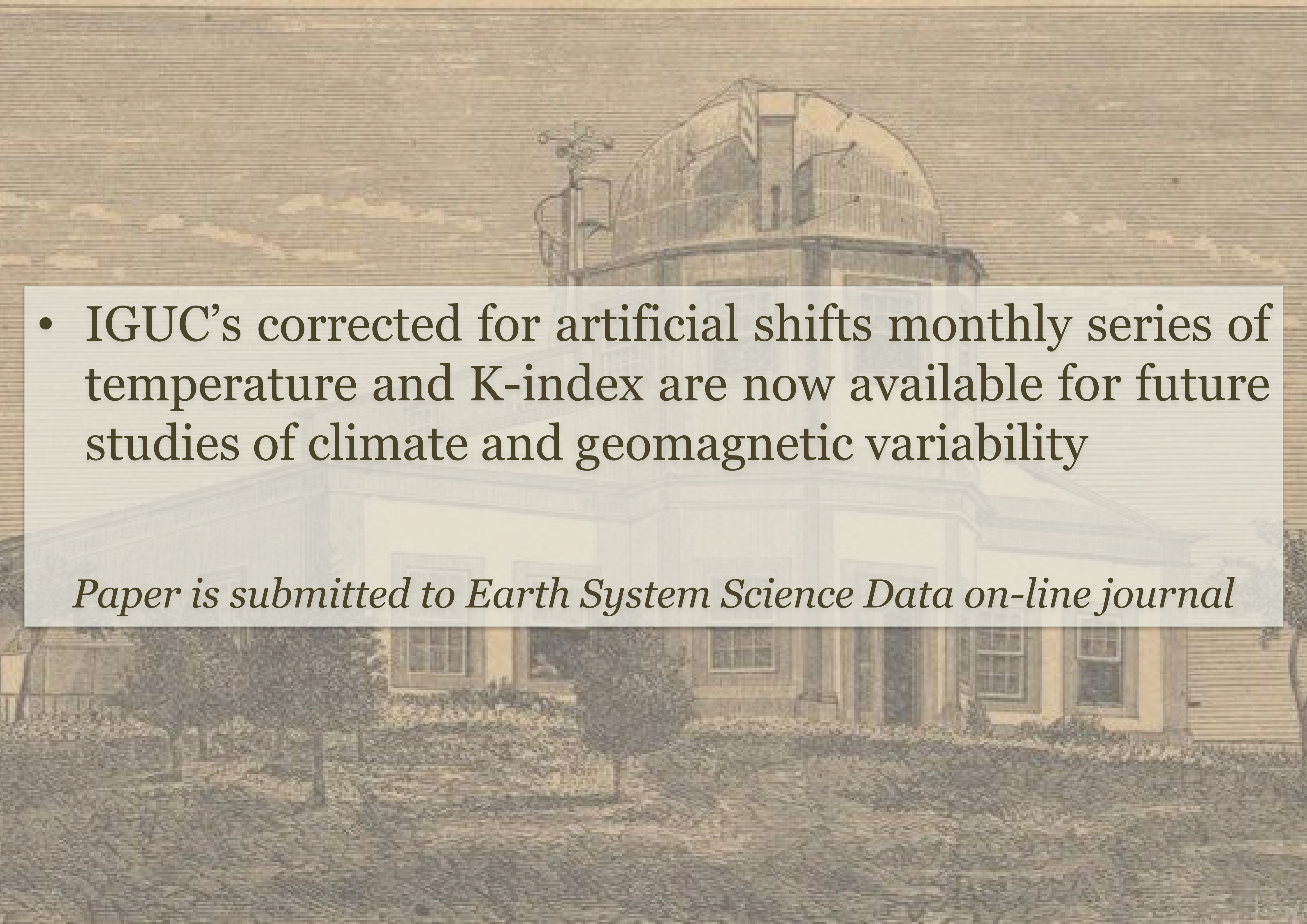


K-index Homogenization: CRMSE changes



K-index Homogenization: Results

- Corrected IGUC K-index series are considered to be free of artificial shifts (significance at least of 95%)
- There are still inhomogeneities caused by natural variability: cycles of solar activity



- IGUC’s corrected for artificial shifts monthly series of temperature and K-index are now available for future studies of climate and geomagnetic variability

Paper is submitted to Earth System Science Data on-line journal

Thank you

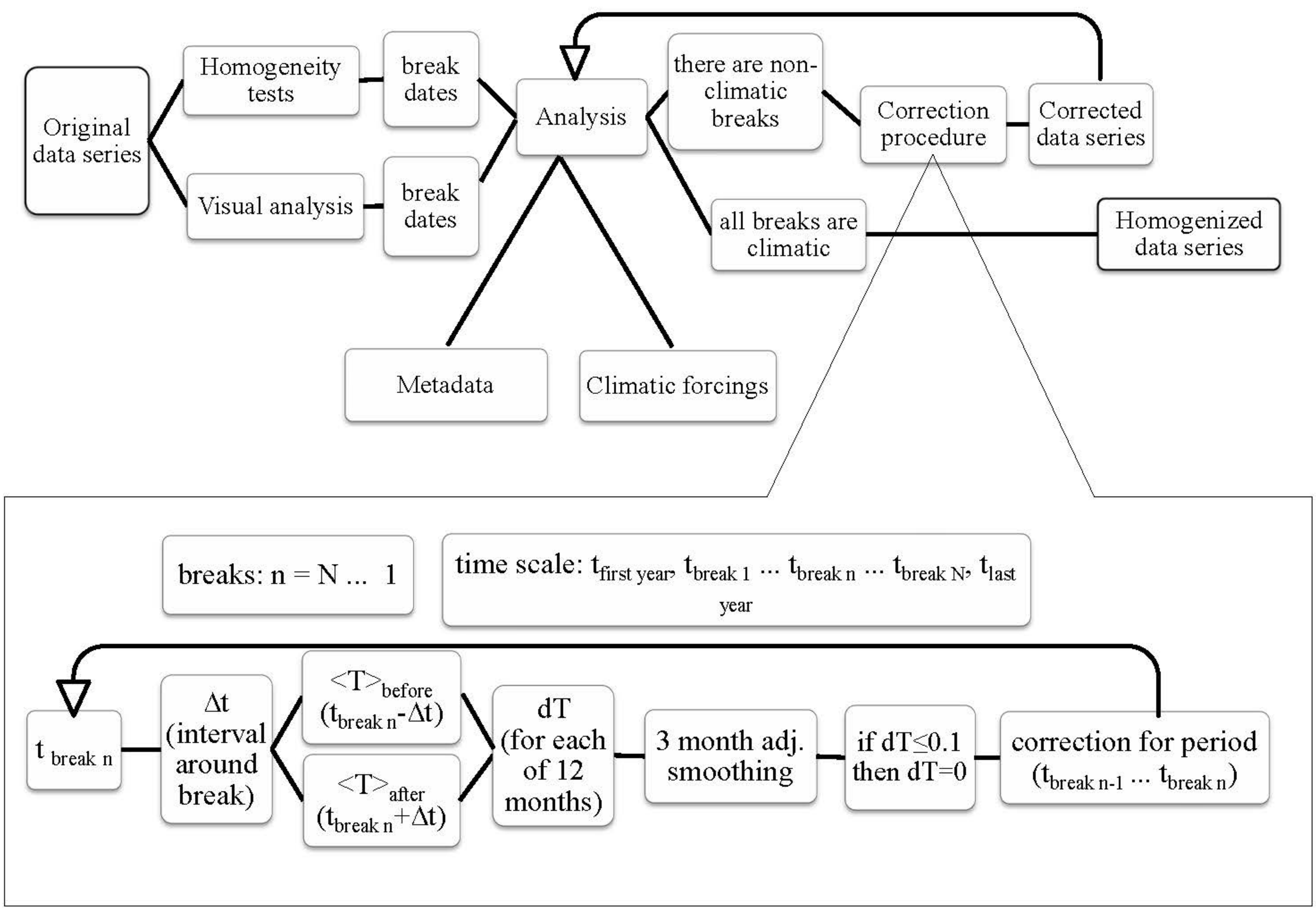
7ª Asamblea Hispano Portuguesa de Geodesia y Geofísica

7ª Assembléia Hispano Portuguesa de Geodesia e Geofísica

Donostia - San Sebastián, del 25 al 28 de junio de 2012

CGUC CENTRO DE GEOFÍSICA UNIVERSIDADE DE COIMBRA

Homogenization Procedure



CRMSE

- $CRMSE^2 = \sigma_D^2 + \sigma_R^2 - 2\,\sigma_D^2\,\sigma_R^2\,r$

- σ_D – standard deviation of analysed data
- σ_R – standard deviation of reference series
- r – correlation coefficient between analysed data and reference series

Buishand test

- parametric test
- 0-hypothesis: data are independent identically normally distributed values
- Adjusted partial sums (Buishand, 1982):

$$S_k^* = n\sum_{i=1}^{k}(Y_i - \bar{Y})/\sum_{i=1}^{n}(Y_i - \bar{Y})^2,\ k = 1\ldots n,\ S_0^* = 0$$

if a break is present in year K, then S_k^* reaches a maximum (negative shift) or minimum (positive shift) near the year $k = K$

- sensitive to the breaks near the middle of the series
- Buishand (1982) gives critical values for statistics for different data set lengths

Standard normal homogeneity test (SNHT)

- likelihood ratio test
- 0-hypothesis: data are independent identically normally distributed values
- statistics T (Alexandersson and Moberg , 1997)

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2, k = 1 \ldots n$$

$$\bar{z}_1 = \frac{1}{k}\frac{\sum_{i=1}^{k}(Y_i - \bar{Y})}{s} \quad \bar{z}_2 = \frac{1}{n-k}\frac{\sum_{i=k+1}^{n}(Y_i - \bar{Y})}{s}$$

$$s = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

if a break is present in year K, then $T(k)$ reaches a max near the year $k = K$.

Standard normal homogeneity test (SNHT)

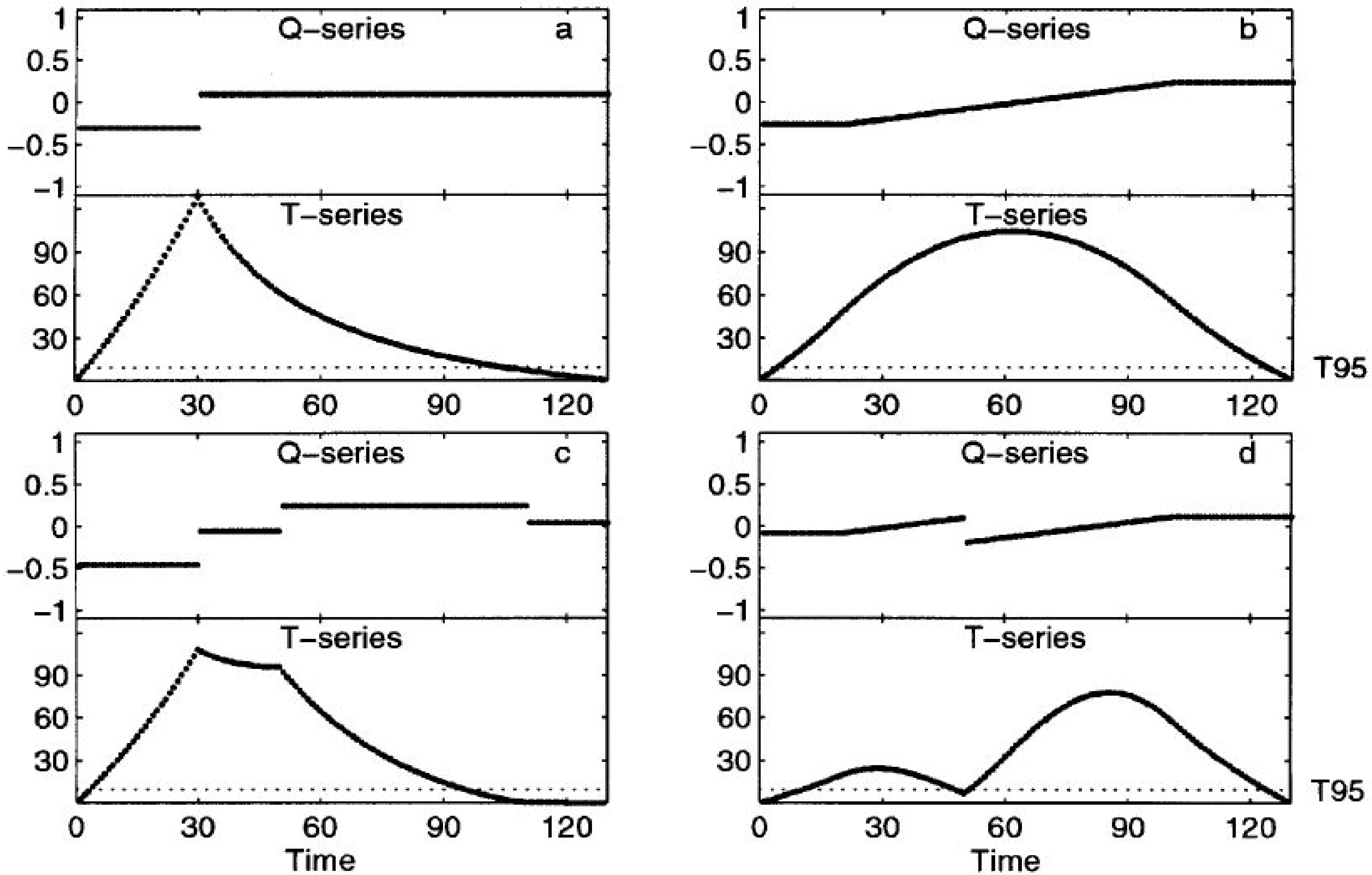


Figure A1. Idealized examples of Q-series and the corresponding T-series for the single shift test. The 95 per cent critical level, T_{95}, is indicated with small dots. (a) A single shift. (b) A perfect trend. (c) Three distinct shifts. (d) A perfect trend interrupted by a single shift

Standard normal homogeneity test (SNHT)

- sensitive to the breaks near the beginning and the end of the series
  - *could generate "false alarm"*
- critical values for different data set lengths are given in Khaliq and Ouarda (2007)

Pettitt test

- non-parametric rank test
- 0-hypothesis: data are independent identically normally distributed values
- The ranks $r_1 \ldots r_n$ of the $Y_1 \ldots Y_n$ are used to calculate the statistics (Pettitt, 1979)

$$X_k = 2\sum_{i=1}^{k} r_i - k(n+1),\ k = 1 \ldots n$$

  if a break is present in year K, then statistic is maximal or minimal near the year $k = K$.

- sensitive to the breaks near the middle of the series
- statistical significance for probability level α

$$X_{K\alpha} = [-\ln\alpha\,(n^3 + n^2)/6]^{1/2}$$

References:

- von Neumann, J.: Distribution of the Ratio of the Mean Square Successive Difference to the Variance, Ann. Math. Stat., 12, 4, 367-395, 1941.
- Buishand, T.A.: The analysis of homogeneity of long-term rainfall records in The Netherlands, R. Neth. Meteorol. Inst. (K.N.M.I.), De Bilt, Sci. Rep. No. 81-7, 77 pp., 1981.
- Buishand, T.: Some methods for testing the homogeneity of rainfall records, J. Hydrol., 58, 1-2, 11-27, 1982.
- Alexandersson, H., Moberg, A: Homogenization of swedish temperature data. Part I: Homogeneity test for linear trends, Int. J. Climatol., 17, 1, 25-34, 1997.
- Khaliq, M.N., Ouarda, T.B.M.J.: On the critical values of the standard normal homogeneity test (SNHT): Int. J. Climatol., 27, 5, 681-687, 2007.