On homogenization of Portuguese meteorological and geophysical data series

A.L. Morozova¹, P. Pribeiro¹, M.A.Valente², M.A. Pais¹

¹ Centro de Geofísica da Universidade de Coimbra
² Instituto Dom Luiz, University of Lisbon
Data description

- Geophysical Institute of the University of Coimbra (IGUC): $\varphi = 40^{\circ}12^\prime$ N, $\lambda = 8^{\circ}25^\prime$ W, $h = 141$ m
- digitized monthly series of temperature parameters ($T_{\text{min}}$, $T_{\text{max}}$ and mean daily $T$): from 1865 to 2005
- digitized daily series of temperature parameters ($T_{\text{min}}$, $T_{\text{max}}$ 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 a study of homogeneity level of the meteorological and geomagnetic data series and a correction 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: $T_{min}$
  – maximum temperature: $T_{max}$
  – daily temperature range ($DTR$): $DTR = T_{max} - T_{min}$

• Reference series: temperature series from Lisbon (1856-2008) and Porto (1888-2001)
Temperature Homogenization: Visual Analysis

a) Tmin, annual
   - (Tmin - Tmin Porto)
   - Tmin
   - (Tmin - Tmin Lisbon)

b) Tmax, annual
   - (Tmax - Tmax Porto)
   - Tmax
   - (Tmax - Tmax Lisbon)

d) DTR, annual

years

°C

°C

°C

°C

°C

°C

°C
Temperature Homogenization: Homogeneity tests

d) DTR, mean of 12 monthly data

**Significance:**
- 99%
- 95%

**Instrumental changes**

**Volcanic eruptions**
• **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 break)
Temperature Homogenization: Correction

Graphs showing temperature data over years, with different markers and lines indicating original and corrected values for minimum and maximum temperatures.
Temperature Homogenization: Homogeneity tests of corrected data

d) DTR, mean of 12 monthly series

significance:

--- 99%

--- 95%
Temperature Homogenization: CRMSE changes

- Tmin
- Tmax

Coimbra vs Porto

Coimbra vs Lisbon
• 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 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=39°33’N, λ=4°21’W)
    Sodankylä, Finland (φ=67°22’N, λ=26°38’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

smoothed adj. aver. 5 pts.:
- Ki Toledo
- Ki Sodankyla
- Ap
- Kp
- Ki Coimbra IGUC
- Ki Coimbra corrected: 1980 +/- 5 yrs.

years from 01.1952
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

smoothed adj. aver. 5 pts.:
- Ki Toledo
- Ki Sodankyla
- Kp
- Ki Coimbra IGUC
- Ki Coimbra corrected: 1980
- Ki Coimbra corrected: 1980 + 1985

K indices

years from 01.1952
K-index Homogenization: Homogeneity tests 2
K-index Homogenization: CRMSE changes

- Ki Coimbra vs AA
- Ki Coimbra vs Ap
- Ki Coimbra vs Kp
- Ki Coimbra vs Ki Toledo
- Ki Coimbra vs Ki Sodankyla
• 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*
Homogenization Procedure

1. Original data series
   - Homogeneity tests
     - Visual analysis
   - Break dates
2. Analysis
   - There are non-climatic breaks
   - All breaks are climatic
     - Metadata
     - Climatic forcings
3. Correction procedure
   - Corrected data series
4. Homogenized data series

Breaks: \( n = N \ldots 1 \)

Time scale: \( t_{\text{first year}} \ldots t_{\text{break } 1} \ldots t_{\text{break } n} \ldots t_{\text{break } N} \ldots t_{\text{last year}} \)

- \( t_{\text{break } n} \) (interval around break)
- \( \Delta t \) (for each of 12 months)
- 3 month adj. smoothing
- if \( dT \leq 0.1 \) then \( dT = 0 \)
- Correction for period \( t_{\text{break } n-1} \ldots t_{\text{break } n} \)
• CRMSE$^2 = \sigma_D^2 + \sigma_R^2 - 2 \sigma_D^2 \sigma_R^2 r$

• $\sigma_D$ – standard deviation of analysed data
• $\sigma_R$ – 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):

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

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)
  \[ T(k) = k\bar{Z}_1^2 + (n-k)\bar{Z}_2^2, \quad k = 1 \ldots n \]
  \[ \bar{Z}_1 = \frac{1}{k} \frac{\sum_{i=1}^{k} (Y_i - \bar{Y})}{s} \quad \bar{Z}_2 = \frac{1}{n-k} \frac{\sum_{i=k+1}^{n} (Y_i - \bar{Y})}{s} \]
  \[ s = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \]

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

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 $r_1...r_n$ of the $Y_1...Y_n$ are used to calculate the statistics (Pettitt, 1979)
  \[ X_k = 2 \sum_{i=1}^{k} r_i - k(n + 1), \quad k = 1...n \]
  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 $\alpha$
  \[ X_{K\alpha} = \left[-\ln \alpha (n^3 + n^2)/6\right]^{1/2} \]
References:


