Homogeneity Analysis of Portuguese SST Time Series

Anabela Jorge de Carvalho[1], Alice Soares dos Santos[2]

Instituto de Meteorologia, Rua C Aeroporto, 1749-077 LISBOA, PORTUGAL

[1] anabela.jc@meteo.pt, [2] alice.soares@meteo.pt

Many of the time series of climatic parameters exhibit spurious non-climatic jumps and/or gradual shifts due to changes in station location, environment, instrumentation or observing practices. The presence of such heterogeneities (artificial irregularities) mask natural trends and variability. These signals overlapping (natural and artificial) difficult time series analysis and requires a pre-processing methodology to homogenise the series and to obtain reliable data, where the spurious effects were eliminated.

In this context, three statistical tests were chosen for testing homogeneity (absolute) of all annual SST (Sea Surface Temperature) time series of the Portuguese Coast. Here are presented some of these results.

Methodology

The adopted methodology denotes Y as a selected serie (in this case, the annual time serie of SST), and Y_i the current value in year i ($i = 1,...,n$), which is assumed, under null hypothesis H_0, have the same mean. For the used tests Y_i are independents, with a *Gaussian distribution.*

Statistical Analysis of Homogeneity

Standard Normal Homogeneity Test (absolutely) – SNHT (Alexandersson, 1986)

The exploitation of Alexanderson-test gives great advantages because it permits to identify the temporal location of the heterogeneity and also its intensity (size). This is appropriate to study artificial trends in climatic temporal series.

This statistical test (T_0) used for a single shift permits no more than find the most significant inhomogeneity in a time serie. Using it in an interactive procedure it is required.

$$T_0 = \max_{1\le a<n} T(a) = \max_{1\le a<n}\left\{a z_1^2 + (n-a) z_2^2\right\} \qquad a = 1,...,n$$

$$z_1 = \frac{1}{a}\sum_{i=1}^{a} (Y_i - \bar{Y})/s \qquad z_2 = \frac{1}{n-a}\sum_{i=a+1}^{n} (Y_i - \bar{Y})/s$$

where $\bar{Y}$ denotes the mean and *s the* standard deviation of the sample. ***T(a)*** is used to relate the mean of the first *a* years of the record with that of the last *n-a years.*

It is assumed that, under the null hypothesis H_0, Y_i have the same mean, however, under the alternative hypothesis H_A, SNHT test denotes shifts, i.e., when ***T(a)*** has a maximum in year $a = A$, it is means that is located in year *A* a probable change (deviation). At what time ***T(a)*** has small values to all *a* years the null hypothesis will be valid, besides when ***T(a)*** takes large values it will be more probable the alternative hypothesis.

The null hypothesis will be rejected when T_0 be above a specified level, which is dependent on the sample size (in this case was used 1% critical values for a single shift SNHT as a function of *n* (Jarusková, 1994)).

Range-test (Buishand, 1982)

This homogeneity test is based on the adjusted partial sums or cumulative deviations from the mean:

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

where $\bar{Y}$ is the average value of the Y_i's and $S_n^* = 0$.

If the selected series are homogeneous the expected values of S_k^* will fluctuate around zero. However, when S_k^* reach a maximum (negative deviation) or a minimum (positive deviation) in the year, it means that a possible shift may occurred in year *K*.

The significance of the shift can be tested using the *rescaled adjusted range R*, which is defined by:

$$R = \left(\max_{0\le k\le n} S_k^* - \min_{0\le k\le n} S_k^*\right)/s$$

Usually, large values of R indicate shifts.

von Neumann Ratio (von Neumann, 1941)

This method defines if the tested serie is homogeneous or inhomogeneous, giving no information about the location of the shift. The *von Neumann Ratio* test is defined by:

$$N = \sum_{i=1}^{n-1} (Y_i - Y_{i+1})^2 \Big/ \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

where $\bar{Y}$ represents the mean value of Y_i.

Under the null hypothesis H_0, it is assumed that Y_i have the same mean. If the serie is homogeneous, the expected value of *N* is two (***E(N) = 2***), moreover if the serie is inhomogeneous, i.e., if the serie contains a shift, the value of *N* tends to be lower (***E(N) < 2***). However, series with rough variations in their mean can cause values rise above two.

Validation

Study of Statistical Significance

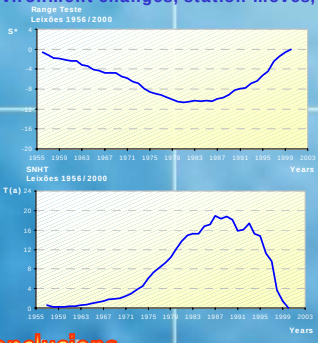
Note that, any statistical test, related to the null hypothesis H_0 and to the alternative hypothesis H_A, is a rule that allows (taking into account the sample point $(X_1, X_2, \ldots, X_N)$) to accept H_0 (reject H_A) or to reject H_0 (accept H_A).

Metadata

Historical information is a useful tool to find explanations and justify the inhomogeneities that are found statistically. Many times it allows justify the obtained results by the knowledge of possible anthropogenic causes, like: environment changes, station moves, instrument changes, etc.

Table 1

SST	E(N)
Portugal Mainland	
Leixões (1956 - 2000)	0.9
Peniche (1956 - 2000)	1.5
Forte do Cavalo (1987 - 2000)	1.3
Sines (1956 - 2000)	0.7
Praia da Rocha (1956 - 1993)	0.9
Azores	
Ponta da Barca (1988 - 2000)	1.4
Pelas (1989 - 2000)	1.2
Arnel (1988 - 2000)	1.5
St.ª Cruz das Flores (1978 – 2000)	0.7
Madeira	
Vila Baleira (1988 - 2000)	1.3

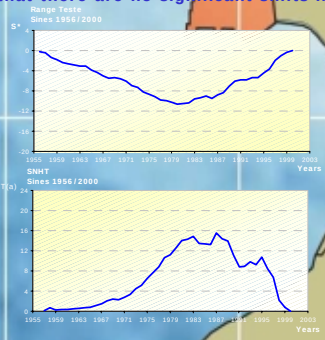


Results Evaluation

The homogeneity tests have identified a small number of suspect shifts in the SST time series, but in two of them were recognized significant shifts.

In the case of Leixões it was located a maximum value around 1986/87 and for Sines it was also identified a maximum values located in 1981/82 and 1986/87. All this values causes the rejection of null hypotesis (H_0) at the 1% level, with strongly significance. *von Neumann Ratio results* (Table 1) also show the existence of a shift in this time series (***E(N) <2***), but it is not known if this is due to artificial or natural changes.

St.ª. Cruz das Flores (Azores) and Vila Baleira (Madeira) maximum values causes the acceptation of null hypotesis (H_0) at the 1% level, which means that there are no significant shifts in these time series.



Conclusions

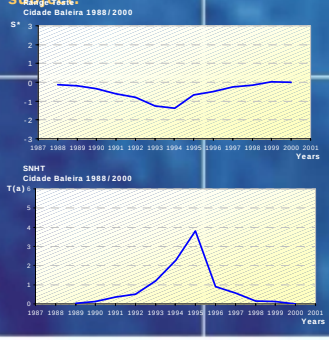
The application of the three homogeneity statistical test, not only allows heterogeneities identification in SST Portuguese time series, but also permits the statistical significance validation.

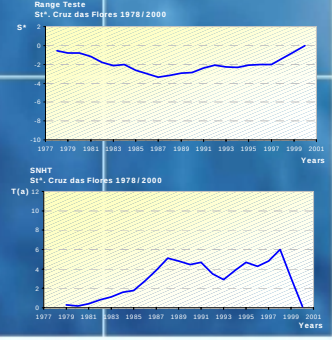
At all tested SST time series, a small number of suspect shifts with strongly significance was identified (p. ex. Leixões and Sines).

Unfortunately, it was not possible justify all-important shifts, because there was no information available of metadata yet that will provide the identification and correction of these inhomogeneities.

However this was an important and necessary control process of SST quality, which let identify and denotes heterogeneities as doubtful to future applications.

Many of questions being maintain. Presently, to a few series it can be possible apply the estimation of the statistical correction with absolute confidence, though they are plans for further work in this subject.





Referências

Alexandersson, H. (1986): "A homogeneity test applied to precipitation data". *Journal of Climatology*, 6, 661-675.

Buishand, T. A. (1982): "Some methods for testing the homogeneity of rainfall records". *J. Hydrology*, 58, 11-27.

Conrad, V. and L.W. Pollak, (1950): "Methods in Climatology". Harvard Univ. Press, Massachusetts.

Jarusková, D. (1994): "Change-point detection in meteorological measurement. Mon. Wea. Rev., 124, 1535-1543.

Von Neumann, J. (1941): "Distribution of the ratio of mean square sucessive difference to the variance". *Ann. Math. Stat.*, 12, 367-395.