Workshop on Drought Predictability and Prediction in a Changing Climate. 2-4 March 2011, Barcelona, Spain

Regional Frequency Analysis of Droughts: Portuguese Case

João Filipe Santos[1], Maria Manuela Portela[2] and Inmaculada Pulido-Calvo[3]

Contact: joaof.santos@estig.ipbeja.pt

[1] ESTIG, Instituto Politécnico de Beja, Rua Pedro Soares, 7800-295 Portugal; [2]SHRH, Instituto Superior Técnico (Lisboa), Avda. Rovisco Pais, 1049-001, Portugal; [3]EPS, Campus de La Rábida, Universidad de Huelva, 21819 Palos de la Frontera, Spain.

1. Introduction

A common problem in drought risk analysis relates to the assessment of the rarity of the events, such as long duration droughts or high magnitude droughts. Being a frequent phenomena in the Southern Europe and in others regions of the world drought constitute a primary natural hazard for human activities. For this reason, and for an improved drought risk management, the preparation of drought hazard maps is an important and urgent task.

The drought definition based on deviations from normal conditions or from reference stages implies that droughts can occur in any hydroclimatological region and at any time of the year with the same probability. In order to due so, a large number of quantitative drought characteristics must been considered. Two common approaches to select extreme events from a drought index time series were analyzed: the annual maximum series (AMS) and the partial duration series (PDS) approaches.

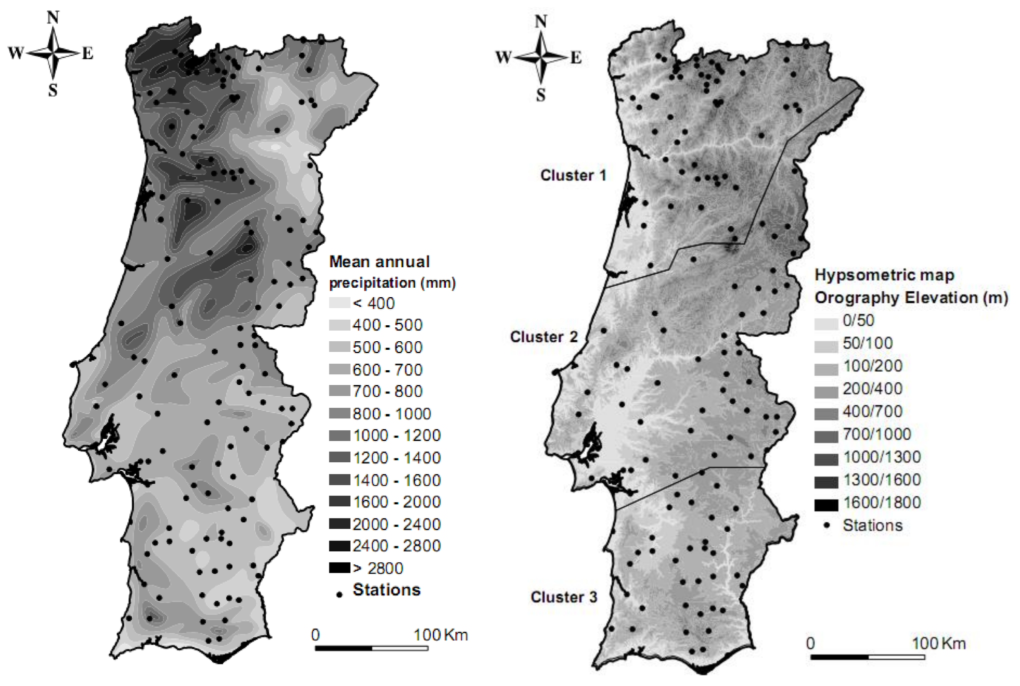


Figure 1 - Mean annual rainfall in Portugal and Three regions with different drought temporal pattern obtained by cluster analysis.

2. Study area and data processing

The study presented herein was accomplished for **Portugal**. In *Santos et al.* [2010] the general spatial pattern of droughts in Portugal was identified by using the standardized precipitation index (SPI). Based on 94-years of monthly precipitation records (from 1910/11 to 2003/04) at the 144 rain gages schematically located in Figure 1, the SPI was calculated for several time scales. By using principal component analysis and K-means clustering *Santos et al.* [2010] proved that three well defined sub regions with droughts with different patterns could be considered: the North, the Center and the South of Portugal (clusters 1, 2 and 3 of Figure 1, respectively).

Using those sub regions the temporal pattern of the droughts are characterized so that the frequency analysis of droughts in each one can be accomplished. Drought data was extracted from the 1, 3, 6 and 12-month SPI time series, **SPI1, SPI3, SPI6** and **SPI12**, respectively.

According to *McKee et al.* [1993] the concept of drought comprehends a beginning date, an ending date, a current drought intensity and a drought magnitude -Figure 2-. The study was focused on drought magnitude. For that purpose the time series of the SPI at the 144 rain gages of Figure 1 were established and analyzed by means of the two following approaches:

- **AMS approach** which considers the series built upon the severest event in each year: one drought per year, the one with the highest magnitude, given that such drought event is larger than a given threshold. The years without droughts are assigned to zero value.
- **PDS approach**, where the magnitudes of all drought events lower than a given threshold are taken into account resulting in a much more comprehensive series with any number of events per year: the first, the second, third, etc. largest drought events , provided those events are mutually independents.

To identify the droughts a **threshold of -0.84 was adopted** [Agnew, 2000].

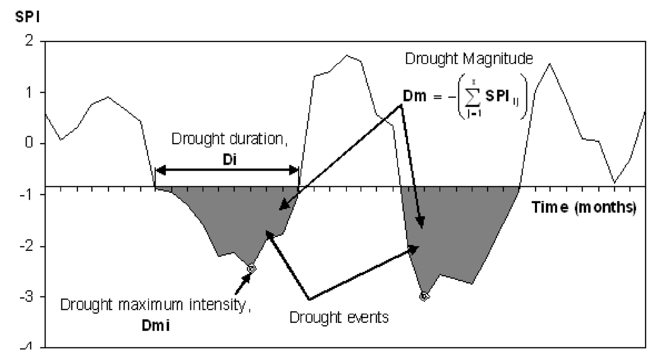


Figure 2 - Definition of drought properties based on the SPI index.

3. Methodology

3.1. Return period

In the present study the traditional concept of return period was considered, Haan [1977]. Since some probability distribution functions are not defined for zero (x = 0), the **AMS** used a corrected non-exceedance probability F', according to the following expression:

$$F' = q + (1-q)\,F \;\Rightarrow\; T = \frac{1}{1-F'}$$

where q is the probability of zero values. The **PDS** approach assumes that the frequency of the events in a given time period is random. Under the Poisson assumption, the return period, was calculated as:

$$T = \frac{1}{\lambda\,(1-F)}$$

where λ is the average number of occurrences per year.

3.2. Regional frequency analysis

The **regional frequency analysis** was applied separately in each one of the **three cluster regions** represented in Figure 1. The number of rain gages grouped in each one of the three cluster regions was **56, 53** and **35** (**cluster 1, 2** and **3**, respectively) **144** in **total**. The **regional frequency analysis** adopted the Hosking and Wallis methodology [Hosking and Wallis, 1997], and follows basically two steps: (i) Screening of the data and testing of regional homogeneity; (ii) Identification of the regional probability distribution function.

3.3. Screening the data and testing the regional homogeneity

Hosking and Wallis [1993] developed one statistic based on the L-moments, called **discordancy measure (S_i)**, with the purpose of identifying those sites from a group of given sites that are grossly discordant with the group as a whole. The measure, S_l, for the station l is given by:

$$S_l = \frac{N_s}{3(N_s-1)}\left(u_l-\bar{u}\right)^{T_m} S_{cm}^{-1}\left(u_l-\bar{u}\right)$$

It is used the **limit** of **S_i = 3** for a station to be considered part of a region as well as the restriction of regions with $N_s \geq 7$ must be considered.

To identify what would be expected in a homogeneous region a number of 500 data sets are generated using a four parameter distribution, i.e. kappa distribution. For each dataset or region a statistic V_k is calculated and based on the vector of the V_k values the mean, μ_{vk} , and the standard deviation, σ_{vk} , are calculated. Then the **heterogeneity measure H_k** is computed as stated in *Hosking and Wallis,* 1997:

$$H_k = \frac{V_k - \mu_{vk}}{\sigma_{vk}}$$

A group of sites is "**acceptably homogeneous**" if H_1 <1, "**possibly heterogeneous**" if 1 ≤ H_1 <2, and "**definitely heterogeneous**" if H_1 ≥ 2.

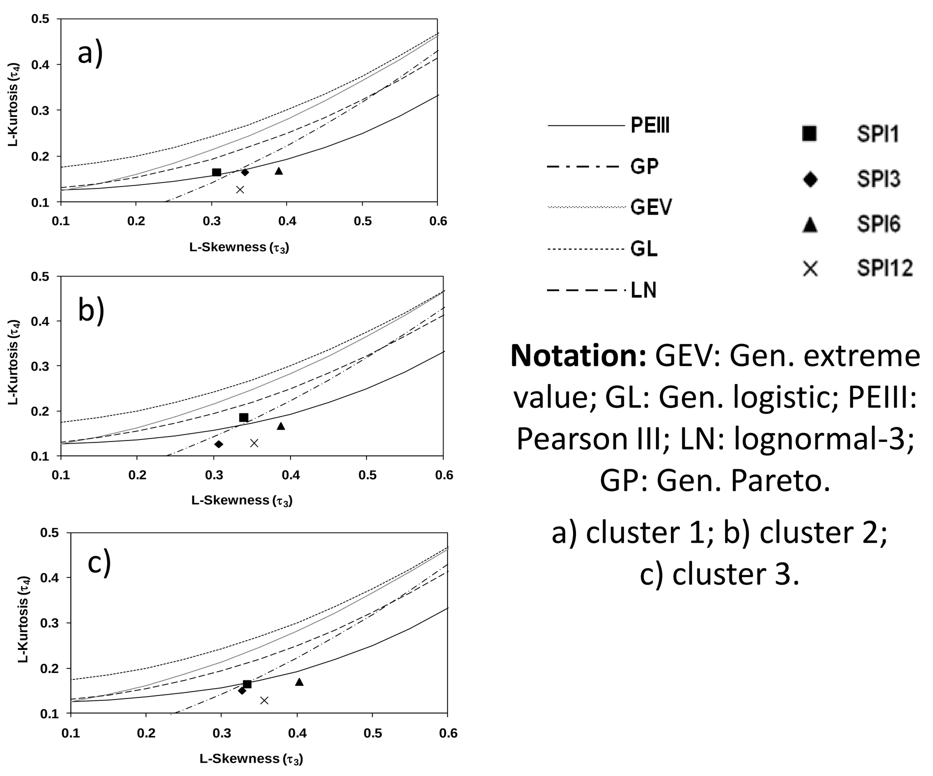
3.4. Identification of regional distribution models

Different **probability distribution models** were applied to the absolute values of the **drought magnitudes**. **Five three-parameter distributions**, i.e., Generalized Logistic (GL), Generalized Extreme Value (GEV), Generalized Pareto (GP), Log Normal (LN) and Pearson Type III (PEIII) were assessed in each region.

To identify the best probability distribution function in each region the **L-moments ratio diagrams** and a goodness of fit measure known as **Z-statistic** were used, *Hosking and Wallis* [1997].

$$Z = \frac{\left(\bar{t}_4^{R} - t_4^{dist}\right)}{\sigma_{t_4}}$$

If based on a given probability distribution |Z|≤1.64 is obtained, then it is assessed to be an acceptable distribution for representing sample data at α=10% (bilateral test).



Notation: GEV: Gen. extreme value; GL: Gen. logistic; PEIII: Pearson III; LN: lognormal-3; GP: Gen. Pareto.

a) cluster 1; b) cluster 2; c) cluster 3.

Figure 3 - AMS.

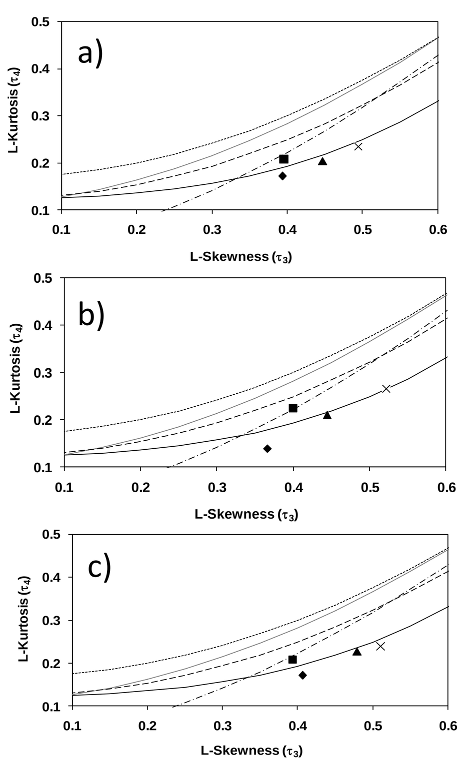


Figure 4 - PDS.

4. Data analysis and results

4.1. Regional homogeneity tests and regional frequency distributions

According to the **heterogeneity measure** , **H_1** , comparatively to the PDS approach, the AMS led to better results in all clusters. **The ratio L-moments diagrams** are represented in Figure 3 and Figure 4 for AMS and the PDS, respectively. For the **AMS approach** (Figure 3), the better choices for cluster 1 are the PEIII distribution, for the time scales of 1, 3 and 6 months, and the GP distribution, for 12 months. For cluster 2, the PEIII fits closely the location of SPI6 and SPI12 while the GP distribution better describes SPI1 and SPI3. The results for cluster 3 are similar to those of cluster 2.

For the **PDS approach** (Figure 4) a wider distribution of the average points was obtained. In every sub region there is a close match between the SPI1 and the GP while the PEIII distribution provides the better model for the remaining. The values of the **Z statistic** for the PEIII and GP were obtained. Whenever |Z| is larger than 1.64 the four parameter kappa distribution was adopted. To compare between drought magnitudes provided by the AMS and the PDS approaches an error measure, **RMSE**, was also applied. The results achieved suggest the better performance of the AMS approach, with better results.

4.2 Drought frequency analysis

Maps of drought magnitude were created for the return periods of 50 and 100 years - Figures 5 to 8 -. A **quantitative analysis** of the **drought duration** applied to the AMS magnitude series was accomplished. The results are represented in the histograms of Figure 9 [a) Northern region, cluster 1; b) Central region, cluster 2; and c) Southern region, cluster 3].

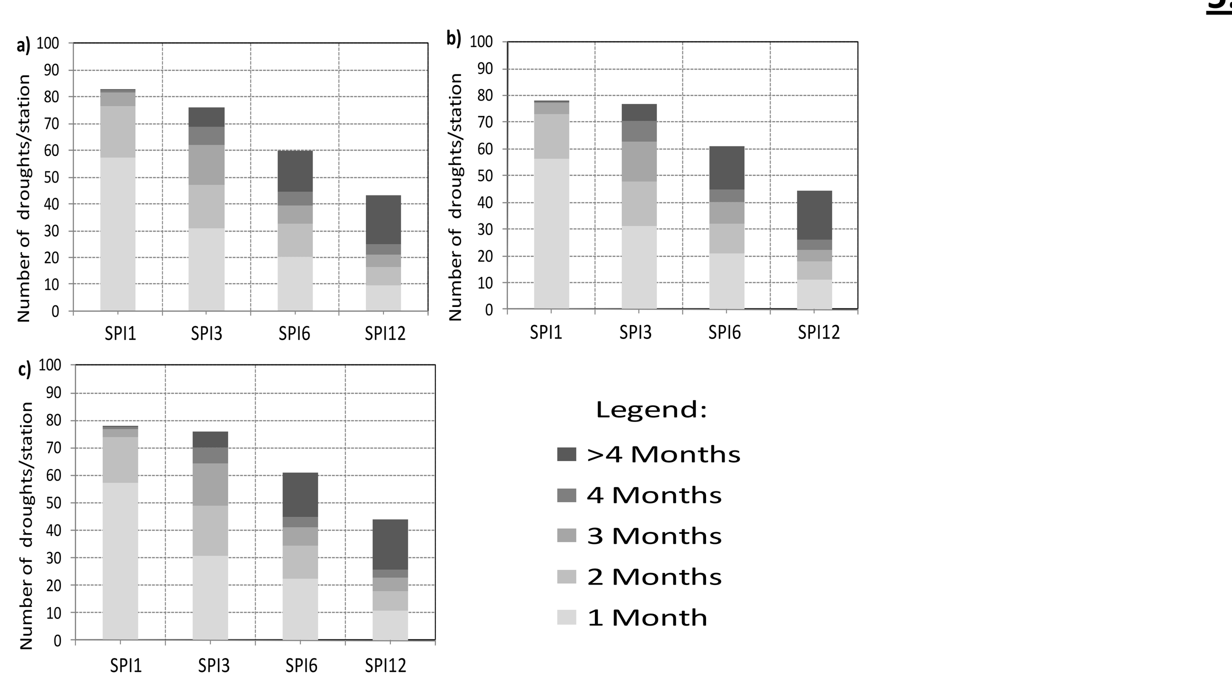


Figure 9 - Weighted number of drought occurrences per station.

5. Discussion of the Results

In general terms several authors state the PDS approach as being more efficient than the AMS which in fact was not the result here obtained. However, Madsen et al. (1997) have shown better performance of the AMS approach when they applied the generalized extreme values distribution with positive shape parameters, according to the probability weighted moments estimation method (PWM), which was also our case for all the regional parent distributions applied. The results also reflect the threshold adopted of -0.84. Some authors found that the final quantile estimates vary significantly when only small changes are made in the threshold. When drought evaluation studies are envisaged the values of the threshold are fixed which compromise the discussion about the PDS approach. So it is not clear the advantage of using the PDS approach when modeling the spatial distribution of drought characteristics.

a) 50 years; and b) 100 years

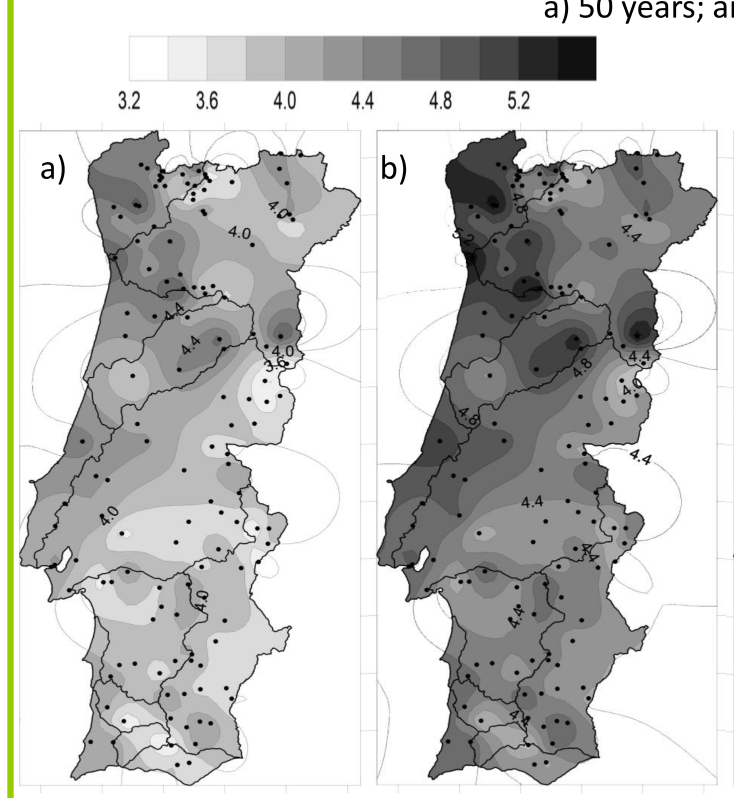


Figure 5 - AMS approach applied to SPI1.

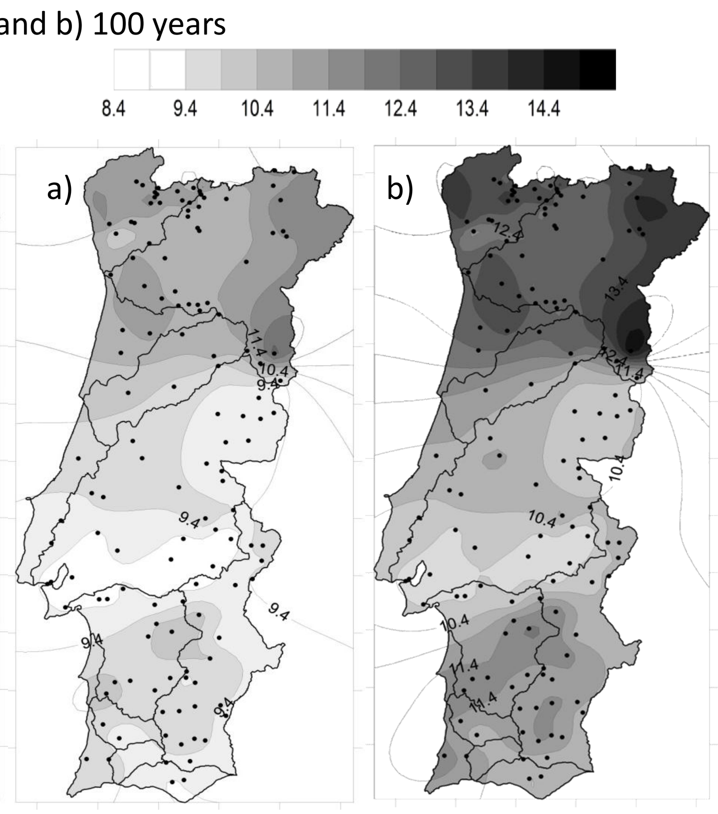


Figure 6 - AMS approach applied to SPI3.

a) 50 years; and b) 100 years

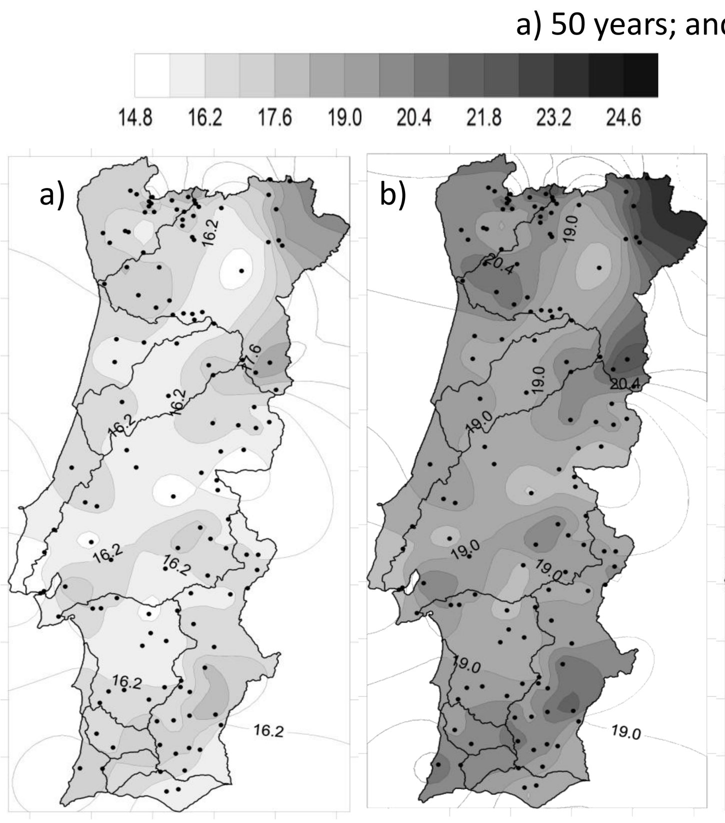


Figure 7 - AMS approach applied to SPI6.

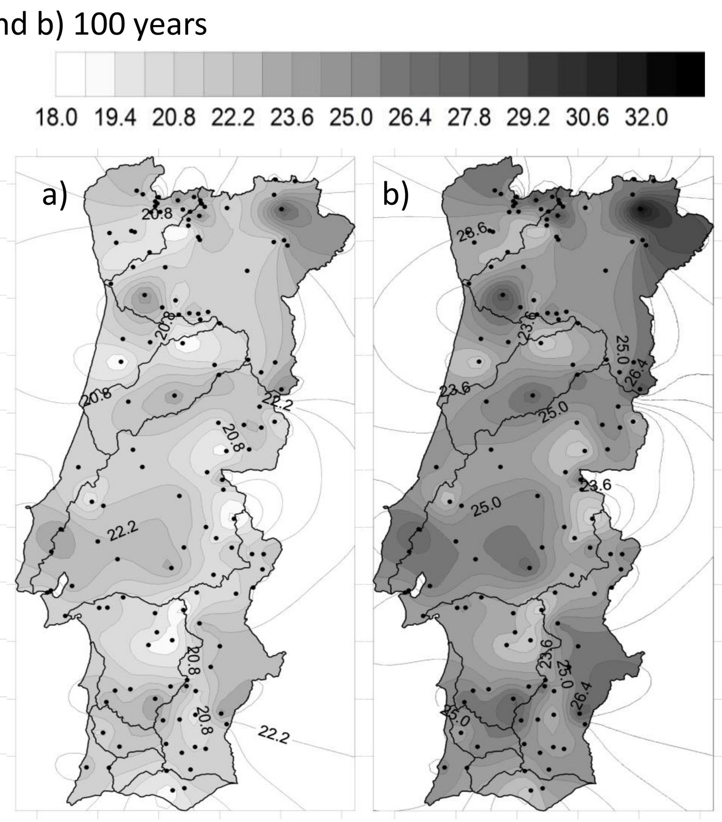


Figure 8 - AMS approach applied to SPI12.

6. References

Hosking, J.R.M. and J.R. Wallis (1993), Some statistics useful in regional frequency analysis, WRR, 29(1), 271-281.

Hosking, J.R.M. and J.R. Wallis (1997), *Regional frequency analysis – An approach based on L-moments*, 224 pp., Cambridge University Press, UK.

Santos, J.F., I. Pulido-Calvo and M.M. Portela (2010), Spatial and temporal variability of droughts in Portugal, WRR, 46, W03503.

Agnew, C.T. (2000), Using the SPI to identify drought, Drought Network News, 12, 6-12.

McKee, T. B., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, in Proceedings of the 8th Conf. on Ap. Clim., *Ame. Met. Soc.*, 179-184.

Haan, C.T. (Ed.) (1977), Statistical Methods in Hydrology, 378 pp., The Iowa State University Press, Iowa, USA.