# MEHODS FOR PREDICTING DROUGHT OCCURRENCES

Isabella Bordi[1], Klaus Fraedrich[2], Marcello Petitta[1] and Alfonso Sutera[1]
[1]*Department of Physics, University of Rome "La Sapienza", P.le Aldo Moro 2 00185 Rome, Italy*
[2] *Meteorologisches Institut, Universität Hamburg, D-20146 Hamburg, Germany*

**Abstract**: A comprehensive assessment of future dry events in a region is essential for finding sustainable solutions for water-related problems concerning water management and preventive risk assessment of drought. It is also a first step in developing a common European strategy for facing up to the negative impacts of present and future dry climatic events. A useful index for drought monitoring, based only on monthly precipitation, is the Standardized Precipitation Index (SPI); here we assume that it properly describes the climatic condition of a particular region. By applying an appropriate forecast method to the precipitation time series and then computing the SPI, it is possible to predict future drought occurrences. Forecasting time series is a well-known field of Statistics and methods are well documented in the scientific literature. In the present paper we apply a standard technique and we contrast it with a new method here presented. As a test case, we use rain gauge data for Sicily, which may be considered a key region for understanding climatic conditions in the Mediterranean basin.

**Key words**: water resources - precipitation - drought indexes – forecasting methods

## 1. INTRODUCTION

Predicting future dry events in a region is an important first step for finding sustainable solutions to water management and risk assessment of drought occurrences. However, it remains a complex task, because the random character of precipitation field, which is the basic variable commonly used for drought assessment. Drought, in fact, is a normal recurrent feature of climate occurring in all climatic zones and it originates from a deficiency of precipitation over a given period of time: short time scales (months) characterize meteorological drought, while longer time scales (years) hydrological drought.

Recently, an index based only on monthly precipitation, the Standardized Precipitation Index (SPI), has been proposed to monitor dryness and wetness on multiple time scales (McKee, 1993). As pointed out in many papers (Hayes et al., 1999, Keyantash and Dracup, 2002, Bordi and Sutera, 2001), the SPI, because its standardized nature (it is based on an equal-probability transformation of the empirical probability density distribution of precipitation for a particular month of the year into the Normal distribution), quantifies properly the meaning of relative dryness and wetness, allowing comparing climatic conditions of areas governed by different hydrological regimes.

Assuming that the SPI describes all the facets of a dry condition, it may be possible to exploit long time series of this index to forecast future occurrences of drought. However, there are at least two main hindrances that do not allow a readily solution to the forecast by standard time series methods (Box and Jenkins, 1970):

i) Precipitation is usually not correlated on the time scale on which drought manifests itself;

ii) The time correlations found in the SPI on long time scales are simply an effect of cumulating precipitation on the selected time scale, unless SPI's multiyear periodicities are revealed.

In summary, the application of forecast methods to SPI time series appears to be a wrong approach from the out start. Thus, we must consider precipitation and exploit the knowledge acquired by sampling it for a long time.

A technique, commonly used in Statistics for predicting the future behavior of a given time series is the Auto Regressive model (AR). So that, if a long time series of precipitation is at hand, we may use AR model to estimate its future behavior. It is intuitive, however, that an AR method would extract the seasonal cycle of precipitation as the leading prediction. However, this forecast may have a small skill in predicting drought, since the latter depends strongly on the departure from the seasonal behavior of precipitation. An alternative approach, here proposed, may be the estimation of the probability function of the monthly precipitation for forecasting (with a known probability) future precipitation values and, thereby, the SPI.

In the present paper we illustrate these two forecast methods for predicting drought on short time scales, so that no artificial correlations are introduced. Specifically, we consider the SPI at one-month time scale (SPI-1) as forecast target. Then, we quantitatively compare the forecast skills obtained using the two approaches: a standard AR model and a new approach, here denoted as the Gamma Highest Probability (GAHP) method. As a test case, we use precipitation data averaged over 36 stations in Sicily, a region which may be considered a key area for studying climatic conditions of the Mediterranean basin.

The paper is organized as follow. Section 2 describes the data and the climatology of the region, while theoretical fundaments of the two forecast methods are presented in section 3. Section 4 is focused on the comparison between the results obtained from the AR and GAHP methods. In the final section conclusions and suggestions for future investigations are discussed.

## 2. DATA AND CLIMATOLOGY

To evaluate the ability in predicting future values of the SPI-1, we use monthly precipitation time series from 36 stations in Sicily covering the period 1926-2000. The stations have been extracted from a larger set according to Alecci et al. (2000) criterions, which are mainly the record length, data quality and homogeneous spatial distribution. For illustrative purposes, we decided to consider the monthly precipitation averaged over the 36 stations, neglecting local variations since they add very little to the understanding of the problem. In computing the SPI on 1-month time scale we employ the algorithm presented in Bordi and Sutera (2001).

To better understand the procedure used in the following section, it is useful to describe the main features of the climatic conditions in Sicily. The climate in Sicily is characterized by precipitation in the autumn and the winter seasons with dry summer. As shown in Bordi et al. (2005), extreme wet events are similarly probable in every month, instead dryer events are concentrated during the months when the larger precipitation events occur, i.e. fall and winter. The precipitation occurring in Sicily are produced by advective systems coming from the Atlantic Ocean, while the South-Eastern part of the island is affected by rainfall originated in the Levantine Mediterranean region. All this events are driven by the seasonal variability and they are not related with the precipitation observed in the previous months. This is because the water cycle has large horizontal scale features that are driven by synoptic meteorological disturbances interacting with orography. Thus, there are no correlations between precipitated water and next precipitation event as it occurs at small spatial scales in some continental areas of the U.S., where precipitation may be originated by the water previously fallen. Thus, for the region of our interest, we should expect very weak time correlations (if any at all) in the precipitation time series.

Moreover, the nature of the probability density function of the precipitation for each month of the year is well fitted by the Gamma-2 function, as shown in figure 1, though some deviations are noticeable. The highly skewed nature of these distributions already reveal the difficulty of assessing extremes and the merit of using the SPI, which, by means of an equal-probability transformation, standardizes the random variable (precipitation).

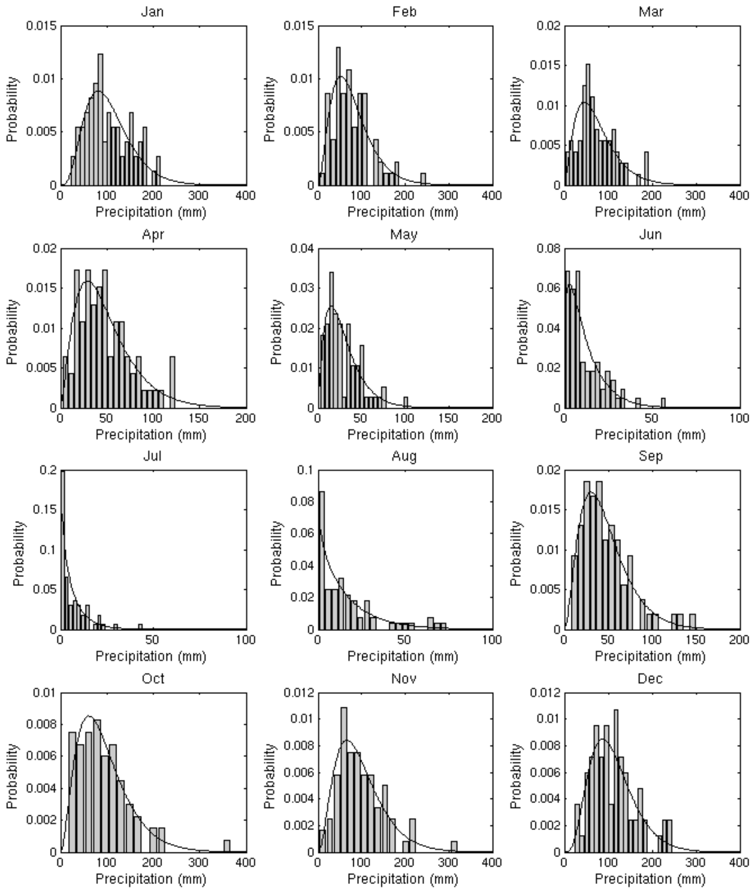


*Figure 1*. Probability density distributions of the observed precipitation for each month of the year. Solid lines denote the Gamma functions fitting the distributions.

## 3. AR AND GAHP FORECAST METHODS

To forecast the SPI-1 values we use two different approaches: the AR and the GAHP method. In this section we briefly describe the theoretical assumptions of these techniques.

*AR method:* A common approach for modeling multivariate time series is the autoregressive model of order $p$ (AR($p$) model):

$$x_t = w + \sum_{l=1}^{p} A_l x_{t-l} + \varepsilon_t$$

where $x_t$ is the $m$-dimensional state vector that have been observed at equally spaced instants $t$. The matrices $A_1,...A_p \in \Re^{m\times m}$ are the coefficient matrices of the AR model and the $m$-dimensional vector $\varepsilon_t = noise(C)$ are uncorrelated random vectors with mean zero and covariance matrix $C \in \Re^{m\times m}$. The $m$-dimensional vector $w$ is a vector of intercept terms, which allows for a nonzero mean of the time series.

First, the approach needs the estimation of the following parameters: the order $p$ of the AR model, the intercept vector $w$, the coefficient matrices $A_1 \ldots A_p$ and the noise covariance matrix $C$. For the selection of these parameters, the stepwise least squares algorithm is usually implemented (Neumaier and Schneider, 2001). Basing finite-sample inferences on the asymptotic distribution of the least squares estimator makes it possible to construct approximate confidence intervals for the intercept vector and for the coefficient matrices. In the present paper we construct the 95% confidence intervals.

Then, before the AR model is used for predictions, it is necessary to assess whether the fitted model provides an adequate representation of the given time series. Various tests of the adequacy of a fitted model are described by Brockwell and Davis (1991) and by Wei (1994). The primary statistical tool for most process modelling applications is the analysis of the residuals, i.e. the differences between the responses observed in the data set and the corresponding prediction of the response computed using the regression function:

$$\hat{\varepsilon}_t = x_t - \hat{w} - \sum_{l=1}^{p} \hat{A}_l x_{t-l}, \quad t = 1,\ldots,N$$

where the hat-accent designates an estimate of the quantity $w$ and $A$. A principal assumption intrinsic to AR models is that the noise vectors be uncorrelated. Thus, the autocorrelation function of the residuals can be estimated graphically.

Finally, a realisation of an AR process can be simulated by substituting Gaussian pseudorandom vectors with covariance matrix $C$ for the noise vector $\varepsilon_t$ in the AR model.

Thus, we start our analysis by computing the SPI on 1-month time scale (see figure 2a), trying to forecast its future values. However, it must be noted that the SPI-1 time series cannot be predicted using AR method because, as expected, it has a white noise power spectrum as clearly shown in figure 2b.

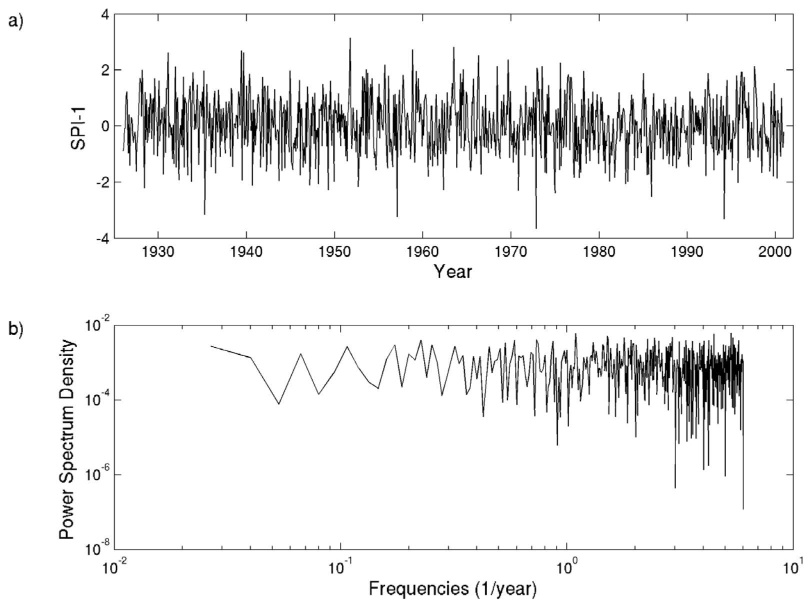


*Figure 2*. a) SPI-1 for Sicily from 1926 to 2000 computed using the precipitation averaged over 36 stations; b) Power spectrum density of the SPI-1.

The inability of the AR method to forecast any future value of the SPI-1 could be overcame by applying the AR method to precipitation time series. Using AR method to evaluate future values of precipitation we find, however, more implications. The AR method gives, in fact, a regression of 12$^{th}$ or 13$^{th}$ order. Comparing the weights of the coefficients, it is possible to note as the first, the eleventh, the twelfth and the thirteenth are the highest ones. The AR method finds correlation in two main components: the seasonal cycle, mixing among the 11$^{th}$, 12$^{th}$ and the 13$^{th}$ coefficients, and a simple regression with the previous event (the first coefficient), typical of the linear AR method. This result is clearly illustrated in figure 3 where the precipitation time series (3a) and the associated power spectrum (3b) are shown. The power spectrum has a peak corresponding to the annual component, though mixed with the nearest frequencies. It follows that AR is just extracting the seasonal cycle, thus preventing to have a good knowledge of large deviations from it. Recalling that the latter ones lead to extreme conditions, it appears that the skill of this method should be low, especially for extremely dry (or wet) occurrences.

*GAHP method:* In avoiding these shortcomings, we propose a method that we denote as GAHP (GAmma Highest Probability) method. This approach forecasts the precipitation for a future month as the most probable value described by the probability density function of the precipitation for that month.

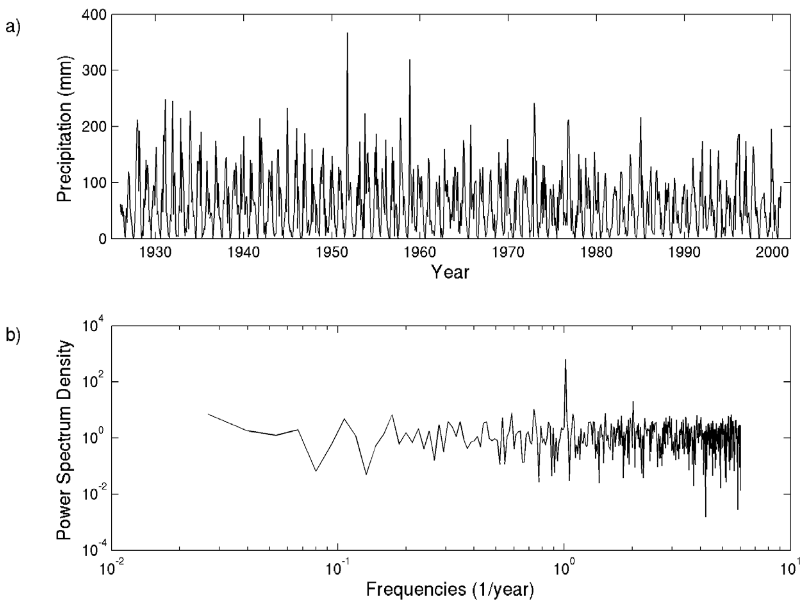


*Figure 3.* a) Monthly precipitation averaged over the 36 stations in Sicily from 1926 to 2000; b) Power spectrum density of the precipitation time series. Note the peak at 1 year period.

Thus, the method needs the estimation of the parameters of the Gamma distribution function that best fits the frequency histogram of the observed precipitation for a given month of the year. Then, the precipitation predicted for the next month is the *mode* of the fitted distribution. This approach rests on three hypotheses:

1. two consecutive precipitation events are not correlated;
2. the relation between precipitation values are only related to the seasonality;
3. future events will be the most probable ones.

The first two assumptions are both related to the phenomenology of the precipitation for a particular month in the Mediterranean regions as discussed in the section above. The third has an empirical nature and may be easily substituted by other location measures of a given distribution.

## 4. SPI-1 FORECAST FOR SICILY WITH AR AND GAHP METHOD

In this section we compare the forecast obtained with the AR and the GAHP method. The procedure used to perform the forecast of the SPI-1 for the next year can be summarized as follows:

- extraction of a subset from the entire precipitation dataset (for example, from January 1926 to December 1999);
- computation of the 12 predicted values for monthly precipitation (one year) with the two methods;
- evaluation of the SPI-1 with the new precipitation time series;
- comparison between the predicted SPI-1 and the one obtained with the entire observed precipitation time series 1926- 2000.

To estimate the goodness of the prediction we use the mean squared prediction error (MSE) given by:

$$MSE(t_i) = \left[SPI_o(t_i) - SPI_F(t_i)\right]^2$$

where $SPI_o$ is the observed value, $SPI_F$ is the forecast value and $i$ denotes the month of the year.

*Results:* In figure 4 we show the precipitation predicted using the two methods and the real observations for the year 2000 (Black bars). Gray bars refer to the precipitation obtained with the AR method. In this case it is possible to note as the annual cycle and the relationship between the previous events are the main features of the predicted signal. As a matter of fact, the AR method estimates the future event, in first approximation, as the mean value of the precipitation observed in that month because this method is based on hypothesis that the observed variables have a Gaussian distribution. However, the precipitation is not a variable Gaussian distributed in any month of the year and, moreover, the mean value of precipitation is not the most probable value.

White bars refer to the forecast obtained from the GAHP method. The difference between the observed and the predicted precipitation is smaller than in the previous case. In fact, the GAHP method appears to perform a better prediction both for the seasonal variability and the deviation thereof.

In order to have a larger statistics and to better validate our claimed skill, we apply the two methods to different time segments of precipitation time series and compute the mean MSE of the resulting SPI-1. Thus, we form a set of 75 random permutations of the precipitation data that preserve the right sequence of the 12 precipitation values registered in a year, i.e. the seasonal variability. For each permutation we evaluate the coefficients of the AR method and we compute the forecast for the last year. It must be noted that for our method it is not necessary to do any permutation because the estimation of the most probable value does not depend on the sequence of the value registered in a certain month. In this case, to obtain a test statistically significant we evaluate the MSE of the SPI-1 taking the squared difference between the predicted value of the index for a particular month and all the values for that month in the previous years.

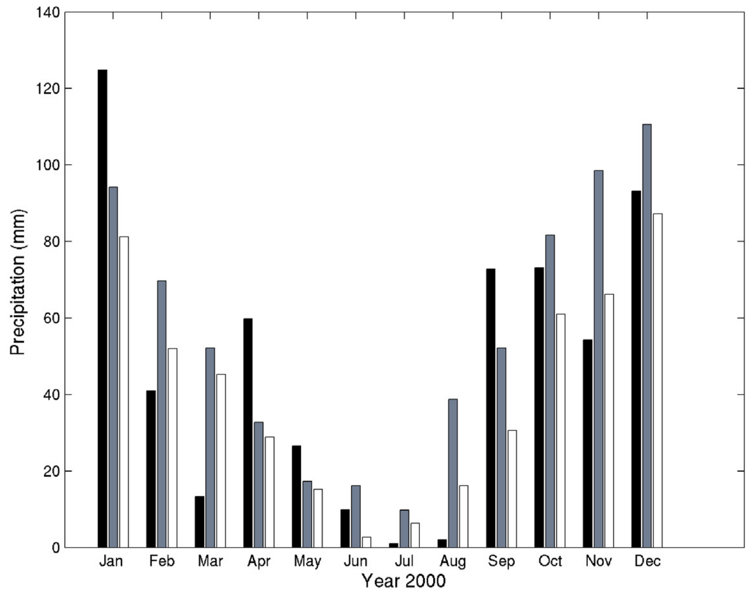


*Figure 4.* Monthly precipitation in Sicily (black) for year 2000, the one predicted with the AR model (gray) and the one predicted with the GAHP method (white).

Results are shown in figure 5. In this case the GAHP method gives consistently better results than the AR method, especially for spring and summer. In fall and winter, when the precipitation has highest values, the AR estimation seems as accurate as the GAHP one, although for these cases the difference between the observed and the forecasted values is too small to make any firm conclusion.

## 5. CONCLUSIONS

Recently, an index based only on monthly precipitation, the SPI, has been proposed to assess dry and wet conditions of a particular location on multiple time scales. It quantifies properly the meaning of relative dryness and wetness allowing comparing climatic conditions of areas governed by different hydrological regimes. Here we assume that the SPI is a good index for drought assessment and try to forecast its future values relative to the next twelve months. We focus the paper on meteorological droughts occurring on short time scale, so that no artificial correlations are introduced. In particular, we consider the SPI on 1-month time scale. By applying an appropriate forecast method to the precipitation time series and then computing the SPI, it is possible to predict future drought occurrences.

In the present work we propose a new method for forecasting dry events, the GAHP, and we compare the results obtained with those resulting from the application of a standard prediction method, say the AM model. As a test case, we use monthly precipitation averaged over 36 stations in Sicily, a region that may be considered a key area for understanding the climatic regime characterising the Mediterranean basin.

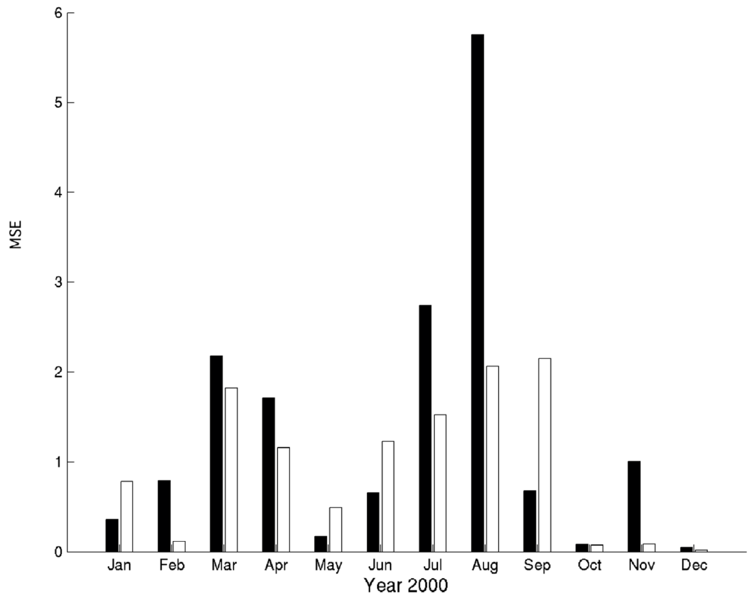


*Figure 5.* MSE of the SPI-1 computed forecasting precipitation with the AR (black bars) and GAHP method (white bars) averaged over the 75 realizations.

The AR method applied to precipitation just extracts the seasonal cycle, preventing to have a good knowledge of large deviations that, generally speaking, are responsible for drought occurrences. In first approximation, the method estimates the future precipitation as the mean value of precipitation observed for that month, because it is based on the hypothesis that the observed variable is Gaussian distributed. The new method, instead, assumes that the precipitation predicted for the next month is the mode of the distribution that best fits the empirical distribution of precipitation for a given month of the year. The GAHP method seems to provide better results for the estimation of future SPI-1 values when compared with those obtained from observed precipitation time series, especially for spring and summer.

Future efforts should be devoted to forecast drought occurrences on longer time scales, which is important for agriculture and hydrologic applications.

## ACKNOWLEDGEMENTS

The financial supports provided by the EU INTERREG IIIB-MEDOCC SEDEMED II project (Bordi, Petitta and Sutera) and the EU NEST E2-C2 project (Fraedrich) are acknowledged.

We thank Eng. Giuseppe Geraci and Eng. Teodoro Vega from the Hydrographic Office of Palermo in making the precipitation time series for Sicily available.

## REFERENCES

Alecci S, Arcidiacono F, Bonaccorso B, Rossi G (2000) Identificazione delle siccità regionali e sistemi di monitoraggio, Atti del Convegno Siccità: monitoraggio, mitigazione, effetti. Villasimius, 21-23 Settembre 2000

Bordi I., Fraedrich K., Petitta M. and Sutera A. (2005) Extreme value analysis of wet and dry periods in Sicily. Submitted to Theor Appl Climatol

Bordi I, Sutera A (2001) Fifty years of precipitation: some spatially remote teleconnections. Water Resour Manag 15: 247-280

Box GE, Jenkins GM (1970) Time series analysis: forecasting and control, Holden-Day, Inc., San Francisco, CA

Brockwell PJ, Davis A (1991) Time series: theory and methods. 2nd ed. Springer-Verlag, New York, NY

Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV (1999) Monitoring the 1996 drought using the standardized precipitation index. Bull Amer meteor Soc 80: 429-438

Keyantash J, Dracup JA (2002) The quantification of drought: an evaluation of drought indices. Bull Amer Meteor Soc 83: 1167-1180

McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales, Preprints, 8th Conference on Applied Climatology, 17-22 January, Anaheim, CA, pp. 179-184

Neumaier A, Schneider T (2001) Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Trans Math Softw 27, 1 (Mar.), 27-57

Wei WWS (1994) Time series analysis. Addison-Wesley Publishing Co., Inc., Redwood City, CA