STOCHASTIC FORECASTING OF STANDARDIZED PRECIPITATION INDEX

ANTONINO CANCELLIERE[1], GIUSEPPE DI MAURO[2],
BRUNELLA BONACCORSO[3] and GIUSEPPE ROSSI[4]

[1]Assistant Professor, Department of Civil and Environmental Engineering,
University of Catania, V.le A. Doria 6, 95125 Catania, Italy
(Tel: +39-095-7382718, Fax: +39-095-7382748, e-mail: acance@dica.unict.it)
[2] PhD student, Department of Civil and Environmental Engineering,
University of Catania, V.le A. Doria 6, 95125 Catania, Italy
(Tel: +39-095-7382711, Fax: +39-095-7382748, e-mail: gdimau@dica.unict.it)
[3]Assistant Researcher, Department of Civil and Environmental Engineering,
University of Catania, V.le A. Doria 6, 95125 Catania, Italy
(Tel: +39-095-7382711, Fax: +39-095-7382748, e-mail: bbonacco@dica.unict.it)
[4]Full Professor, Department of Civil and Environmental Engineering,
University of Catania, V.le A. Doria 6, 95125 Catania, Italy
(Tel: +39-095-7382702, Fax: +39-095-7382748, e-mail: grossi@dica.unict.it)

Abstract

An improved drought management must rely on an accurate monitoring and forecasting of the phenomenon in order to activate appropriate mitigation measures, based on the risk associated with the possible evolution of a current drought condition.

In the paper, a stochastic model to forecast SPI values at short-medium term has been developed. The forecast is based on the expected value of future values of SPI, conditioned on past observations of precipitation and the resulting analytical expression is a function of past precipitation and of its statistics. Also, analytical expressions of the Mean Square Error (MSE) and of prediction confidence limits of fixed probability are derived, which enable to assess the forecast accuracy. Validation of the model has been carried out with reference to precipitation series observed in 43 stations located in Sicily, Italy, making use of a moving window scheme for parameters estimation. Results show a good agreement between observed and forecasted values, thus suggesting the suitability of the proposed procedure as a tool for drought management.

Keywords: Drought forecasting; Standardized Precipitation Index; MSE; Confidence intervals; Precipitation

1. INTRODUCTION

Drought is a frequent natural phenomenon affecting many regions of the world. It arises from a significant reduction in precipitation with respect to normal values (e.g. long term mean), which has consequences on all components of the natural hydrological cycle, producing negative environmental, economic and also social impacts (Yevievich et al., 1983; Rossi et al., 1992, Wilhite et al., 2000). Regardless of the specific climatic conditions of the interested area, drought impacts can be reduced or even amplified by the interaction with human factors. Indeed, the amount of water demand, the efficiency of the hydraulic infrastructures and the effectiveness of water management criteria can sensibly modify the vulnerability of water supply systems to drought events.

Moreover, since drought consequences are generally perceived with some delay with respect to its inception and persist long after its ending, an effective mitigation activity can be implemented, more than in the case of other extreme hydrological events (e.g. floods), provided a timely perception of an incoming drought and its monitoring are available (Rossi, 2003). Thus, an accurate selection of tools for drought identification, like indices providing a synthetic and objective description of drought conditions, is essential. Furthermore, such tools must enable the forecast of the probable evolution of a current drought, in order to implement appropriate mitigation measures and policies for water resources management.

Within this framework, Karl et al. (1987) assessed the amount of precipitation to restore normal condition after a drought event, with reference to the Palmer index (PHDI). Cancelliere et al. (1996) proposed a procedure for short-medium term forecasting of the Palmer Index and tested its applicability to Mediterranean regions. In particular, they computed the probability that an ongoing drought will end in the following months. Other authors (Lohani et al., 1998) proposed a forecasting procedure of the Palmer index based on first-order Markov chains, which enables to forecast drought conditions for the future months, based on the current drought class of PHDI.

Among the several proposed indices for drought monitoring, the Standardized Precipitation Index (SPI) has found widespread application (McKee et al., 1993; Heim, 2000; Wilhite et al., 2000; Rossi and Cancelliere, 2002). The SPI is able to take into account the different time scales at which the drought phenomenon occurs, and because of its standardization, it is particularly suited to compare drought conditions among different time periods and regions with different climatic conditions (Bonaccorso et al., 2003). Besides, due to its intrinsic probabilistic nature, the SPI is the ideal candidate for carrying out drought risk analysis (Guttmann, 1999). The index is based on an equi-probability transformation of accumulated monthly precipitation into a standard normal variable. In practice, computation of the index requires fitting a probability distribution to accumulated monthly precipitation series (e.g. k= 3, 6, 12, 24 months, etc.), computing the non-exceedence probability related to such accumulated values and defining the corresponding normal standardized quantile as the SPI. McKee et al. (1993) assumed accumulated precipitation gamma distributed and used maximum likelihood method to estimate the parameters of the distribution.

In the present paper, short-medium term forecast of the Standardized Precipitation Index is addressed by means of stochastic techniques. In particular, analytical expressions of SPI forecasts are derived as the expectation of future SPI values conditioned on past monthly precipitation, under the hypothesis of normally distributed precipitation aggregated at different time scales k. The forecast accuracy is evaluated in terms of the Mean Square Error of prediction (Brockwell and Davis, 1996), which allows confidence intervals for prediction to be computed. Validation of the model is carried out with reference to precipitation series observed in 43 stations located in Sicily, Italy, making use of a moving window scheme for parameters estimation.

2. FORECASTING THE STANDARDIZED PRECIPITATION INDEX

From a stochastic point of view, the problem of forecasting future values of a random variable can be seen as the determination of the probability density function of future values conditioned by past observations. Once the conditional distribution is known, the forecast is usually defined as the expected value or the median and confidence intervals of the forecast values can be computed.

In practice, however, the derivation of the conditional probability distribution of future values can be cumbersome in most cases; therefore, usually a function of the past observations that forecast future values is sought instead. More formally, let's consider a sequence of random variables Y_1, Y_2, ..., Y_t. The interest lies in determining a function $f(Y_1, Y_2,..., Y_t)$ that forecast a future value Y_{t+M} with minimum error (see Fig. 1). The latter is usually expressed as the Mean Square Error (MSE) of prediction, defined as (Brockwell and Davis, 1996):

$$\mathrm{MSE}=\mathrm{E}[(\mathrm{Y}_{t+M} - \mathrm{f}(\mathrm{Y}_1, \mathrm{Y}_2,...,\mathrm{Y}_t))^2] \tag{1}$$

It can be shown, that the function $f(\cdot)$ that minimizes the MSE is the expected value of Y_{t+M} conditioned on Y_1, Y_2, ..., Y_t, i.e.:

$$\mathrm{f}(\mathrm{Y}_1, \mathrm{Y}_2,...,\mathrm{Y}_t)=\mathrm{E}[\mathrm{Y}_{t+M} \mid \mathrm{Y}_1, \mathrm{Y}_2,...,\mathrm{Y}_t] \tag{2}$$

The above property allows to derive the "best" forecast (in MSE sense), provided the conditional expectation can be computed. Also, it may be worthwhile to note that if Y_{t+M} is independent of Y_1, Y_2, ..., Y_t, the best predictor of Y_{t+M} is its expected value, and furthermore, the MSE of prediction is just the variance of Y_{t+M}.

Application of the above results to the SPI index can be carried out by making the assumption that the underlying precipitation aggregated at k months is normally distributed. In this case, the SPI value in the year ν and month τ takes the following simple form:

$$Z_{\nu,\tau}^k = \frac{\sum_{i=0}^{k-1} X_{\nu,\tau-i} - \mu_\tau^k}{\sigma_\tau^k} \tag{3}$$

where $X_{\nu,\tau}$ is the precipitation in year ν and month τ, and μ_τ^k and σ_τ^k are, respectively, the mean and the standard deviation of monthly precipitation aggregated at k months, preceding the month τ. Note that the mean μ_τ^k can be written as a function of the monthly means as:

$$\mu_\tau^k = \sum_{i=0}^{k-1} \mu_{\tau-i} \tag{4}$$

The best predictor of the SPI index M months ahead $\widetilde{Z}_{\nu,\tau+M}^k$, given observations up to month τ, will be:

$$\widetilde{Z}^k_{\nu,\tau+M} = E\left[Z^k_{\nu,\tau+M}\left|Z^k_{\nu,\tau}, Z^k_{\nu,\tau-1}, \mathrm{K}\right.\right] = E\left[Z^k_{\nu,\tau+M}\left|\sum_{i=0}^{k-1} X_{\nu,\tau-i}, \sum_{i=0}^{k-1} X_{\nu,\tau-1-i}, \mathrm{K}\right.\right] \tag{5}$$

Since conditioning on aggregated precipitation values is equivalent to conditioning on single precipitation values, eq.(5) can be written as:

$$\widetilde{Z}^k_{\nu,\tau+M} = E\left[Z^k_{\nu,\tau+M}\left|X_{\nu,\tau}, X_{\nu,\tau-1}, \mathrm{K}\right.\right] \tag{6}$$

which upon substitution of eq. (3) becomes:

$$\widetilde{Z}^k_{\nu,\tau+M} = E\left[\frac{\sum_{i=0}^{k-1} X_{\nu,\tau+M-i} - \mu^k_{\tau+M}}{\sigma^k_{\tau+M}}\left|X_{\nu,\tau}, X_{\nu,\tau-1}, \mathrm{K}\right.\right] \tag{7}$$

By assuming monthly precipitation serially independent, the above expression simplifies as:

$$\widetilde{Z}^k_{\nu,\tau+M} = \frac{\sum_{i=0}^{k-M-1}\left(x_{\nu,\tau-i} - \mu_{\tau-i}\right)}{\sigma^k_{\tau+M}} \tag{8}$$

where $x_{\nu,\tau-i}$ represent the observed precipitation at year ν and month τ–i. Note that the predictor in eq. (8) is unbiased, as can be easily verified by taking expectations on both sides. The corresponding MSE can be computed as:

$$MSE^k_{\tau+M} = E\left[\left(\widetilde{Z}^k_{\nu,\tau+M} - Z^k_{\nu,\tau+M}\right)^2\right] \tag{9}$$

which, upon substitution of eqs. (3) and (8) into (9), after some algebra becomes:

$$MSE^k_{\tau+M} = E\left[\left(\frac{\sum_{i=0}^{M-1}\left(X_{\nu,\tau+M-i} - \mu_{\tau+M-i}\right)}{\sigma^k_{\tau+M}}\right)^2\right] \tag{10}$$

By assuming monthly precipitation serially uncorrelated, the above expression simplifies as:

$$MSE^k_{\tau+M} = \left(\frac{\sigma^M_{\tau+M}}{\sigma^k_{\tau+M}}\right)^2 \tag{11}$$

where $\sigma^M_{\tau+M}$ is the standard deviation of precipitation aggregated in the M months preceding the month $\tau+M$.

Besides MSE, a practical way of quantifying the accuracy of the forecast is by estimating the confidence interval of prediction, i.e. an interval that contains the future observed value with a fixed probability α. Obviously, the wider the interval, the less is the accuracy of prediction and vice-versa. Confidence intervals of prediction for SPI can be estimated by capitalizing on the intrinsic normality of the index and by observing that, since the predictor is unbiased, its variance coincides with the MSE. Thus, the upper and lower confidence limits $Z_{1,2}$ of fixed probability α, can be computed as:

$$Z_{1,2} = \hat{Z} \pm RMSE \cdot u\left(\frac{1-\alpha}{2}\right) \tag{12}$$

where for brevity $\hat{Z}$ represents the generic forecast, *RMSE* is the square root of the corresponding MSE and $u(.)$ is the normal standard quantile.

3. APPLICATIONS

The proposed methodology has been applied to monthly precipitation series observed in 43 stations in Sicily (Italy) during the period 1921-2003. In particular, several aggregation time scales have been considered, k=6, 9, 12 and 24 months, as well as different forecasting time horizons M=3, 6, 9 and 12 months.

First, theoretical *MSE* values (see eq. (11)) have been computed for all the 43 stations. Fig. 2 illustrates the boxplots of monthly MSE values for four combinations of the time scales k and time horizons M. The overall height of each boxplot indicates the variability of MSE as different stations are considered in each month. Except for the case of k=24 and M=12 months, the effect of seasonality appears evident, although with different patterns according to the scale of aggregation k and of the time horizon M. As expected, the performance of the forecasting model get worse (higher MSE's) as the ratio between time horizon M and aggregation time scale k increases.

Then, the forecasting model has been validated by comparing observed and forecasted SPI during a period not used for parameter estimation. Such validation is usually carried out by splitting the available sample into two sub-samples to be used for parameter estimation and model validation respectively (Klemes, 1986). Here a slightly different approach is proposed, where the generic SPI value at a given time interval is compared with the corresponding forecast, computed by estimating the parameters on the previous N years. Thus, a moving window of N years is considered for parameters estimation every year. This is consistent with the fact that when the model has to be applied for real time forecast, its parameters are usually computed on the basis of the last decades of observation. In order to better select the size of the moving window, an investigation has been carried in order to determine how the correlation coefficient r and the empirical root mean square error (RMSE) between forecasts and observations are affected by the window size. In Fig. 3, the correlation coefficients and the RMSE between forecasts and observations are plotted versus the time window size, for a few combinations of the aggregation time scale k and the time horizon M. Note that, in order to reduce the effect of sampling variability, both measures of performance have been computed on the same period of fixed length (from 1971 to 2003) in all cases. It can be inferred from the figure that correlation coefficients and RMSE's present a relative maximum and minimum respectively, when a moving window of 20 years is considered. Hence, a moving window of 20 years has been applied for model validation.

As an example, Fig. 4 shows the comparison between observed and forecasted values SPI for one of the 43 stations, namely Caltagirone, for different combinations of the time scales k and M. On the same plots, 95% confidence intervals are also shown. Note that although the forecasts have been computed using eq. (8), strictly valid under the normality assumption for aggregated precipitation, observed SPI values have been computed by means of the equiprobability transformation originally proposed by McKee et al. (1993), based on gamma distributed precipitation. From the figure it can be inferred a fairly good agreement between observed and forecasted SPI values as is also evident from the fact that almost all of the observed values lie within the confidence intervals.

4. CONCLUSIONS

Drought monitoring and forecasting are essential tools for implementing appropriate mitigation measures in order to reduce negative impacts. The availability of forecasts of drought indices, and of the related confidence intervals for a given site or region, can help to improve the decision making process for drought mitigation, since appropriate measures can be selected based on the risk associated with the possible evolution of a current drought condition.

In the paper, a stochastic model for forecasting future SPI index values on the basis of past precipitation, is presented. Analytical expressions of the Mean Square Error of prediction are also presented, which enable to derive confidence intervals for the forecasted values. Validation of the model has been carried out by comparing SPI values computed on precipitation observed in 43 stations in Sicily and the corresponding forecasts, making use of a moving window scheme to estimate parameters of the model. The results show a fairly good agreement between observations and forecasts, which indicates the suitability of the model for medium-short term forecast of drought conditions.

Ongoing research is being carried out in order to improve the forecasting capabilities of the model, by including information related to large scale climatic indices.

Finally, it is worth underling that, although drought forecasting models are essential tools for drought management, nonetheless they provide only partial information which need to be properly integrated with other data (i.e., economical and environmental data), in order to help decision makers to successfully select appropriate mitigation measures.

ACKNOWLEDGEMENTS

The present research has been developed with the partial financial support of the EU Projects "Mediterranean drought preparedness and mitigation planning" (MEDROPLAN, EuropeAid MEDA, contract no. ME8/AIDCO/2001/0515/59770-P027) and Interreg III B Medocc SEDEMED II "Sécheresse et Désertification dans le bassin Méditerranée II" (contract no. 2003-03-4.4-I-010), and of the Italian project MIUR COFIN 2003 "Caratterizzazione delle siccità e gestione dei serbatoi" (contract no. 2003081738-002).

REFERENCES

Bonaccorso, B., Bordi, I., Cancelliere, A., Rossi, G., and Sutera, A. (2003). "Spatial variability of drought: an analysis of SPI in Sicily". *Water Resources Management*, Vol. 17, pp. 273-296.

Brockwell, P.J. and Davis, R.A., (1996). *Introduction to Time Series and Forecasting*, Springer-Verlag ed., New York.

Cancelliere, A., Rossi, G. and Ancarani, A., (1996). "Use of Palmer Index as Drought Indicator in Mediterranean Regions". *Proc. IAHR Congress "From Flood to Drought"*, Sun City, South Africa, August 5-7, 1996, pp. S4.12. 1-25.

Guttman, N.B., (1999). "Accepting the Standardized Precipitation Index: a calculating algorithm", *J. Amer. Water Resources Association*, Vol. 35(2), pp. 311-323.

Heim, R.R., (2000). "Drought indices: a review". In D.A. Wilhite (Ed.), *Drought: A Global Assessment*, Routledge, London and New York.

Karl, T., Quinlan, F. and Ezell, D.S., (1986). "Drought termination and amelioration: its climatological probability", *J. Climate Appl. Meteor.*, Vol. 26, pp. 1198-1209.

Klemes, V., (1986). "Operational Testing of Hydrological Simulation Models", *Hydrological Science Journal*, Vol. 31(1), pp. 13-24.

Lohani, V.K., Loganathan, G.V. and Mostaghimi, S., (1998). "Long-term analysis and short-term forecasting of dry spells by Palmer Drought Severity Index", *Nordic Hydrology*, Vol. 29(1), pp. 21-40.

McKee, T.B., Doesken, N. J. and Kleist, J., (1993). "The relationship of drought frequency and duration to time scales", *Proc. 8th Conference on Applied Climatology*, Anaheim, California, pp. 179-184.

Rossi, G., (2003). "Requisites for a Drought Watch System". In: Rossi, G., Cancelliere A. et al. (eds), *Tools for Drought Mitigation in Mediterranean Regions*, Kluwer Academic Publishing, Dordrecht, pp. 147-157.

Rossi, G., Benedini, M., Tsakiris, G. and Giakoumakis, S., (1992). "On regional drought estimation and analysis", *Water Resources Management*, Vol. 6, pp. 249-277.

Rossi, G., Cancelliere, A., (2002). "Early warning of drought: development of a drought bullettin for Sicily", *Proc. 2nd International Conference "New trends in Water and Environmental Engineering for Safety and Life: Eco-compatible Solutions for Aquatic Environments"*, Capri, Italy, June 24-28, 2002, pp. 1-12.

Wilhite, D.A., Hayes, M. J. and Svoboda, M. D., (2000). "Drought monitoring and assessment: status and trends in the United States". In: J. V. Vogt, F. Somma (Eds.), *Drought and Drought Mitigation in Europe*, Kluwer Academic Publishers, pp. 149-160.

Yevjevich, V., Da Cunha, L. and Vlachos, E., (1983). *Coping with Droughts*, Water Resources Publications, Littleton, Colorado, 417 pp.

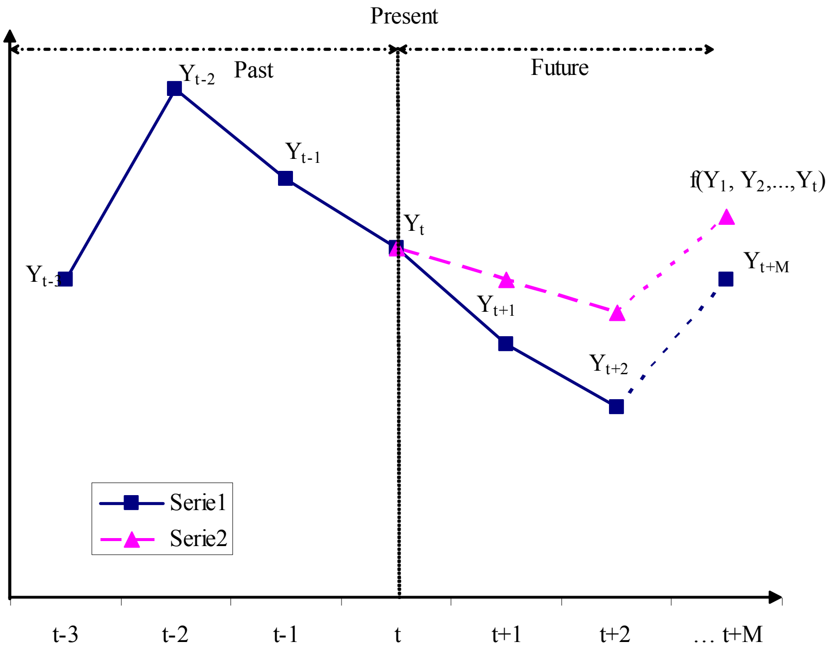

Fig. 1 Sketch of the forecast scheme

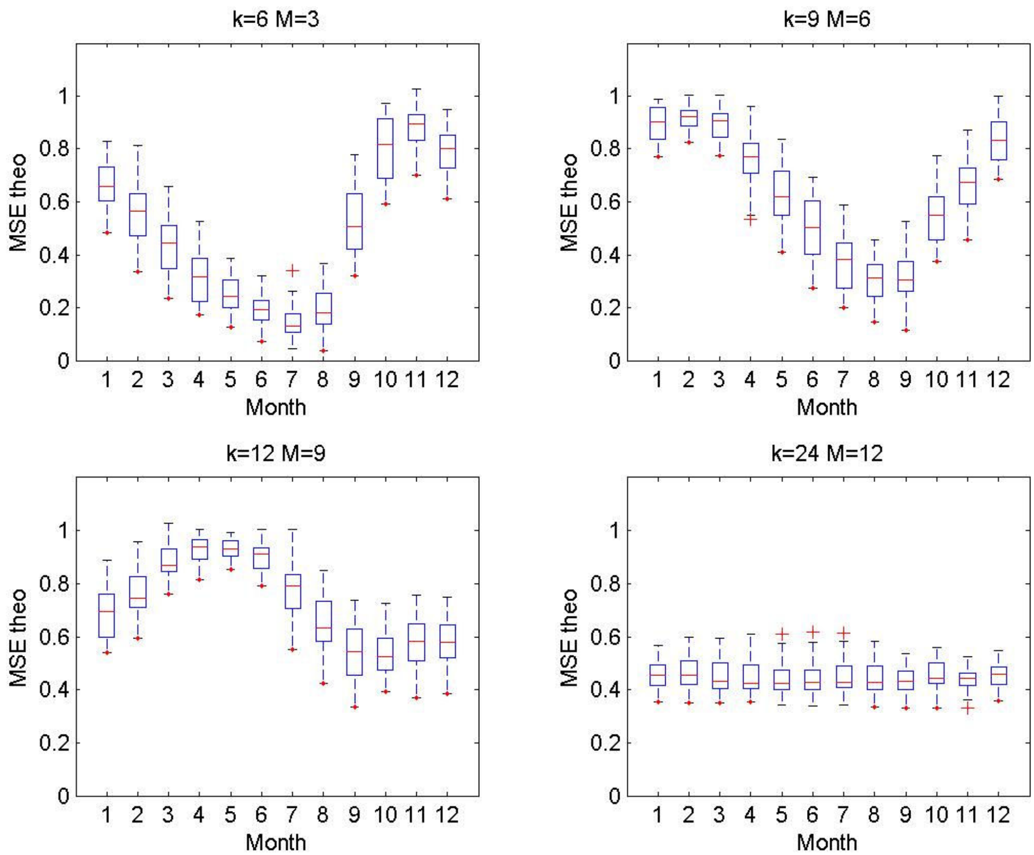

Fig. 2 Theoretical values of SPI's MSE under the hypothesis of normally distributed aggregated precipitation for the considered 43 precipitation stations

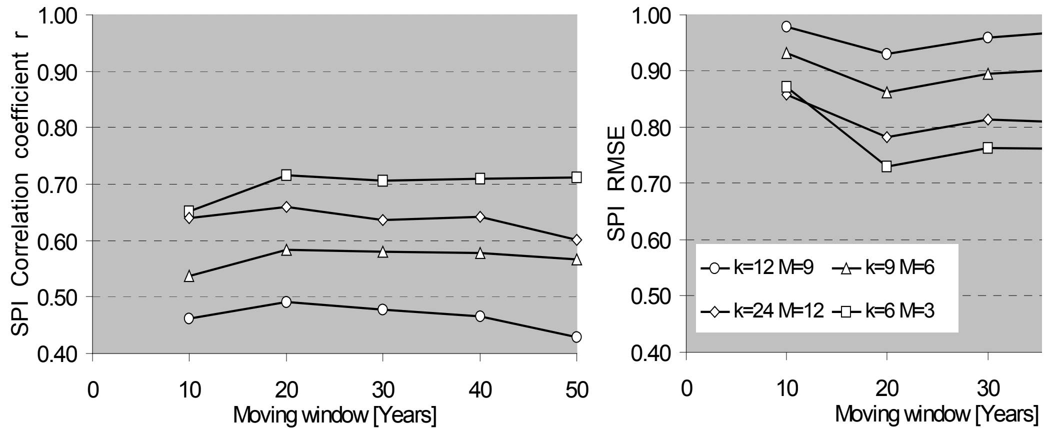

Fig. 3 Correlation coefficient and RMSE of observed and forecasted SPI values for different moving windows (Caltagirone station)

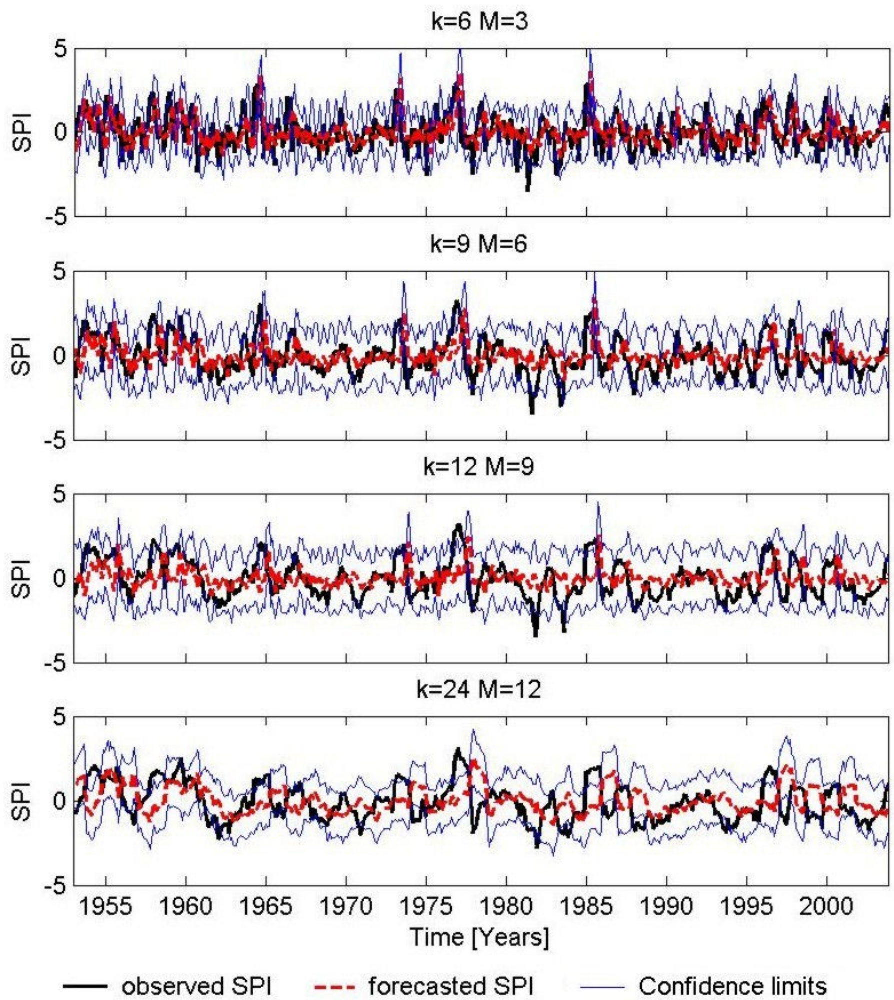

Fig. 4 Model validation: comparison between observed and forecasted SPI for Caltagirone station (moving window: 20 years)