Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(1) 2016 71-81

Forecasting Meteorological Drought for a Typical Drought Affected Area in India using Stochastic Models

N.M. Alam[1], Susheel Kumar Sarkar[2], C. Jana[1], A. Raizada[3], D. Mandal[1], R. Kaushal[1], N.K. Sharma[1], P.K. Mishra[1] and G.C. Sharma[1]

[1]*ICAR-Indian Institute of Soil and Water Conservation, Dehradun*

[2]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

[3]*ICAR-Indian Institute of Soil and Water Conservation, Research Centre, Bellary, Karnakata*

Received 10 February 2016; Revised 06 April 2016; Accepted 19 April 2016

SUMMARY

The Standardized Precipitation Index (SPI) is used throughout the world as a meteorological drought index to identify the duration and/or severity of drought. Early forecasting of drought is a critical issue to mitigate the adverse effects of drought of varying intensities. To address this issue, linear stochastic models, such as ARIMA and SRIMA have been used in this study. We studied ARIMA and SARIMA models to identify the most appropriate model to describe the SPI series at 3, 6, 9, 12 and 24 month time scale for the Ballary region in Southern India. Temporal characteristics of droughts based on SPI as an indicator of drought severity indicated that the region has been affected by a prolonged drought during the study period (1968–2012). Our study followed ARIMA calibration approach using time series data of SPI series for drought forecasting. The best model among different data sets has been identified using minimum Akaike Information Criteria (AIC), Schwarz-Bayesian Information Criteria (SBC) criteria along with the independency and normality criteria of the residuals. For 3-month SPI series ARIMA was observed to be appropriate while SARIMA model series is promising for the remaining SPI series. The stochastic models developed to predict drought were observed to give reasonably good results with 3 month lead time. Since drought prediction plays an important role in conservation of water resources, water storage management and mitigating drought severity, stochastic models has been observed to be the best and is recommended for drought forecasting in this region of India.

Keywords: Auto regressive integrated moving average, Drought forecasting, Linear stochastic model, Seasonal auto regressive integrated moving average, Southern India, Standardized precipitation index.

1. INTRODUCTION

Rising global temperatures are predicted to lead to an intensification of the hydrological cycle, resulting in dryer dry seasons and wetter rainy seasons, and subsequently heightened risks of more extreme, longer and frequent floods and droughts (IPCC, 2008, Jana *et al*. 2015). India has experienced changes in climate variability and extremes of weather and climate events in the recent years. Drought affects the natural environment of an area when it persists for a longer period. Research has shown that the lack of a precise and objective definition in specific situations has been an obstacle in understanding drought which has led to indecision and inaction on the part of land managers and policy-makers (Alam *et al*. 2012, Alam *et al*. 2014, Alam *et al*. 2015a, Alam *et al*. 2015b, Wilhite *et al*. 1985, Wilhite *et al*. 1986, Mishra *et al*. 2011). Comprehensive planning for developing optimal

Corresponding author: N.M. Alam
E-mail address: alam.nurnabi@gmail.com

strategies to deal with drought situations is becoming an increasingly important subject of concern for researchers and water resource managers in order to protect the affected community from drought. Drought has a direct impact on water resources management, thus water resource decision makers must be prepared to anticipate such situations and accept the challenges and complications that are involved in dealing with drought related problems. Monitoring of drought implies the ability to assess the current conditions and predict future drought occurrence, which is the key for developing any water resource management plan during drought periods.

To identify drought, several indices has been developed earlier (Mishra and Singh 2010). Some of the widely used drought indices are the Palmer Drought Severity Index (PDSI), Reconnaissance Drought Index (RDI), Standardized Precipitation Index (SPI), Standardized Precipitation Evapo-transpiration Index (SPEI), Surface Water Supply Index (SWSI) etc. of which Standardized Precipitation Index (SPI) (McKee *et al.* 1993) has been used widely because of relatively higher prediction accuracy (Guttman 1998). Drought forecasting plays an important role in the mitigation of the impact of drought on water resources. While many methods and approaches for formulating forecasting models are available in literature, this paper exclusively deals with time series forecasting models, in particular, the seasonal and non-seasonal Auto Regressive Integrated Moving Average (ARIMA) (Box *et al*. 1994). They have been successfully applied in various water and environmental management applications and are the most widely used stochastic models for the purpose of drought forecasting (Alam *et al*. 2014a, Durdu 2010, Haan 2002, Mishra and Desai 2005, Mishra *et al*. 2007).

The impacts of drought in the low and variable rainfall region are widespread, affecting diverse sectors as agriculture, irrigation, and energy and can be classified as short term and long term. The consequence of drought in the short run adversely affects food grain production which can lead to drop in employment and income, and in the long run, it leads to distress sale of assets and out migration of affected households. Another short–term effect of drought is decline in food stock-it leads to increase in food grain prices, and thus, there is reduction in the intake of food, and in the long run, it affects the health of people and leads to starvation (Patnaik 2010).

India is one of the most vulnerable countries to climate change (FAO 2002) and is considered as one of the most drought-prone countries in the world (Shetty *et al*. 2013). Within the country, the rainfed areas, which constitute 55 per cent of the net sown area and support two-thirds of livestock and 40 percent of the human population of the country is assumed to be the most vulnerable to climate change. The options to take care of variability in rainfall (timing and duration) are also limited (Rao *et al*. 2011). Moreover, farmers in the these regions are acutely vulnerable to climate variability and change due to their limited natural and financial resources coupled with poor infrastructure, institutional support, and governance (World Bank 2008). Among different farm categories, expectedly smallholder farmers, mostly poor, are highly vulnerable, who are highly dependent on agriculture and on natural resources, for their livelihood (Conway 2008). It will therefore be prudent to develop a methodology that will predict the occurrence of a meteorological drought in the drought prone areas, so that suitable contingency measures are employed to reduce the impact.

The study was undertaken in the Ballary region of India which falls under the category of an arid region (aridity ratio 0.05-0.20). This region is characterized by an average annual rainfall < 500 mm, with 16% of net sown area (Alam *et al.* 204b, Patnaik 2010, Alam *et al.* 2015b) and was declared to be affected by metrological drought 4 times in every 10 years (Ministry of Agriculture/ Drought Management 2008).

Alam *et al.* (2014) and Durdu (2010) reported drought forecasting up to two months in advance using ARIMA model for the Bundelkhand region in central India and Buyuk Menderes river basin in semi-arid climatic condition of western Turkey, respectively. Our hypothesis is that simple linear stochastic models can be useful in predicting the occurrence of drought with a sufficient time advantage. No systematic study has been conducted on the application of stochastic models to forecast drought for the sem-arid tracts in India, which is historically known to be a drought prone region. The present investigation was undertaken with the objective to use linear stochastic model (ARIMA) as a potential tool for drought forecasting which will be useful for implementing appropriate drought mitigation strategies in areas susceptible to frequent droughts of moderate intensities.

2. MATERIALS AND METHODS

2.1 Dataset

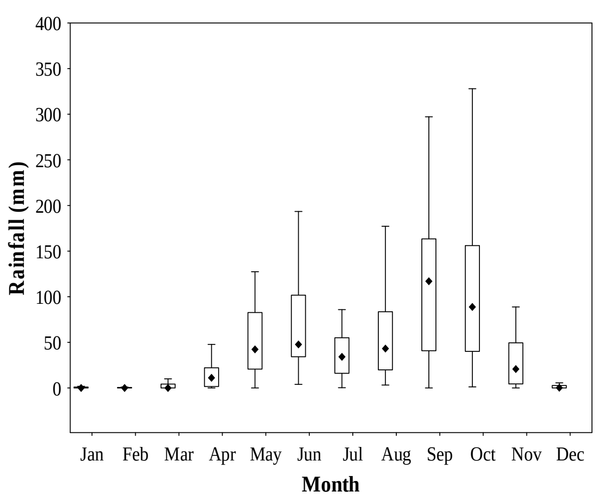


Fig. 1. Distribution of monthly rainfall for Bellary region of India, vertical bars represents upper and lower limit while box represents 75th and 25th percentile and dot represents mean value (Alam *et al*. 2015)

The daily rainfall (mm) data of the Ballary region was collected for 45 years and 3 months (Jan, 1968–March, 2013) from a Class A meteorological observatory located at the Indian Institute of Soil and Water Conservation, Research Centre, Bellary, Karnataka, India which is situated at an elevation of 580 AMSL ($15^0$15' N Lat. and $76^0$93$'$ E Long.). This region falls in the southern part of Karnataka, which is the 9th largest state in India, covering an area of 191976 sq.km, but has the 2nd largest arid zone after the state of Rajasthanin India. The distribution of monthly rainfall for the region has been shown in Fig. 1 (Alam *et al*. 2015b).

The Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (SPI) is a tool, which was primarily developed to identify meteorological drought and wet events by using only series of monthly rainfall (McKee *et al*. 1993). Mathematically, the standardized precipitation index is simply the difference of precipitation from the mean for a specified time period divided by the standard deviation, where the mean and standard deviation are determined by past records. The computation of SPI becomes complicated, when the SPI is normalized so as to reflect the variable behaviour of precipitation for time steps shorter than 12 months (Shahid and Hazarika 2010). To overcome this problem, the long-term precipitation records of a station are fitted to a gamma distribution, since the gamma distribution has been observed to adjust to the precipitation distribution quite well. This is done through a process of maximum likelihood estimation of the gamma distribution parameters, α and β. Then, the cumulative probability of an observed precipitation event for each time scale of interest is deduced. The cumulative distribution is transformed to a normal distribution with a mean of zero and standard deviation of one, since the probability distribution is determined by fitting an incomplete gamma distribution to the data, which is the value of the SPI.

In simple terms, SPI is a normalized index representing the probability of occurrence of an observed rainfall amount when compared with the rainfall climatology at a certain geographical location over a long-term reference period. Negative SPI values represent rainfall deficit, whereas positive SPI values indicate rainfall surplus. Intensity of drought event can be classified according to the magnitude of negative

SPI values such that larger the negative SPI values are, the more serious the event would be. For example, $-0.99 \leq \text{SPI} \leq 0$ is classified as mild drought, $-1.49 \leq \text{SPI} \leq -1.00$ and $-1.99 \leq \text{SPI} \leq -1.5$ are classified as moderate and severe drought respectively, whereas negative SPI values greater than or equal to 2 are classified as extremely dry conditions. SPI enables rainfall conditions to be quantified over different time scales (e.g. 3-, 6-, 9-, 12-, or 24-month rainfall), facilitating the analyses of drought impact on various water resource needs. A 3-month SPI reflects short- and medium-term moisture conditions and provides a seasonal estimation of precipitation. In agricultural regions, a 3-month SPI might be more applicable in highlighting available moisture conditions. A 6-month SPI indicates medium-term trends in precipitation and is considered to be more sensitive to conditions at this scale than the Palmer Index. A 9-month SPI provides an indication of precipitation patterns over a medium time scale. SPI values below -1.5 for these time scales are usually a good indicator which reflects significant adverse impacts in agriculture and is applicable in other sectors as well. A 12-month SPI reflects long-term precipitation patterns and is related to stream flows, reservoir levels, and even groundwater levels at the longer time scales. In some locations the 12 month SPI is most closely related with the Palmer Index, and the two indices reflect similar conditions. A 24-month SPI explains long-term drought in a given area.

2.2 ARIMA Model

One of the most widely used time series models is the ARIMA model. In early 1970's, Box and Jenkins pioneered in evolving methodologies for time series modelling in the univariate series often referred to as Univariate Box-Jenkins (UBJ) ARIMA modelling. The acronym ARIMA stands for "Auto-Regressive Integrated Moving Average". Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series (Ghafoor and Hanif 2005). Random-walk and random-trend models, autoregressive models and exponential smoothing models (i.e., exponential weighted moving averages) are all special cases of ARIMA models.

In general, a non-seasonal ARIMA model is characterized by the notation ARIMA (p, d, q), where 'p' is the number of autoregressive terms, 'd' is the number of non-seasonal differences and 'q' is the number of lagged forecast errors in the prediction equation. In ARIMA parlance, TS is a linear function of past actual values and random shocks. For instance, given a time series process (Y_t), a first order auto-regressive process is denoted by ARIMA (1,0,0) or simply AR (1) and is given by:

$$Y_t = \mu + \varphi_1 * Y_{t-1} + \varepsilon_t \quad (1)$$

where the auto regressive coefficient is denoted by φ.

A first order moving average process is denoted by ARIMA (0,0,1) or simply MA(1) is given by:

$$Y_t = \mu - \theta_1 * \varepsilon_{t-1} + \varepsilon_t \quad (2)$$

where θ, the coefficient of the lagged forecast error.

Alternatively, the model ultimately derived, may be a mixture of these processes and of higher orders as well. Thus a stationary ARIMA (p, d,q) process is defined by the equation:

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} \\ &\quad - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \end{aligned} \quad (3)$$

where ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for $t = 1, 2, \ldots, n$.

Seasonal ARIMA (SARIMA) model

The ARIMA process can be extended to include seasonal terms, giving a non-stationary seasonal ARIMA (SARIMA) process. Seasonal ARIMA models are powerful tools in the analysis of time series as they are capable of modelling a very wide range of series. Identification of relevant models and inclusion of suitable seasonal variables are necessary for seasonal analysis. SARIMA model is characterized by the notation SARIMA (p, d, q) (P, D, Q)$_s$ model, as reported in Shumway and Stoffer (2000), is defined by:

$$\underbrace{\left(1-\phi_1 B-\phi_1 B^2-\cdots.-\phi_p B^p\right)}_{AR(p)}$$

$$\underbrace{\left(1-\beta_1 S^s-\beta_2 S^{2s}-\cdots-\beta_P S^{Ps}\right)}_{AR_S(P)}$$

$$\underbrace{(1-B)^d}_{I(d)}\underbrace{(1-B^s)^D}_{I_S(D)} y_t \qquad (4)$$

$$= c + \underbrace{\left(1-\psi_1 B-\psi_1 B^2-\cdots-\psi_p B^p\right)}_{MA(q)}$$

$$\underbrace{\left(1-\theta_1 S^s-\theta S^{2s}-\cdots-\theta_P S^{Ps}\right)}_{MA_S(Q)}\varepsilon_t$$

where,

AR(p) Autoregressive part of order p,

MA(q) Moving average part of order q,

I (d) differencing of order d,

ARs (P) Seasonal Autoregressive part of order P,

MAs (Q) Seasonal Moving average part of order Q,

I_s (D) seasonal differencing of order D,

s is the period of the seasonal pattern appearing,

B is the backshift operator (i.e. B $y_t = y_{t-1}$, $B^2 y_t = y_{t-2}$ and so on),

The SPI series of different time scale was fitted using time series modelling approach which involves the following steps: model identification, parameter estimation, and diagnostic checking (Alam 2014a, Guttman 1998, Durdu 2010, Mishra and Desai 2005, Modarres 2007). The data set from 1968 to 2005 was used for model building for all the five SPI series.

2.3 Diagnostics of ARIMA and SARIMA Model

Different models can be obtained for various combinations of AR and MA individually and collectively (Khattree 2003). The best model is obtained with the following diagnostics:

2.3.1 Low Akaike Information Criteria (AIC)/ Bayesian Information Criteria (BIC)/ Schwarz-Bayesian Information Criteria (SBC)

AIC is given by AIC = (-2 log L + 2 m) where $m = p + q + P + Q$ and L is the likelihood function. Since -2 log L is approximately equal to {n (1+log 2π) + n log σ^2} where σ^2 is the model MSE, and AIC can be written as AIC= {n (1+log 2π) + n log σ^2 +2m} and because the first term in this equation is a constant, it is usually omitted while comparing between models. As an alternative to AIC, sometimes SBC is also used which is given by SBC = log σ^2 + (m log n) /n.

2.3.2 Plot of Residual's ACF

Once the appropriate ARIMA model has been fitted, one can examine the goodness of fit by means of plotting the ACF of the residuals of the fitted model. If most of the sample autocorrelation coefficients of the residuals are within the limits of ±1.96√N where, N is the number of observations upon which the model is based then the residuals are white noise indicating that the model is good.

2.3.3 Non-significant Autocorrelations of Residuals

To check the independence of the residuals i.e., to test the null hypothesis that a current set of autocorrelations is white noise, Ljung-Bix-Pierce statistic (Q) which is a function of autocorrelations of residuals is given by:

$$Q = n(n+2)\sum_{j=1}^{k} r^2(j)/(n-j) \qquad (5)$$

The Q statistic is compared to critical values from chi-square distribution. A significant value of Q indicates that the chosen model does not fit well.

The model having been identified and the parameters estimated, diagnostic checks are then applied to the fitted model to verify that the model is adequate. Several tests are employed for diagnostic check that consists of tests viz. Portmantateau Lack-of-fit Test and Anduson-Darling Goodness of Fit Test for Normality.

2.3.4 Portmantateau Lack-of-fit Test to Check the Independence of Residuals

Portmantateau lack-of-fit test is modified Ljung–Box–Pierce statistics proposed by Ljung and Box (Ljung and Box 1978) employed to check the independence of residuals. In order to test the null hypothesis that a current set of autocorrelations is white noise, test statistics are calculated for different total numbers of successive lagged autocorrelations using the Ljung–Box–Pierce corrected statistics ($Q^*(r)$ test) to test the adequacy of the model. The $Q^*(r)$ statistic is formulated as follows (Durdu 2010).

$$Q^* = n(n+2)\sum_{k=1}^{L} \frac{r_k^2(\varepsilon)}{n-k} \qquad (6)$$

where L is the total number of lagged autocorrelations under investigation, r_k is the sample, and autocorrelation of the residuals at lag k. $Q^*(r)$ values are compared to a critical test value of χ^2 distribution with respective degree of freedom at a 5% and significant level, n is total observation.

2.3.5 Anderson-Darling Goodness of Fit Test for Normality

In present study Anderson-Darling (AD) test is used for goodness of fit for testing the normality with 5% level of significance. The test statistics of AD test is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln \ln F(X_i) + \ln\{1 - F(X_{n-i+1})\}\right] \qquad (7)$$

where n is the number of observations and F(X) is the Cumulative Density Function (CDF) for the data. For a chosen significance level α, if A^2 is greater than the critical value D_{tab}, the null hypothesis related to normality is rejected for the chosen level of significance.

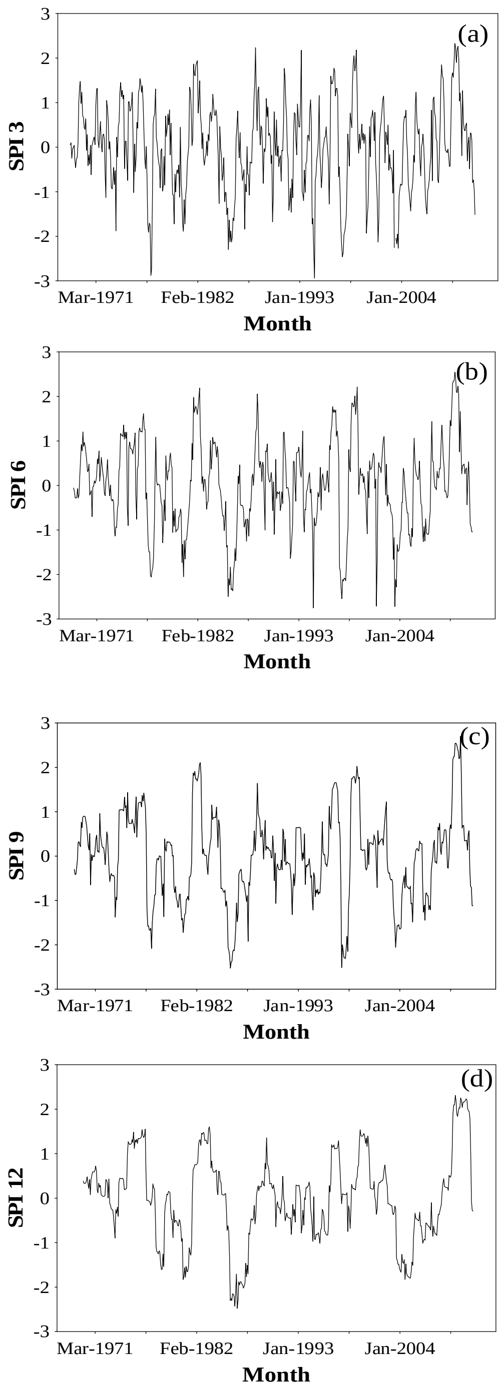
Also, basic statistical properties are compared between observed and forecasted data using Z-test for the means and F-test for standard deviation (Haan 1977). Since Z_{cal} values related to means were between Z critical table values (±1.96 for two tailed at a 5% significance level) and similarly, the Fcal values of standard deviation were smaller than the Fcritical values at a 5% significant level.

Detail description of ARIMA models can be referred in Box *et al.* (1994). A number of computer programs are available to compute predictions. In the present study ARIMA of SAS/ETS was used to estimate models using SAS 9.3 software.

3. RESULTS AND DISCUSSION

3.1 Drought Characteristics in the Region

The SPI has been calculated for 45 years at five different time scalesviz. 3-month, 6-month, 9-month, 12-month and 24-month to quantify both short as well as long term drought. The SPI series of the region for different timescales are shown in Fig. 2, from which it is clear that the region experiences both severe and extreme drought at all-time scales and period of drought increases with increase with the higher SPI time scale. From SPI at 3-month time scale it is clear that region experienced short term drought in all the months except February and March (Table 1) where SPI-3 values were positive i.e. non-drought. Annual minimum SPI-3 shows that in Bellary region it was -3.34 during May 2003. The frequency analysis of occurrence of annual minimum SPI at higher time scale viz. SPI-6, SPI-9, SPI-12 and SPI-24 showed that the region experiences moderate and severe drought for all the months of the year (Table 1). From the time series of monthly SPI series, it is clear that the region experienced frequent droughts and several severe and extreme drought events were detected at multiple time scale during the period under study Alam *et al*. (2015b). A minimum SPI at 6-month time scale for the region was observed in May, 2003 (-2.63). An extreme drought with lowest 9-month time scale SPI (-2.75) was observed in September 1995 for the region. In the region, minimum 12-month and 24-month SPI were observed during October, 1985 and July, 1986 with a drought magnitude of -2.52 and -2.48, respectively. Similar results have been observed by the studies in the Buyuk Menderes river basin, western Turkey by Durdu (2010) and for Bundelkhand region in central India by Alam *et al.* (2012) and Alam *et al.* (2014).



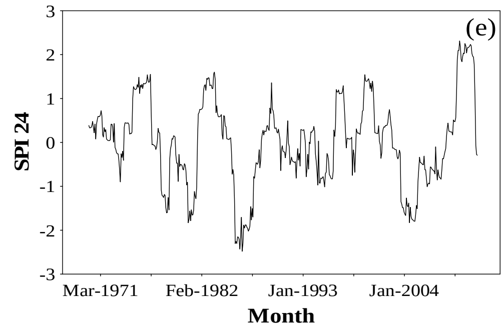


Fig. 2. SPI at different time scale (a, b, c, d, and e) based on monthly rainfall in bellary region

Table 1. Frequency percentage of occurrence of moderate and severe drought at different SPI

Month	SPI 3	SPI 6	SPI 9	SPI 12	SPI 24
Jan	10.70	7.20	8.50	6.60	8.30
Feb	0	7.20	7.30	6.60	8.30
Mar	0	8.40	9.80	7.90	8.30
Apr	9.30	9.60	7.30	7.90	8.30
May	10.70	6.00	8.50	10.50	8.30
Jun	10.70	10.80	9.80	10.50	9.70
Jul	13.30	9.60	7.30	9.20	8.30
Aug	12.00	9.60	8.50	7.90	6.90
Sep	8.00	9.60	8.50	9.20	8.30
Oct	10.70	8.40	8.50	7.90	8.30
Nov	5.30	6.00	7.30	7.90	8.30
Dec	9.30	7.20	8.50	7.90	8.30

3.2 Model Calibration

For model calibration SPI series for the period 1968- 2005 was taken for SPI-3, -6, -9 and -12 while 1968-2001 was taken for SPI-24 series. Stationarity of these SPI series have been checked using Dickey Fuller test. Dickey Fuller statistic and corresponding probability level for SPI-3, 6, 9, 12 and 24 series were -7.37 ($p < 0.01$), -7.63 ($p < 0.01$), -7.68 ($p < 0.01$), -5.66 ($p < 0.01$) and -3.84 (< 0.02), respectively. Since all the test statistic are significant at 5% level of significance, we can conclude that all the five SPI series are stationary and does not need differencing. One of the main reasons of being stationary is that SPI values are standardized values. After checking the stationarity of the series, identification of best model has been carried out. The identification of best model for the different SPI series based on minimum AIC and SBC criteria is demonstrated in Table 2, which indicates that other than SPI-3, all series performed well in SARIMA model. SPI–3 performs best under MA (2) model i.e. the stochastic model for SPI 3 is $y_t = \theta_1 \varepsilon_{t-1} +$

$\theta_2 \varepsilon_{t-2} + \varepsilon_t$, which indicates for SPI 3 series the stochastic model is weighted moving average over past errors. Lowest AIC and SBC value of 1012.06 and 1024.33 respectively, has been obtained among the other candidate models for MA (2) series of SPI 3. For all other SPI series, as the seasonal patterns are strong and consistent, seasonal ARIMA models has performed well. For SPI-6 series as the auto-correlation at the seasonal period is significantly positive at lag 1, we obtained best model as SARIMA (1, 0, 0)(1, 0, 0)6 with AIC and SBC values as 809.13 and 821.38 respectively. For SPI-9 and SPI-12 series the seasonal autocorrelation is significantly negative we obtained best fitted model as SARIMA (1, 0, 0)(0, 0, 1)9 and SARIMA (1, 0, 0)(0, 0, 1)12 respectively. SARIMA (1, 0, 0)(1, 0, 2)24 outperformed all the other candidate models for SPI 24 series with minimum AIC and SBC value of 30.45 and 10.24 for SPI 24 series (Table 2).

Table 2. Best selected ARIMA and SARIMA model based on minimum AIC and SBC criterion

SPI Series	Model	AIC	SBC	Adj R Square
SPI 3	MA (2)	1012.06	1024.33	0.61
SPI 6	SARIMA (1, 0, 0) (1, 0, 0)6	809.13	821.38	0.64
SPI 9	SARIMA (1, 0, 0) (0, 0, 1)9	597.66	609.89	0.77
SPI 12	SARIMA (1, 0, 0) (0, 0, 1)12	243.65	255.86	0.78
SPI 24	SARIMA (1, 0, 0) (1, 0, 2)24	30.45	10.24	0.94

The parameters estimates with associated standard errors, t-ratio and probabilities for the standard errors for the best fitted ARIMA and seasonal ARIMA model are listed in Table 3. For all the SPI series, stochastic model with no-intercept was found to be the best. Results also indicate that in comparison to the parameter values, the standard errors estimated for the model parameters are small. As most of the parameters are significant at 5% level of significance, associated parameters can be judged as significantly different from zero (Alam 2014, Durdu 2010, Mishra and Desai 2005). This indicates that the estimates of parameters are statistically significant and these parameters should be included in the models.

Table 3. Parameter estimates for best ARIMA and SARIMA models

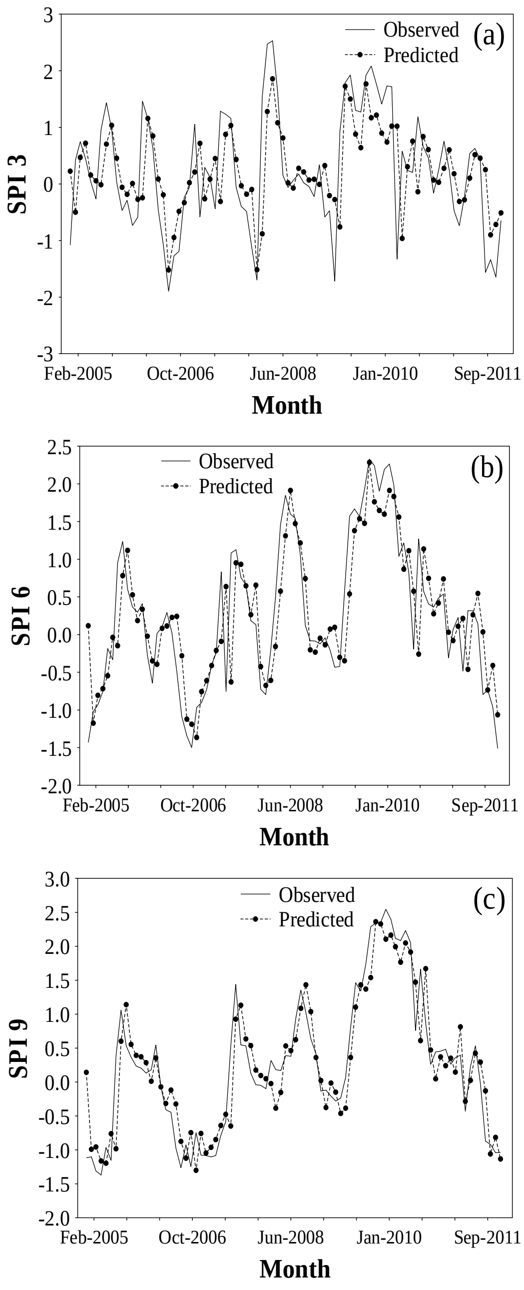
SPI Series	Model Parameters	Parameter Estimates	Standard Error	t-ratio	Prob>|t|
SPI 3	θ_1	-0.655	0.042	-15.54	< 0.001
	θ_2	-0.517	0.040	-12.93	< 0.001
SPI 6	φ_1	0.818	0.028	29.46	< 0.001
	Φ_1	-0.294	0.046	-6.37	< 0.001
SPI 9	φ_1	0.887	0.022	39.75	< 0.001
	Θ_1	0.458	0.044	10.31	< 0.001
SPI 12	φ_1	0.956	0.015	65.32	< 0.001
	Φ_1	0.761	0.037	20.40	< 0.001
SPI 24	φ_1	0.968	0.013	0.7409	< 0.001
	Φ_1	-0.893	0.118	-7.55	< 0.001
	Θ_1	-0.159	0.144	-1.10	< 0.027
	Θ_2	0.769	0.109	7.08	< 0.001

3.3 Model Validation

After selecting the best time series model, the model was validated using the SPI series for the period 2005 to 2012for the SPI3, SPI6, SPI 9 and SPI 12 while for the SPI 24 series data set from 2001 to 2012was used. Fig. 3 describes the close fit between the observed and predicted data using the selected best ARIMA/SARIMA models for the entire time series for the region. It was observed that the predicted data and observed data have similar characteristics in terms of the SPI series. The residuals of the validated series was tested to check whether they are independently and normally distributed. Portmantateau lack-of-fit test was done as a check of independence of the residuals. Testing normality of residuals by Anderson Darling (AD) goodness of fit test indicated that residuals for all the SPI series follow normal distribution as the test statistic was non-significant (Sharda and Das 2005). The test statistic along with probability level for Portmantateau lack-of-fit test and Anderson Darling test have been presented in Table 4 to describe validity (independent and normally distributed) of the residual of the models which is a statistical prerequisite for model validation. Portmantateau test shows that the calculated value is less than the actual χ^2 value, which signifies that the present models are adequate on the available data. Since all the SPI series test statistic AD test are non-significant (p<0.05) which indicates that residuals are normally distributed.

Table 4. Test for normality and independence of residuals

SPI Series	Test for Independence			Test for Normality	
	Portmantateau Test			Anderson Darling Test	
	***Q*(*r*)**	**df**	**Probability**	**A^2**	**Probability**
SPI 3	11.86	10	0.156	0.601	0.863
SPI 6	8.90	7	0.089	0.688	0.730
SPI 9	1.40	9	0.514	0.612	0.848
SPI 12	5.48	7	0.073	1.165	0.133
SPI 24	0.073	6	0.830	1.94	0.051



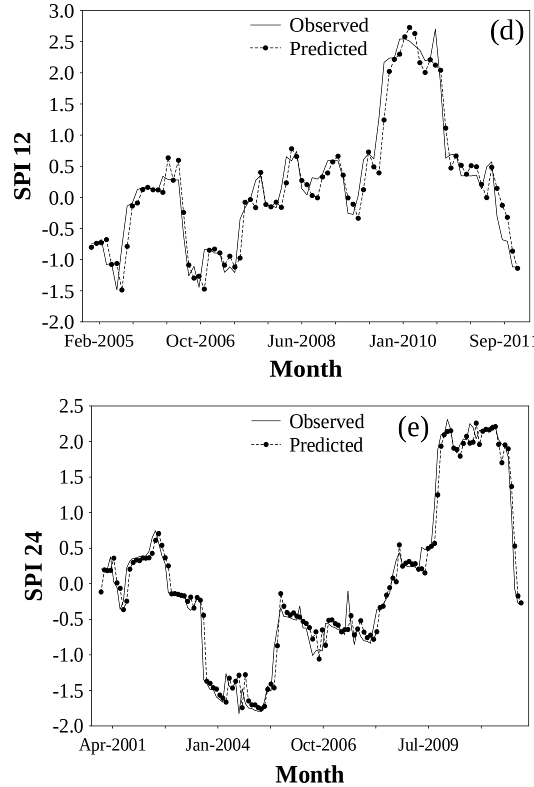


Fig. 3. Comparison of observed and predicted SPI data at different time scale (a, b, c, d and e) using the Best ARIMA models for bellary region, india

3.4 Drought Forecasting using ARIMA Model

We proceeded with the validation of model as per the flow chart following standard procedure. The one, two and three-step-ahead forecasting have been done using the best fitted ARIMA model for January, 2013 to March, 2013. The observed and forecasted series at different lead time is presented in Table 5. The basic statistical properties are compared between observed and forecasted data for 3 month lead time using ARIMA approach, based on t-test for the mean and F-test for the standard deviation (Table 6). Since t_{cal} values were found to be lower than t-critical table value for two tailed at a 5% level of significance level, it can be concluded that there is no significant difference between observed and forecasted values. Similarly the F_{cal} values of standard deviation were smaller than the F critical values at a 5% significance level. Thus, the results showed that forecasted data follow the basic statistical properties of the observed series. Results of Table 5 and 6 indicate

Table 5. Observed and forecasted data up to 3 month lead time using the best ARIMA/ SARIMA model

SPI Series	Lead Period	Observed SPI	Forecasted SPI
SPI 3	Jan-2013	-1.338	-0.899
	Feb-2013	1.647	0.716
	Mar-2013	0.646	0.509
SPI 6	Jan-2013	-0.736	-0.733
	Feb-2013	-0.960	-0.409
	Mar-2013	-1.512	-1.063
SPI 9	Jan-2013	-0.918	-1.059
	Feb-2013	-1.041	-0.815
	Mar-2013	-1.040	-1.133
SPI 12	Jan-2013	-0.705	-0.321
	Feb-2013	-1.117	-0.864
	Mar-2013	-1.119	-1.139
SPI 24	Jan-2013	-0.071	-0.032
	Feb-2013	-0.274	-0.171
	Mar-2013	-0.291	-0.267

Table 6. Statistics result for 1 to 3 month lead time of all SPI Series using ARIMA model

SPI Series	Mean Observed	Mean Forecasted	Probability	Variance Observed	Variance Forecasted	Probability	MPE
SPI 3	-0.556	-0.142	0.288	0.263	0.081	0.236	36.85
SPI 6	-0.288	-0.267	0.953	0.323	0.024	0.069	29.17
SPI 9	-0.163	-0.365	0.652	0.284	0.230	0.446	-0.86
SPI 12	0.123	0.093	0.469	0.0003	0.004	0.066	25.11
SPI 24	-0.212	-0.157	0.093	0.0003	0.001	0.052	33.59

MPE=Mean percentage error(%)

that ARIMA model can predict meteorological drought using SPI as drought indicator 3-month in advance with reasonable accuracy so it can be considered as the best analytical tools for drought forecasting. Similar results have also been reported by Alam *et al.* (2004a) and Durdu (2010).

4. CONCLUSION

This study indicates that the SPI is a valuable tool for quantifying meteorological drought at different scale and is able to detect different level of severities. Linear stochastic models (ARIMA) successfully demonstrated drought forecasting for Bellary region and proved that it is one of the worst affected drought prone regions of India. Temporal characteristics of the droughts indicated that the region experienced frequent moderate and severe droughts (i.e. SPI <-1) for almost all the months of the year. The stochastic models developed to predict drought were found to give reasonably good result up to 3 month in advance. Linear stochastic models can be used for the aridregions for predicting SPI time series of multiple time scale to detect the drought severity in future which is useful information for the local administration and water resource planners to take safety measures considering the severity of drought well in advance. The results of this study suggest that the linear stochastic models can be used for other hydro-meteorologically similar watersheds or regions for predicting SPI time series of multiple time scales to detect drought severity in future. The stochastic models validated for the region can be employed for the development of drought mitigation to ensure sustainable water resources management in the region and elsewhere. Model validation studies however also indicated that database from more number of sites is required to come into a meaningful interpretation.

ACKNOWLEDGEMENTS

The authors are grateful to Mr. B.N. Seshadri (T-4: Met. Observer) for collecting and providing data on long term climatic parameters. The authors declare that there is no conflict of interest related to this study.

REFERENCES

Alam, N.M., Adhikary, P.P., Jana C., Kaushal R., Sharma, N.K., Avasthe, R.K., Ranjan, R. and Mishra, P.K. (2012). Application of Markov model and standardized precipitation index for analysis of droughts in Bundelkhand region of India. *J. Tree Sci.*, **31(1-2),** 46-53.

Alam, N.M., Mishra, P.K., Jana, C. and Adhikary, P.P. (2014). Stochastic model for drought forecasting for Bundelkhand region in Central India. *Ind. J. Agric. Sci.*, **84(1)**, 79-84.

Alam, N.M., Raizada, A., Jana, C., Meshram, R.K. and Sharma, N.K. (2015b). Statistical modeling of extreme drought occurrence in Bellary district of Eastern Karnataka. *P. Natl. A. Sci. India*, **B85(2)**, 423-430.

Alam, N.M., Sharma, G.C., Jana C., Patra S., Sharma, N.K., Adhikary, P.P. and Mishra, P.K. (2015a). Probabilistic drought analysis of weekly rainfall data using markov chain model. *J. Rehab. Statist. Studies*., **8(1)**, 105-114.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis Forecasting and Control*. PernticeHall, Englewood Cliffs, NJ, USA.

Conway, G. (2008). The science of climate change in Africa: impacts and adaptation. Department of International Development, U.K.

Durdu, O.F. (2010). Application of linear stochastic models for drought forecasting on the Buyuk Menderes river basin, western Turkey. *Stoch. Env. Res. Risk.,* **A24,** 1145-1162.

FAO (Food and Agricultural Organisation) (2002). Report of FAO-CRIDA Expert Group Consultation on Farming System and Best Practices for Drought-prone Areas of Asia and the Pacific Region Food and Agricultural Organisation of United Nations Published by Central Research Institute for Dryland Agriculture, Hyderabad, India.

Ghafoor, A. and Hanif, S. (2005). Analysis of the trade pattern of pakistan: past trends and future prospects. *J. Agric. Soc. Sci.,* **1(4),** 346-349.

Guttman, N.B. (1998). Comparing the palmer drought index and the standardized precipitation index. *J. Amer. Water Resour*., **34(1)**, 113-121.

Haan, C.T. (2002) *Statistical Methods in Hydrology*. The Iowa State University Press, Ames, Iowa.

IPCC (2008). *Climate Change and Water*. IPCC Technical Paper VI., Cambridge Univ. Press, London.

Jana, C., Alam, N.M., Shrimali, S.S., Kumar, G., Ghosh, B.N. and Mishra, P.K. (2015). Rainfall extremity in Doon valley of Uttarakhand-reorienting agricultural management. *Inter. J. Agric. Statist. Sci*., **11(1)**.

Khattree, R. and Rao, C.R. (2003). *Statistics in Industry*. Handbook of Statistics, Elsevier, Amesterdam Boxton. **xxi,** 1187.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *J. Biom.*, **65,** 297–303.

McKee, T.B., Doesken, N.J. and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Preprints, Eighth Conference on Applied Climatology, Anaheim, California, 179-184.

Ministry of Agriculture/ Drought Management (2008). Drought management strategies in arid region. *Bulletin*, **22,** 38.

Mishra, A.K. and Desai, V.R. (2005). Drought forecasting using stochastic model. *Stoch. Env. Res. Risk.,* **19,** 326-339.

Mishra, A.K., Desai, V.R. and Singh, V.P. (2007) Drought forecasting using a hybrid stochastic and neural network model. *J. Hydrol. Engg.,* **12(6),** 626-638.

Mishra, A.K. and Singh, V.P. (2010). A review of drought concepts. *J. Hydrol.*, **391(1-2)**, 202-216.

Mishra, A.K. and Singh, V.P. (2011). Drought modeling-A review. *J. Hydro.,* **l403(1-2),** 157-175.

Modarres, R. (2007). Stream flow drought time series forecasting. *Stoch. Env. Res. Risk.,* **21,** 223-233.

Patnaik, I. (2010). Distress situation in drylands area: impacts on livelihood pattern and the coping strategies: A review. Centre for Economic and Social Studies. Begumpet, Hyderbad. *Bulletin*, **6,** 1-38.

Rao, V.U.M., Rao, A.V.M.S., Kumar, P.V., Desai, S., Saikia, U.S., Srivsatava, N.N. and Venkateswarlu, B. (2011). Agricultural Drought: Climate change and rainfed agriculture. Lecture notes of the 5th SERC school Central Research Institute for Dryland agriculture, Hyderabad, India, 324.

Shahid, S. and Hazarika, M.K. (2010). Groundwater drought in the north-western districts of bangladesh. *Water Resour. Manage.*, **24,** 1989-2006.

Sharda, V.N. and Das, P.K. (2005), Modelling weekly rainfall data for crop planning in a sub-humid climate of India. *Agric. Water Manage.,* **76,** 120-138.

Shetty, P.K., Ayyappan S. and Swaminathan, M.S. (2013). *Climate Change and Sustainable Food Security*. Published by the National Institute of Advanced Studies, Bangalore and Indian Council of Agricultural Research, New Delhi.

Shumway, R.H. and Stoffer, D.S. (2000), *Time Series Analysis and its Applications*. Springer texts in statistics. Springer, New York, **xiii,** 549.

Wilhite, D.A. and Glantz, M.H. (1985). Understanding the drought phenomenon: the role of definitions. *Water Int.,* **10,** 111-120.

Wilhite, D.A., Rosenberg, N.J. and Glantz, M.H. (1986). Improving federal response to drought. *J. Clim. Appl. Meteorol*., **25,** 332-342.

World Bank (2008). Agriculture for development. World Bank, Washington DC, USA.