Stochastic model for drought forecasting for Bundelkhand region in Central India

N M ALAM[1], P K MISHRA[2], C JANA[3] and PARTHA PRATIM ADHIKARY[4]

Central Soil and Water Conservation Research and Training Institute, Dehradun, Uttrakhand 248 195

ABSTRACT

In the present study, standardized precipitation index (SPI) series at 3-month, 6-month, 9-month, 12-month and 24-month time scale has been used to assess the vulnerability of meteorological drought in the Bundelkhand region of Central India. SPI values revealed that the droughts in the region over the study period vary from moderately high to extremely high. Suitable linear stochastic model, viz. seasonal and non seasonal autoregressive integrated moving average (ARIMA) developed to predict drought at different time scale. The best model was selected based on minimum Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). Statistical analysis revealed that non-seasonal ARIMA model was appropriate for 3-month SPI series while seasonal ARIMA models have been found promising for SPI series at 6-, 9,-12 and 24-month time scale. Parameter estimation step indicates that the estimated model parameters are significantly different from zero. The predicted data using the best ARIMA model were compared to the observed data for model validation purpose in which the predicted data show reasonably good agreement with the actual data. Hence the models were applied to forecast drought in the Bundelhand region up to 3 months advanced with good accuracy.

Key words : Auto regressive integrated moving average, Drought, Linear stochastic model, Seasonal Auto regressive integrated moving average, Standardized Precipitation Index

Because of widespread effect of drought in the USA in 1987-88 (Ahmed, 1991) and its devastating effect in the Somalia (1991-93), and Ethiopia (few years ago), the word "drought" has been drawn world-wide attention in recent years. Drought is considered by many researchers to be the most complex but least understood of all natural hazards, affecting more people than any other hazard (Sivakumar *et al*. 2005). It is not possible to avoid drought but drought preparedness can be developed and drought impacts can be managed. The success of both depends, amongst others, on how well the droughts are defined and drought characteristics quantified (Smakhtin and Hughes 2004). Drought can be defined in terms of both supply reduction and demand increase, and there are numerous definitions of hydrological, agricultural, ecological, and economic drought. Some of the widely used drought indices are the Deciles (Kininmonth *et al*. 2000), the SPI (McKee *et al*. 1993) and the RDI (Tsakiris and Vangelis 2005). All these indices have their own capability to assess drought under different situations.

The Standardized Precipitation Index, known as SPI,

seems to be the most popular among the existing simple indices for the estimation of drought because it is simple (low data requirements), spatially consistent in its interpretation, probabilistic so that it can be used in risk management and decision analysis, and can be tailored to time periods of user's interest (Edossa *et al*. 2010). The SPI is widely used for defining and monitoring meteorological droughts especially to know the behaviour of precipitation. It is also used worldwide due to its unique relation to probability, and normally distributed so it can be used to monitor wet as well as dry periods (Tsakiris *et al*. 2007). Patel *et al*. (2007) focused on investigation variability of seasonal drought events in Gujarat where it was concluded that SPI at a 3 month time scale was found effective in capturing seasonal drought patterns over space and time. Angelidis *et al*. (2012) computed SPI using lognormal and normal probability distribution.

For mitigating the impact of drought on water resources systems, forecasting of drought plays an important role. One of the basic deficiencies in mitigating the effects of drought is the inability to forecast drought well in advance either by few months or seasons (Mishra and Desai 2005). It is obvious, when drought occurs during climate season, it affects both crop seasons as well as increasing the food shortage and misery of the people (Ahmed 1995). Accurate drought forecasting plays an important role in planning and

[1]Scientist (email: alam.nurnabi@gmail.com), [2]Director (email: pkmbellary@rediffmail.com), [3]Scientist (email: chayna_4503@yahoo.co.in), [4]Scientist (email: ppadhikary@gmail.com), Central Soil and Water Conservation Research and Training Institute, Research Centre, Koraput, Odhisa 763 002

management of natural resources by adopting appropriate mitigation measures and policies for water resource management. Rao and Padmanabhan (1984) investigated the stochastic nature of yearly and monthly palmer's drought index (PDI) and to characterize those using valid stochastic models to forecast and to simulate PDI series. Kim and Valdes (2003) used PDSI as drought parameter to forecast drought in the Conchos River Basin in Mexico. Durdu (2010) used ARIMA model to predict drought in the Buyuk Menders river basin in western Turkey.

Very less attention has been focused on the application of stochastic model to forecast drought for Bundelkhand region in central India which is historically a drought prone area. Recognizing the above concerns and needs, the present study is envisaged to develop reliable and promising linear stochastic model of SPI for multiple time scale to forecast drought for Bundelkhand region in Central India. Linear stochastic model known as autoregressive integrated moving average (ARIMA) and multiplicative seasonal auto regressive integrated moving average (SARIMA) is taken into consideration to predict drought for Bundelkhand region using SPI as drought index.

MATERIALS AND METHODS

The daily rainfall (mm) data of Bundelkhand Region was collected for 44 years (1968–2011) from a USWB Class A meteorological observatory located at Central Soil and Water Conservation Research and Training Institute, Research Centre, Datia, Madhya Pradesh, India, which is situated at an elevation of 222 m with 25.7° N latitude and 78.43° E longitude. The observatory is located in Bundelkhand region Dounder the network of IMD, Pune, India. The area has a flat to rolling topography (300-450 msl) with hard rock formations, poor aquifer recharge and high run off potential.

McKee *et al*. (1993) developed the Standardized Precipitation Index (SPI) to quantify the precipitation deficit for multiple time scales, reflecting the impact of precipitation deficiency on the availability of various water supplies. The SPI provides a quick and handy approach to drought analysis. Computation of the SPI involves fitting a gamma probability density function to a given time series of precipitation, whose probability density function is defined by the expression:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \tag{1}$$

Where $\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale parameter, and $x > 0$ is the amount of precipitation. $\Gamma(\alpha)$ is the gamma function, which is defined as:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \tag{2}$$

Fitting the distribution to the data requires á and â to be estimated.

Integrating the probability density function with respect to x yields the following expression G(x) for the cumulative probability:

$$G(x) = \frac{1}{\Gamma(a)} \int_0^\infty t^{\alpha-1} e^{-1} dt \tag{3}$$

It is possible to have several zero values in a sample set. In order to account for zero value probability, since the gamma distribution is undefined for $x=0$, the cumulative probability function for gamma distribution is modified as:

$$\text{H}(x) = q + (1 - q)\, \text{G}(x) \tag{4}$$

where q is the probability of zero precipitation.

Finally, the cumulative probability distribution is transformed into the standard normal distribution to yield the SPI. Following the approximate conversion provided by Abramowitz and Stegun (1965), it results:

$$z = \text{SPI} = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), t = \sqrt{\ln\left(\frac{1}{(1.0 - H)(x)^2} \right)}$$

$$\text{for } 0 < \text{H}(x) < 0.5 \tag{5}$$

$$z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), t = \sqrt{\ln\left(\frac{1}{(1.0 - H)(x)^2} \right)}$$

$$\text{for } 0.5 < \text{H}(x) < 1.0 \tag{6}$$

and $c_0 = 2.515517$; $c_1 = 0.0802853$; $c_2 = 0.010328$; $d_1 = 1.432788$; $d_2 = 0.189269$; $d_3 = 0.001308$. Once standardized, the strength of the SPI, as given in the Table 1 (Lloyd-Hughes and Saunders 2002) can be visualized. Positive SPI values indicate greater than median precipitation and negative values indicate lower than median precipitation. Since the SPI is normalized, wetter and drier climates can be represented in the same way, and wet periods can also be monitored using the SPI.

A 3-month SPI reflects short- and medium-term moisture conditions and provides a seasonal moisture conditions as provides a seasonal estimation of precipitation. In primary agricultural regions, a 3-month SPI might be more applicable in highlighting available moisture conditions. 6-month SPI indicates medium-term trends in precipitation and is still considered to be more sensitive to conditions at this scale than the Palmer Index. 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

Table1   Drought classification by SPI value and corresponding event probability

SPI value	Category
SPI ≥ + 2.0	Extremely wet
1.50 ≤ SPI ≤ 1.99	Severely wet
1.00 ≤ SPI ≤ 1.49	Moderately wet
0.00 ≤ SPI ≤ 0.99	Mildly wet
–0.99 ≤ SPI ≤ 0	Mild-drought
–1.49 ≤ SPI ≤ –1.00	Moderate drought
–1.99 ≤ SPI ≤ –1.5	Severe drought
SPI ≤ –2.00	Extreme drought

indication that fairly significant impacts are occurring in agriculture and may be showing up in other sectors as well. 12-month SPI reflects long-term precipitation patterns. 12-month SPI are probably tied 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 should reflect similar conditions. The 24-month SPI captures the long-term drought.

Box & Jenkins (1976) developed this forecasting technique. The auto regressive integrated moving average ARIMA (p, d, q) model of the time series $\{r_1, r_2, ...\}$ is defined as

$$\varnothing(B) \Delta^d X_t = \theta (B)e_t$$

where X_t and e_t represent the time series and random error terms respectively at time t.

Often time series possess a seasonal component that repeats every s observations. Box et al. (1994) have generalized the ARIMA model to deal with seasonality, commonly known as SARIMA model

$$\varnothing(B) \Delta^d X_t = \theta (B)e_t$$

Where e_t is such that

$$\varnothing(B^3) \Delta_s^d \alpha_t = \varnothing (B^3)e_t$$

Hence

$$\varnothing(B)\Phi (B^3)\Delta_s^D\Delta^d X_t = \theta(B)\Phi(B^s) e_t$$

And we write $X_t \sim$ ARIMA (p, d, q) × (P, D, Q)$_s$. The idea is that SARIMA are ARIMA (p, d, q) models whose residuals et are ARIMA (P, D, Q). With ARIMA (P, D, Q) we intend ARIMA models whose operators are defined on B^s and successive power.

RESULTS AND DISCUSSION

SPI at five different time scale, viz. 3- month, 6- month, 9-month, 12-month and 24- month were used for short as well as long term drought quantification parameter. The SPI series for different timescales are shown in Fig. 1. From SPI at 3- month time scale it is clear that the region experienced moderate and severe drought (i.e. SPI <-1) during June to September (Table 2). Annual minimum SPI 3 show that the most extreme drought at 3-month time scale occurred in September 1979 with a magnitude of –3.22. During the last two decades maximum drought in terms of severity using SPI-3 (–2.20) was observed in July, 2002. The frequency analysis of occurrence annual minimum SPI at higher time scale, viz. 6 and above month showed that the region experienced moderate and severe drought for all the month of the year (Table 2). A minimum SPI at 6 month time scale was observed in September 1979 with an extreme drought (SPI = –3.25). An extreme drought with lowest 9 month time scale SPI (SPI = –3.04) was observed in February, 1980. The annual minimum 12-month SPI observed in May 1980 (SPI = –3.15), whereas the annual minimum 24-month SPI was reported in September 2007 (SP = –2.30).

Table 2　Frequency % for occurrence of moderate and severe drought at different SPI

Month	SPI 3	SPI 6	SPI 9	SPI 12	SPI 24
Jan	0.0	10.0	9.3	9.2	8.0
Feb	0.0	4.3	8.0	5.7	10.2
Mar	0.0	5.7	6.7	6.9	10.2
Apr	0.0	8.6	8.0	6.9	10.2
May	0.0	8.6	6.7	5.7	9.1
Jun	17.6	12.9	8.0	9.2	8.0
Jul	14.7	5.7	6.7	9.2	6.8
Aug	20.6	8.6	6.7	9.2	8.0
Sep	14.7	11.4	12.0	10.3	6.8
Oct	20.6	8.6	9.3	9.2	8.0
Nov	11.8	8.6	9.3	9.2	6.8
Dec	0.0	7.1	9.3	9.2	8.0

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 (Box and Jenkins 1974, Mishra and Desai 2005, Modarres 2007, Durdu 2010). The data set from 1968 to 2005 was used for model building for all the five SPI series.

The foremost step in identification was to check for the stationary of the time series, as the estimation procedures are available only for stationary series. Dickey Fuller test was used for checking the stationary of the series. At the identification stage, more than one models are tentatively chosen that seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model are obtained by least squares as advocated by Box and Jenkins. Out of the different models chosen, the best model was obtained with the following diagnostics:

Low Akaike Information Criteria (AIC)/ Schwarz-Bayesian Information Criteria (SBC) and non-significance of auto correlations of the residuals i.e. independence of residuals via Portmonateau test (Q-tests based on Chisquare statistics)-Box-Pierce or Ljung-Box tests.

A mathematical formulation for the AIC is developed as AIC = –2 log L + 2m where m= p + q+ P + Q and L is the likelihood function. Since -2 log L is approximately equal to n (1+ log 2ð) + n log ó^2 where ó^2 is the model MSE (Bozdogan 2000). As an alternative to AIC, sometimes SBC is also used which is given by

$$SBC = \log \sigma^2 + (m \log n)/n.$$

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:
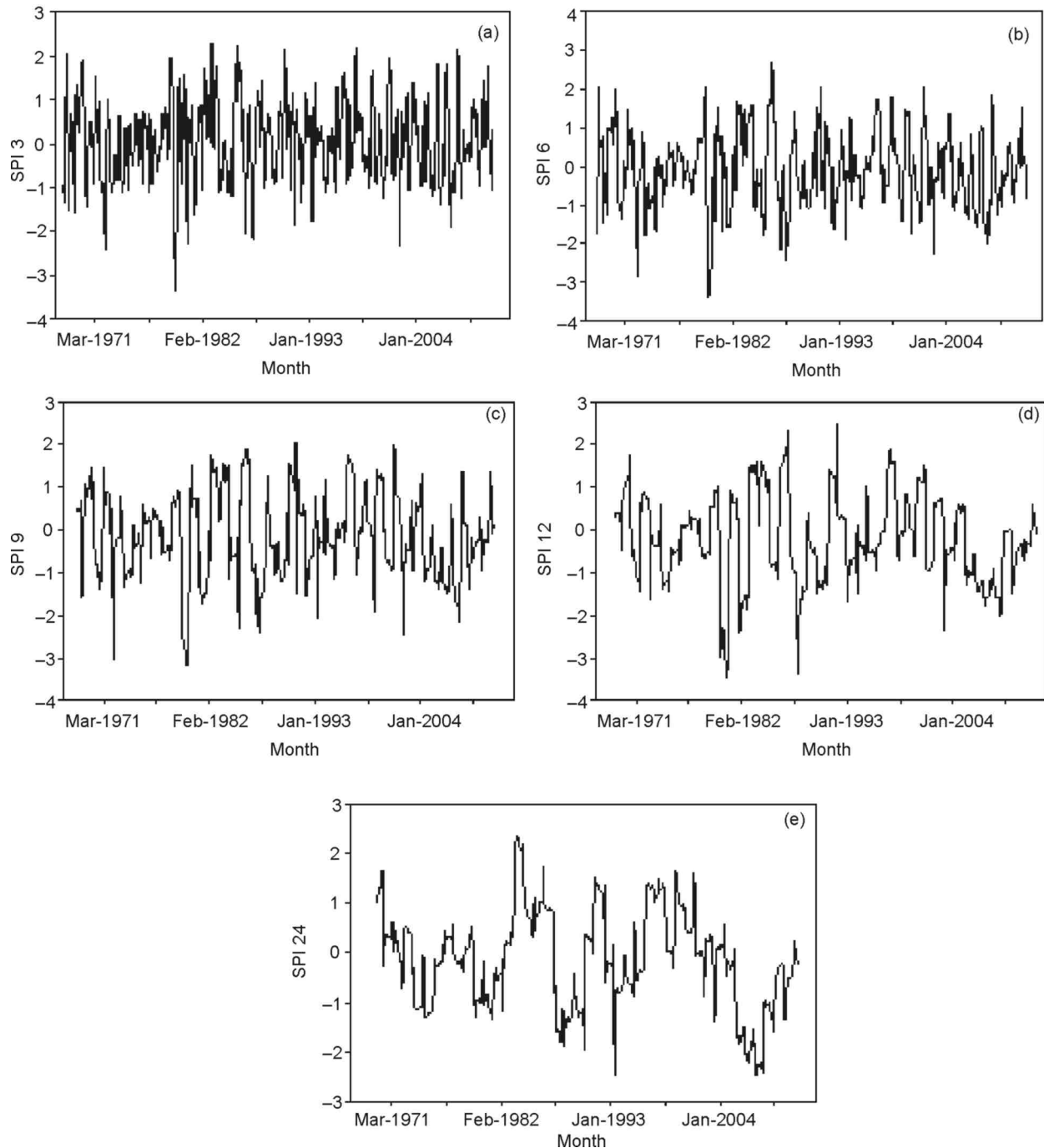
$$\sum_{i=1}^{k} r^2 (j)/(n-j)$$

Fig 1   SPI at different time scale (a, b, c, d, and e) based on monthly rainfall in Bundelkhand region

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.

Model calibration

Considering the Auto correlation function (ACF) and partial auto correlation function (PACF) graphs of the SPI series, different ARIMA/SARIMA models were identified and the best model was determined out of the candidate models. The model which gave the minimum AIC and SBC, was selected as the best model. The identification of best model for the different SPI series depended on minimum AIC and SBC criteria is demonstrated in Table 3. The table indicates that other than SPI-3 all series performed well in SARIMA model.

After selecting the best model, the parameters of the function was estimated. The value of the parameters,

Table 3   Best selected ARIMA/SARIMA models based on minimum AIC and SBC criteria

SPI Series	Model	AIC	SBC	R Square
SPI 3	MA(2)	975.19	987.55	0.42
SPI 6	SARIMA(1, 0, 0) (2, 0, 0)6	892.42	908.87	0.56
SPI 9	SARIMA(1, 0, 0) (0, 0, 1)9	731.88	744.19	0.70
SPI 12	SARIMA(1, 0, 0) (2, 1, 3)12	465.51	494.01	0.83
SPI 24	SARIMA(1, 0, 0) (0, 0, 1)24	175.20	187.42	0.89

associated standard errors, t-ratio and probabilities for the standard errors are listed in Table 4.

The Box-Jenkins methodology requires examining the residuals of the model to verify that the models are adequate. The residuals are examined to discover if the pattern remains unaccounted for. Ljung-Box-Pierce statistic is employed to check the independence of the residuals using the first 50 ACF of residuals from the model. The Q values are compared to a critical test value distribution with respective degrees of freedom at 5% level of significant level and in all the cases a non-significant value has been observed indicating the residuals from the best models are white noise.

Model validation

After selecting the best time series model, the model

Table 4   Parameter estimates for best ARIMA models

SPI Series	Model Parameters	Parameter estimates	Standard error	t-ratio	Prob>|t|
SPI 3	θ_1	−0.6800	0.0398	−17.10	<0.0001*
	θ_2	−0.5124	0.0422	−12.13	<0.0001*
	Intercept	0.1664	0.0722	2.31	0.0215*
SPI 6	φ_1	0.7495	0.0318	23.56	<0.0001*
	$\Phi1$	−0.3287	0.0478	−6.87	<0.0001*
	$\Phi2$	−0.1936	0.0476	−4.07	<0.0001*
	Intercept	0.0553	0.0801	0.69	0.4901
SPI 9	φ_1	0.8303	0.0266	31.21	<0.0001*
	Θ_1	0.4560	0.0453	10.06	<0.0001*
	Intercept	0.0615	0.0816	0.75	0.4512
SPI 12	φ_1	0.9428	0.0184	51.29	<0.0001*
	$\Phi1$	−1.0307	0.1439	−7.16	<0.0001*
	$\Phi2$	−0.2021	0.0657	−3.08	0.0022*
	$\Theta1$	0.8018	0.1533	5.23	<0.0001*
	$\Theta2$	0.7574	0.2516	3.01	0.0028*
	$\Theta3$	−0.5592	0.1301	−4.30	<0.0001*
	Intercept	0.0028	0.0087	0.32	0.7457
SPI 24	φ_1	0.9459	0.0158	59.78	<0.0001*
	Θ_1	0.7818	0.0353	22.15	<0.0001*
	Intercept	0.1610	0.0669	2.41	0.0165*

was validated using the SPI series for the period 2005 to 2001 for the SPI 3, SPI 6, SPI 9 and SPI 12 while for the SPI 24 series data set from 2001 to 2011 was used. It was observed that the predicted data ad 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. Run test was done for check of independence and Kolmogorov-Smirnov test was used to test the normality of residuals (Table 5).

Table 5   Test for normality and independence of residuals

SPI Series	Run test		Test for Normality	
	Wald-Wolfowitz Z statistic		Shapiro-Wilk Test	
	Statistic	Significant	Statistic	Significant
SPI 3	−1.313	0.189	0.601	0.863
SPI 6	−1.537	0.124	0.688	0.730
SPI 9	0.156	0.876	0.612	0.848
SPI 12	−1.939	0.053	1.165	0.133
SPI 24	−1.779	0.075	1.94	0.051

Drought forecasting using ARIMA model

1, 2 and 3 step ahead predictions of the model were computed using the best fitted ARIMA model. The observed and forecasted series at different lead time is presented in Table 6. 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 shown in Table 7. Since t_{cal} values found to be lower than t-critical table value for two tailed at a 5% level of significance level. Validation of the model revealed that no significant difference between observed and forecasted SPI series. Similarly the F_{cal} values of standard deviation were smaller than the F critical values at a 5%

Table 6   Observed and forecasted data up to 3 month lead time using the best ARIMA model

SPI Series	Lead period	Observed SPI	Forecasted SPI
SPI 3	1-month	−0.824	−0.387
	2-month	−0.880	−0.210
	3-month	0.035	0.171
SPI 6	1-month	−0.304	−0.409
	2-month	0.288	−0.101
	3-month	−0.848	−0.290
SPI 9	1-month	0.189	0.069
	2-month	0.097	−0.283
	3-month	−0.777	−0.879
SPI 12	1-month	0.142	0.057
	2-month	0.113	0.166
	3-month	0.114	0.057
SPI 24	1-month	−0.062	−0.099
	2-month	−0.095	−0.150
	3-month	−0.094	−0.124

Table 7   Statistics result for 1 to 3 month lead time of all SPI series using ARIMA model

SPI Series	Mean observed	Mean forecasted	Signi-ficant	Variance observed	Variance forecasted	Signi-ficant
SPI 3	–0.5562	–0.1419	0.288	0.2626	0.0812	0.236
SPI 6	–0.2879	–0.2667	0.953	0.3226	0.0241	0.069
SPI 9	–0.1634	–0.3645	0.652	0.2842	0.2295	0.446
SPI 12	0.1230	0.0931	0.469	0.0003	0.0039	0.066
SPI 24	–0.0838	–0.1241	0.0925	0.0003	0.0007	0.052

significance level. Thus, the result showed that forecasted data preserves the basic statistical properties of the observed series.

Employing these SPI based drought indices and the developed ARIMA models for drought forecasting played an important role in the lightening of drought damage for Bundelkhand region in Central India.

CONCLUSION

In this paper linear stochastic model (ARIMA) was successfully demonstrated on drought forecasting for Bundelkhand region in Central 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 Bundelkhand region for predicting SPI time series of multiple time scale to detect the drought severity in future which can also be useful for local administrations and water resource planners to take safety measures considering the severity of drought well in advance.

ACKNOWLEDGEMENT

REFERENCES

Abramowitz M and Stegun A. 1965. *Handbook of Mathematical Formulas, Graphs, and Mathematical Tables*. Dover Publications Inc, New York.

Ahmed R. 1991. The 1987-88 drought in selected north-central States of the USA. Gerographical Bulletin **33**, pp 30–36.

Ahmed R. 1995. Probabilistic duration of agricultural drought in Bangladesh during the pre-monsoon season. Six International Meeting on Statistical Climatology. University Collage, Galway, Ireland.

Angelidis P, Maris F, Kotsovinos N and Hrissanthou V. 2012. Computation of Drought Index SPI with alternative distribution functions. *Water Resources Management* **26**: 2 453–73.

Box G E P and Jenkins G M. 1974. Time series forecasting and control. Holden-Day, San Francisco.

Box G E P, Jenkins G M and Reinsel G C. 1994. Time series analysis forecasting and control. Prentice Hall, Englewood Cliffs, NJ, USA.

Durdu O F. 2010. Application of linear stochastic models for drought forecasting on the Buyuk Menderes river basin, western Turkey. *Stoch Environ Res Risk Assess* **24**: 1 145–1 162.

Edossa D C, Babel M S and Gupta A D. 2010. Drought Analysis on the Awash River Basin, Ethiopia. *Water Resources Management* **24**: 1 441–60.

Kim T and Valdes J B. 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural network.

Kinninmonth W R, Voice M E, Beard G S, de Hoedt, G C and Mullen C E. 2000. Australian climate services for drought management. (*In*) *Drought, a Global Assessment,* pp 210–22. Wilhite D A (Ed), Routledge, New York.

Lloyd-Hughes B andSaunders M. 2002. A drought climatology for Europe. *International Journal of Climatology*, **22**: 1 571–92.

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, pp 179–84.

Mishra A K and Desai V R. 2005. Drought forecasting using statistic model, *Stoch Environ Res Risk Asses* **19**: 326–39.

Modarres R. 2007. Streamflow drought time series forecasting. *Stock Environ Res Risk Assess* **21**: 223–33.

Patel N R, Chopra P and Dhdwal V K. 2007. Analyzing spatial patterns of meterological drought using Standardized Precipitation Index, *Meteorol Appl* **14**: 329–36.

Rao A R and Padmanabhan G. 1984. Analysis and modelling of plamar's drought index series. *Journal of Hydrology* **68**: 211–29.

Sivakumar M V K, Motha R P and Das H P. 2005. *Natural Disasters and Extreme Events in Agriculture*. Springer, Berlin, Heidelberg, New York.

Smakhtin V U and Hughes D A. 2004. Review, Automated estimation and analysis of drought Indices in South Asia. Working Paper 83, International Water Management Institute, Sri Lanka, p 24.

Tsakiris G and Vangelis H. 2005. Establishing a Drought Index incorporating evapotranspiration. *European Water* **9/10**: 3–11.

Tsakiris G, Pangalou D and Vangelis, H. 2007. Regional drought assessment based on the Reconnaissance Drought Index (RDI). *Water Resources Management* **21**(5): 821–33.