Technical Note on Probabilistic Assessment of one-step-ahead Rainfall Variation by Split Markov Process

Rajib Maity[*] and Dhanesh Prasad

Department of Civil Engineering, Indian Institute of Technology Kharagpur, W.B., India-721302

Abstract

In this paper, Split Markov Process (SMP) is developed to assess one-step-ahead variation of daily rainfall at a rain gauge station. SMP is an advancement of general Markov Process (MP) and specially developed for probabilistic assessment of change in daily rainfall magnitude. The approach is based on a first-order Markov chain to simulate daily rainfall variation at a point through state/sub-state Transitional Probability Matrix (TPM). The state/sub-state TPM is based on the historical transitions from a particular state to a particular sub-state, which is the basic difference between SMP and general MP. In MP, the transition from a particular state to another state is investigated. However, in SMP, the daily rainfall magnitude is categorized into different states and change in magnitude from one temporal step to another is categorized into different sub-states for the probabilistic assessment of rainfall variation. The cumulative state/sub-state TPM is represented in a contour plot at different probability levels. The developed cumulative state/sub-state TPM is used to assess the possible range of rainfall in next time step, in a probabilistic sense. Application of SMP is investigated for daily rainfall at Khandwa station in the Nimar district in Madhya Pradesh, India. Eighty years of daily monsoon rainfall is used to develop the state/sub-state TPM and twenty years data is used for to investigate its performance. It is

[*] Corresponding author: email: rajib@civil.iitkgp.ernet.in; rajibmaity@gmail.com

observed that the predicted range of daily rainfall capture the actual observed rainfall with few exceptions. Overall, the assessed range, particularly the upper limit, provides a quantification possible extreme value in the next time step, which is very useful information to tackle the extreme events, such flooding, water logging etc.

Keywords: Split Markov Process (SMP); Probabilistic Assessment; Rainfall Variation; Transitional Probability Matrix (TPM); Khandwa, India.

1 Introduction

Rainfall is one of the most complex and difficult component of the hydrologic cycle to model due to the complexity of the atmospheric processes and the wide range of variation in both space and time. However, prior information of rainfall is essential (both at large and small spatio-temporal scale) for proper planning and management of water resources. This is a high priority objective for developmental activities of a country, where the agricultural sector plays a key role for their economic growth. Large spatio-temporal variation of rainfall arises many water-related problems, such as, flood and drought, which seriously affect the crop production. Reasonably accurate rainfall prediction is required, which can help in alleviating such problems by planning for appropriate cropping patterns corresponding to water availability.

At smaller spatio-temporal scale, variation of rainfall has an effect on day-to-day life, such as, water logging, heavy traffic jams, shutdown of airports, blackout problem and so on. Heavy rain may paralyze most of daily activities. High intensity of rainfall at Mumbai on July 26, 2005 causes a complete halt for the city, large number of death (almost 1100) and an enormous loss of housing, trade and commerce, agriculture, cattle (as per the status report published by the government). An early information (at least a day before) could have helped in better management of the disaster. According to scientists at National Centre for

Medium Range Weather Forecasting (NCMRWF), which is a premier institute to provide medium range weather forecast in India, the predictions of severe weather events have enormous limitations (Bohra et al., 2006). Even though such events have a very short life but still cause extensive damage. Thus, even though the prediction of rainfall (spatio-temporal) is possible to achieve from numerical weather model, probabilistic information on of rainfall could be an added advantage for the concerned community. The main purpose is to provide as much advance notice as possible to the people to save the human and animal lives and properties from an impending disaster. The focus of this paper is the variation of point rainfall at a particular station.

Use of probabilistic rainfall prediction has a long history to predict the near-future occurrence of extreme events (Box et al., 1976; Weeks and Boughton, 1987; Wójcik et al, 2003). A framework for probabilistic rainfall forecast using nonparametric kernel density estimator is presented in a series of three papers (Sharma, 2000a; Sharma et al., 2000; Sharma, 2000b). The approach is developed for station rainfall data. However, the temporal resolution is seasonal to interannual. Application of Markov Process (MP) for short-term rainfall forecast through a probabilistic way is well accepted for a long time (Gabriel and Neumann, 1962; Chin, 1977; Fraedrich and Muller, 1983; Stern and Coe, 1984; Rajagopalan et al., 1996; Jimoh and Webster, 1996; Kaseke and Thompson, 1997; Wilks, 1999; Hayhoe, 2000; Kottegoda et al., 2004; Baik et al., 2006; Deni et al., 2009). For instance, Gabriel and Neumann (1962) found that the first-order Markov chain model could be fitted to daily rainfall data at the Tel Aviv in Israel. However, it was argued later that a second order model would fit the data more suitably (Gates and Tong, 1976). Fraedrich and Muller (1983) predicted the probability of weather state by first order of Markov chains by using data of single station and forecasted daily sunshine measurements and rainfall combined with three hourly past weather observations. Stern and Coe (1984) used a nonstationary Markov chain to model the occurrence of daily rainfall along with Gamma distribution to model the amount of rainfall. Fraedrich and Leslie (1987) used a linear combination of probabilistic approach (Markov chain) and numerical weather prediction (NWP) for short-term rainfall prediction. A first-order Markov process is a continuous-time process for which the future behavior, given the past and the present, only depends on the present and not on the past and characterized by set of states and the transition probabilities P_{ij} between the states. Here, P_{ij} is the probability that the state in the next time step is j, given that the same is i at the present time step. Haan et al. (1976) developed the stochastic model which was based on a first-order Markov process and used rainfall data to estimate the Markov transitional probabilities and simulated daily rainfall record of any length which was based on the estimated transitional probabilities and frequency distributions of rainfall amounts and concluded that simulated data had statistical properties similar to those of historical data. Kaseke and Thompson (1997) developed the partially observed Markov process algorithms for rainfall runoff process model and considered the special case of the martingale estimating function approach on the runoff model in the presence of rainfall. Rajagopalan et al. (1996) estimated the daily transition probability matrices nonparametrically and estimated the transition probabilities through a weighted average of transition by kernel estimator. Based on the assumption that the daily rainfall occurrence depends only on the previous days rainfall, first order Markov chain model was reported by Kottegoda et al. (2004) to fit the observed daily rainfall at Italy. However, application of higher order Markov chain model was established to be suitable for stochastic weather generator for daily rainfall characteristics (Wilks, 1999; Hayhoe, 2000). Optimum order of Markov chain model for a particular data was also addressed (Chin, 1977; Jimoh and Webster, 1996). Deni et al., (2009) applied an optimum order model for daily rainfall in Peninsular Malaysia using the Akaike’s (AIC) and Bayesian information criteria (BIC).

Almost all these approaches follow a general path of creating a single set of different states depending on historical record and the probabilities of transition from one state to another is obtained. However, for rainfall variation study, the change in rainfall magnitude, particularly in higher side, is more crucial information as indicated before. Quantifying these changes, through a single set of states, demands large number of defined states. The word 'large' is subjective and implies more number of required states for more the inherent variability. Generally, in the tropical countries, the variation of daily rainfall is very high and application of MP may not perform well. Moreover, probabilistic prediction is more useful than simple point prediction. Defining another set of sub-states, classifying the changes in magnitude of daily rainfall will be helpful for such probabilistic assessment. This is the theme of this paper. The objective of this study is to develop an approach for change prediction daily rainfall through state to sub-state transition, which is achieved through Split Markov Process (SMP). However, the approach considers daily rainfall in which sequential phases within a particular event of rainfall (e.g., initiation, growth, peak, decay and vanish) is not of interest. Rather the total rainfall in a day is considered, which is important from water resources point of view. Thus, the transitions through states to sub-states is computed through state/sub-state Transitional Probability Matrix (TPM) for a daily temporal resolution, which is used for probabilistic assessment of one-step-ahead rainfall variation. The methodology of Split Markov Process (SMP) is explained in next section. The proposed methodology is applied to a station rainfall data at Khandwa raingauge station in the Nimar district in Madhya Pradesh, India. Results and discussions are presented afterwards.

2 Methodology

2.1 General Markov Process

The Markov Process (MP) at discrete time points is characterized by a set of states and the transition probabilities P_{ij} from state i at time step t to state j at time step $t+1$ (Haan et al, 1976; Haan, 2002). The matrix representation of all possible P_{ij} forms the transition probability matrix (TPM) of the Markov chain, denoted as P. The definition of the P_{ij} implies that the sums of all elements in any row equal to 1 as the transitions from a particular state to all possible states are 'mutually exhaustive'.

The property of successive dependence in a time series is modeled through MP. The order of a MP is equal to the number of previous observation(s) on which the present value depends. For example, the conditional probability for m^{th} order Markov Process is expressed as $P\left[X_t = a_j / X_{t-1} = a_i,\ X_{t-2} = a_k, \cdots,\ X_{t-m} = a_l\right]$. Similarly, a first order Markov process is a stochastic process in which the state of the value X_t of the process at time t depends only on the state of X_{t-1} at time $t-1$ and no other previous values. Thus, the transition probability for the first order MP, P_{ij} , is expressed as

$$P_{ij} = P\left[X_t = a_j / X_{t-1} = a_i,\right] \tag{1}$$

The collection of all these probabilities with m different states forms the transition probability matrix (TPM), which provides information of transition from one state to another state, and thus can be synonymously termed as state-to-state TPM or state/state TPM as against state/sub-state TPM in case of SMP

2.2 Split Markov Process (SMP)

Major steps of SMP are shown in a flowchart in Fig. 1. It is a data driven process as in case of a MP. Basic assumption is the first order stationarity of the data. However, homogeneity of the data across different stations is not a necessary requirement if SMP is being applied to a specific station. In order to investigate the daily rainfall variation in a probabilistic way, another sub-state is introduced in addition to the existing states. Thus, the states categorize the daily rainfall amount and the sub-states categorize the daily rainfall variation. The observed rainfall data is classified in different categories depending on its variability and these categories are denoted as different states, say, S_1, S_2, $\cdots$, S_n, n being the total number of states. The amount of variation in daily rainfall magnitude is obtained by first order differencing of original data. These variations in daily rainfall magnitude are classified into different categories depending on the range of their variability. These categories are denoted as sub-states, say, $\bar{s}_1$, $\bar{s}_2$, $\cdots$, $\bar{s}_m$, m being the total number of states. The probability of transitions from a particular state to a particular sub-state is obtained from historical data and denoted as state/sub-state transition probability. The general m^{th} order state/sub-state transition probability is expressed as

$$P^m_{S,\bar{s}(j)} = P\left[r_n = \bar{s}_j / R_{n-1} = S_i,\ R_{n-2} = S_k, \cdots,\ R_{n-m} = S_l\right] \quad (2)$$

where R denotes the daily rainfall magnitude and r denotes the change in daily rainfall magnitude. A first-order state/sub-state transition implies that the change in magnitude for the next time step depends on the state of the system at the current time. Thus, a first-order state/sub-state transition probability is expressed as

$$P^1_{S(i),\bar{s}(j)} = P\left[r_n = \bar{s}_j / R_{n-1} = S_i\right] \quad (3)$$

The first-order state/sub-state TPM is expressed as (omitting the superscript for clarity)

$$P_{S,\bar{s}} = P\begin{bmatrix} P_{S(1),\bar{s}(1)} & P_{S(1),\bar{s}(2)} & \cdots & P_{S(1),\bar{s}(m)} \\ P_{S(2),\bar{s}(1)} & P_{S(2),\bar{s}(2)} & \cdots & P_{S(2),\bar{s}(m)} \\ \vdots & \vdots & \vdots & \vdots \\ P_{S(n),\bar{s}(1)} & P_{S(n),\bar{s}(2)} & \cdots & P_{S(n),\bar{s}(m)} \end{bmatrix} \tag{4}$$

State/Sub-state transition probability matrix is computed by selecting a particular state and counting the number of transition from that state to a particular sub-state. If a particular state, say $S(j)$, is observed for a total n times and m is the number of transition from state $S(j)$ to a particular sub-state $\bar{s}(j)$, then the $(i, j)th$ component of the state/sub-state TPM will be

$$P_{S(i),\bar{s}(j)} = \frac{m}{n} \tag{5}$$

The total number of times a particular state is observed and its transition to different sub-states is obtained from sufficiently long record of daily rainfall series.

Once the state/sub-state TPM is obtained, the cumulative state/sub-state TPM is obtained by row wise summation of column-by-column probabilities. A contour plot of this cumulative state/sub-state TPM will represent the nature of possible variation (probabilistically) in the forthcoming step from all possible states at the current time step. Thus, this contour plot can be used for probabilistic prediction of possible range of daily rainfall in the next step. For instance, from a particular state (current step), the possible variation of magnitude of expected change in next day rainfall (at some probability level) is computed using cumulative state/sub-state TPM. For graphical interpretation, one has to start from that particular state to that probability contour (desired probability level) and magnitude of expected change can be computed using a suitable interpolation technique. The minimum and maximum possible changes (with sign) are added to the rainfall magnitude of the current step to obtain the possible range of rainfall in the next time step. If the minimum possible change turned out to be very high negative value, it might be possible to get the lower limit of predicted rainfall range as negative value. However, the lower bound of the predicted range of possible rainfall should be bounded by zero.

2.3 Numerical example: Calculation of the transitional probability matrix for SMP

Let us consider that there are 100 data points in a series of observed values. Each observed values can be categorized into different states, thus, there are 100 states. First odder differencing $(X_{t+1} - X_t)$ is the (next-step) change in rainfall magnitude for the time step t. These changes can also be categorized into different sub-states and thus, there are 99 sub-states. Finally, paired states and sub-states (one less, i.e., 99) are obtained. Let us further considered that there are 5 states (I, II, …, V) and 5 sub-states (a, b, …, e). It may, however, be noted that number of states and sub-states need not be same). Now, from the record, numbers of different states are as shown in 2nd column if table 1. Again, transition from one particular state to different sub-states is also obtained from the record and shown in 3rd to 7th column of table 1.

Now to compute the state/sub-state TPM, each row should be divided by row-wise total, e.g., 15 for first row, 45 for 2nd row, and so on. This ensures that total probability of transitions from one state to different sub-states is equal to unity. Thus, the state/sub-state TPM is as shown in table 2. Next, the cumulative state/sub-state TPM is obtained as row wise summation of probabilities up to that cell, i.e., cumulative probability of being transited to a particular sub-state or lower than that sub-state. Thus, the cumulative state/sub-state TPM is as shown in table 3.

2.4 Numerical example: Estimation of probabilistic range of daily rainfall using SMP

Computation of probabilistic range of predicted rainfall is computed from a particular row of the state/sub-state TPM. This row refers to the state at which the previous day rainfall belongs to. Let us consider a row as follows, which indicates that at previous time step the rainfall state was in category 3. Possible range of each sub-states (a through e) are also shown in parentheses.

States	**Sub-states**				
	a (<-100)	**b (-100 to -25)**	**c (-25 to 25)**	**d (25 to 100)**	**e (>100)**
State S	0.000	0.291	0.515	0.183	0.011

If we are interested to know the 95% limits of the next day rainfall, we should obtain the lower and upper limits of the predicted change. We should first get the cumulative probability distribution, which is as follows:

States	**Sub-states**				
	a (<-100)	**b (-100 to -25)**	**c (-25 to 25)**	**d (25 to 100)**	**e (>100)**
State S	0.000	0.291	0.806	0.989	1.000

The change in magnitude should be in between states c and d. Lower limits of c and d are -25 and 25 respectively, whereas upper limits of sub-states c and d are 25 and 100 respectively. Thus, to find the limits of changes following interpolations are done. Interpolated values are shown in bold face.

Interpolation for lower limit		Interpolation for upper limit	
a (<-100)	**b (-100 to -25)**	**a (<-100)**	**b (-100 to -25)**
0.291	-100	0.000	-100
0.806	-25	0.291	-25
0.950	**14.34**	0.806	25
0.989	25	0.950	**84.02**
1.000	100	0.989	100

Thus, the lower and upper limits of the projected change are 14.34 and 84.02 unit. These can be subtracted and added from actual observed value of present time step to obtain the limit, subject to the lower bound of the predicted range of possible rainfall should be bounded by zero, as mentioned before. Thus, if the today's observed rainfall is 10 unit, then tomorrow's lower and upper limits of the rainfall will be 0 and 94.02 unit.

3 Application of SMP

The methodology is applied to the daily rainfall at four raingauge stations – Khandwa, Jabalpur, Sambalpur and Puri. Khandwa raingauge station is located in the Nimar district in Madhya Pradesh, India. Similar to the major part of Madhya Pradesh, Khandwa is having more or less plain topography. Average altitude of Khandwa is 316 m above mean see level. Puri is a costal station. It is located on the sea coast of Bay of Bengal and having an almost flat terrain. It is just few meter above the mean sea level. Sambalpur is having an undulating topography with approximate altitude 188 m above the mean sea level. It is about 300 km away from the coastal line. Jabalpur is located on the banks of the perennial Narmada River and approximate altitude is 393 m above mean see level. The entire area is low rocky and barren hillocks with slopes differing in grade from 2 to 30 per cent. Jabalpur and Khandwa are far away from the coast and located in the interior part of Indian land.

The daily rainfall data is collected for the period 1901 to 1999 from Indian Meteorological Department (IMD), Pune. The data set is complete and there is no missing data. The data is for the monsoon period (June to September) only as most of the annual rainfall (above 80%) occurs in this period only. Basic statistics for the rainfall data at all these stations are shown in Table 4. It is found that the station Sambalpur is having maximum mean rainfall whereas the kurtosis (measure of peakedness) is maximum for Jabalpur. For Khandwa station, mean rainfall is lowest with the maximum coefficient of variation.

Data for the period 1901 to 1980 is used for development of state/sub-state TPM and the data for the period 1981 to 1999 is used to test the performance of SMP. Stationarity of the data set is checked and the results are shown in Table 5. The entire period of the data is divided into five parts and the mean daily rainfall is computed for each period. Mean is also computed for entire length of data (1901-1999). The p-value (in parentheses) is obtained for the null hypothesis that the mean is equal to the mean for entire period (1901-1999) for that station at 5% significance level. It is found that almost for all the cases the mean does not differ from the overall mean (except two cases). Thus, it can be safely assumed that the data is first order stationary. The methodology of SMP is applied to a specific station, thus the homogeneity of the data is not checked. On the other hand, being located over different parts of the country, the daily rainfall characteristics need not be homogeneous. However, it can be found later that the SMP is performs almost equally for all these stations.

3.1 *Result and Discussion*

The daily rainfall data (R) is divided into nine different states. The zero rainfall ($R = 0$) is categorized as State 1 and range of other eight states are selected suitably as follows (data in mm):

State 1 → $R = 0$

State 2 → $0 < R \leq 5$

State 3 → $5 < R \leq 10$

State 4 → $10 < R \leq 20$

State 5 → $20 < R \leq 30$

State 6 → $30 < R \leq 45$

State 7 → $40 < R \leq 65$

State 8 → $65 < R \leq 100$

State 9 → $R > 100$

The states are selected in such a way that approximately 70% data falls below state 2, 80% data is below states 3, 85% data below state 4, 90% data below state 5, 95% data below state 6, 97.5% data below state 7 and 99% data below state 8. Thus, it is ensured that higher the magnitude finer the division. However, it is also ensured that minimum 50 data should fall in any state.

The changes in magnitude of daily rainfall are computed by taking first order different of the original series. These magnitudes (r) are classified into another set of nine different sub-states. The categorization is as follows (values are in mm):

Sub-state a → $r \leq -100$

Sub-state b → $-100 < r \leq -50$

Sub-state c → $-50 < r \leq -25$

Sub-state d → $-25 < r \leq -5$

Sub-state e → $-5 < r \leq 5$

Sub-state f → $5 < r \leq 25$

Sub-state g → $25 < r \leq 50$

Sub-state h → $50 < r \leq 100$

Sub-state k → $r > 100$

State/sub-state TPM is computed by selecting one particular state and historical transitions from that state to a particular sub-state are obtained from the available data, as shown in eqn. (5) in the methodology. The state/sub-state TPM is shown in Table 6. Row wise summation of column-by-column probabilities in the state/sub-state TPM results in cumulative state/sub-state TPM. The cumulative state/sub-state TPM is represented in a contour plot (Fig. 2). In this plot, 5%, 50% and 95% probability contours are shown in particular.

Three points can be noticed from the contour plot of cumulative state/sub-state TPM. First, the low probability contour line are almost linear whereas the high contour lines are nonlinear. Second, the low probability contours indicate that a lower state can have a larger change in the next time step, particularly for the low probability contours. For example, if the initial state is 2, at 50% probability level, the change magnitude is somewhere in between sub-states d and e, whereas if the initial state is 4, the change magnitude is some where in between c and d. However, for high probability contours, change magnitude increases with the relatively higher initial states. This can be observed for states 1 though 4 at 95% probability level. The third point is that for all the probability lines, for high initial states, the probability contours are linearly decreasing. This indicates that an extreme event can be followed by reduction in its magnitude in the next step (at daily scale).

As stated before, the cumulative state/sub-state TPM can be used to probabilistically infer the possible change in rainfall magnitude in the next time step. Being in some particular state at the current time-step, computation of the magnitude of expected change in rainfall (at some probability level) in the next time step is carried out using cumulative state/sub-state TPM. Two different values (minimum and maximum possible changes) are computed from the identified state of change by interpolation considering lower and upper boundaries for each sub-states. Results using linear interpolation are presented in this paper. The minimum and maximum possible changes (with sign) are added to the rainfall magnitude of the current step to obtain the possible range of rainfall in the next time step. The prediction performance in investigated for the period 1981 to 1999. The prediction performance varies with the probability level for the next day rainfall. A plot between probability level Vs

Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) is prepared (Fig. 3). These measures are computed between observed and the average of upper and lower limits predicted range. It is found that both the MSE and RMSE remain constant up to 80% probability level. However, these values are actually decreasing up to 80% probability level. MAE is found to gradually increase with the increase in probability. However, considering all these measures, the best performance is obtained at 80% probability level in terms of MSE and RMSE. Thus, the prediction performance is obtained at 80% probability level for the period 1981 to 1999 and shown in fig. 4a for all the raingauge stations. However, for clarity, the prediction performance for the period 1998 to 1999 is shown in fig. 4b all the raingauge stations. Upper and lower limits of possible next day rainfall are shown in these plots along with the actual observed rainfall. It is found that most of the observed rainfall either lie within the predicted range or close to it. However, there are still few cases in which the predicted range fails to capture the observed values. In particular, the upper limit is very high compare to the observed one. This might be due to the non existence of such variation in the historical record. Even though this is an shortcoming of the prediction performance, the overall performance is very useful to the community as an early warning to tackle the extreme events, such flooding, water logging etc. It is also worthwhile to mention here that one of the most important shortcomings of the SMP is the fact that it needs a long historical record to properly capture the historical behaviour of daily rainfall variation through state/sub-state TPM, which is a general shortcoming for almost all data driven approaches.

4 Conclusions

Daily variation of rainfall is one of the highly complex but most important parameter to tackle various hydrologic problems. Split Markov Process (SMP) is introduced in this paper to assess the daily rainfall variation in a probabilistic way. This study attempts to statistically analyze and predict the probabilistic behavior of the station rainfall using SMP. SMP investigates the transition between states and sub-states, as against the general Markov Process (MP), which investigates the transition between different states of the system. In order to assess probabilistic range of variation, sub-states are introduced in addition to the states to obtain state/sub-state transition probability matrix (TPM) in SMP. The state/sub-state TPM is generated for daily rainfall data from different raingauge stations using SMP. The probabilistic behavior of change in daily rainfall magnitude is captured through state/sub-state cumulative TPM, which is finally used to predict the possible range of daily rainfall in the next time step.

Illustration of SMP in this paper deals with first order SMP. The concept can be extended to higher order as well. As explained in equation (2), in general, previous m states are to be considered for m^{th} order SMP to obtain corresponding TPM. For example, TPM for second order SMP should consider two previous states. As it is noticed in the analysis, first order SMP with nine states and nine sub-states constitute a $[9 \times 9]$ TPM, i.e., $[Number\ of\ states \times Number\ of\ sub-states]$. However, for 2^{nd} order SMP the size of TPM will be 81X9. Similarly, for 3^{rd} order SMP three previous states are to be considered and the size of TPM will be 729x9. Thus, the number of rows increases by $(Number\ of\ states)^{order}$.

Using SMP, predictions are provided with a possible range of upper and lower limit of rainfall magnitude. Four raingauge stations are selected including one coastal station (Puri), one station (Sambalpur) is few hundred kilometers interior from the sea coast and other two stations (Jabalpur and Khandwa) are located inland. Topography of each station differs from each other. However, the performance of SMP is found to be uniform for all the stations as revealed in the analysis. The results are very useful for the upper range of prediction. The early notice for the extreme events is possible to communicate to the concerned community. However, as in the other data driven methods, the major drawback of the SMP is that it need a reasonably long historical record to capture the behavior of daily rainfall variation.

Reference

Baik H. S., Jeong H. S., Abraham D. M.: Estimating transition probabilities in Markov chainbased deterioration models for management of wastewater systems. Journal of Water Resources Planning and Management 132(1): 15-24, 2006.

Box G., Jenkins G. M., Reinsel G.: Time Series Analysis: Forecasting and Control. Prentice-Hall: NJ, 1976.

Bohra A. K., Basu S., Rajagopal E. N., Iyengar G. R., Das Gupta M., Ashrit R., Athiyaman B.: Heavy rainfall episode over Mumbai on 26 July 2005: Assessment of NWP guidance. Current Science 90(9): 1188-1194, 2006.

Chin E. H.: Modeling daily precipitation occurrence process in Markov chain. Water Resour Res 13(6):949–956, 1977.

Deni S. M., Jemain, A. A., Ibrahim K.: Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia. Theor Appl Climatol 97:109–121, 1997, DOI 10.1007/s00704-008-0051-3.

Fraedrich K., Mullar K.: On Single Station Forecasting: Sunshine and Rainfall Markov Chains. Beitr. Phys. Atmosph. 56(1): 108-134, 1983.

Fraedrich K., Leslie L. M.: Combining predictive schemes in short-term forecasting. Monthly Weather Review 115: 1640 – 1644, 1987.

Gabriel K. R., Neumann J.: A Markov chain model for daily rainfall occurrence at Tel Aviv. Quart J Roy Meteorol Soc 88:90 – 95, 1962.

Gates F., Tong H.: On Markov chain modeling to some weather data. J Appl Meteorol 15:1145–1151, 1976.

Jimoh O.D., Webster P.: Optimum order of Markov chain for daily rainfall in Nigeria. J Hydrol 185:45–69, 1996.

Haan C. T.: Statistical Methods in Hydrology, 2nd Edition, Iowa State Press: Iowa, 2002.

Haan C. T., Allen D. M., Street J. O.: A Markov chain model for daily rainfall. Water Resources Research 12(3): 443-449, 1976.

Hayhoe H. N.: Improvements of stochastic weather data generators for diverse climates. Clim Res 14:75–87, 2000.

Kaseke T. N., Thompson M. E.: Estimation for rainfall-runoff modeled as a partially observed Markov Process. Stochastic Hydrology and Hydraulics 11(1): 1-16, 1997.

Kottegoda N.T., Natale L., Raiteri E.: Some considerations of periodicity and persistence in daily rainfalls. J Hydrol 296:23–37, 2004.

Rajagopalan B., Lall U., Tarbotan D. G.: Nonhomogeneous Markov Model for daily precipitation. Journal of Hydrologic Engineering 1(1): 33-40, 1996.

Sharma A.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 — A strategy for system predictor identification. Journal of Hydrology 239(1-4): 232-239. DOI:10.1016/S0022-1694(00)00346-2, 2000a.

Sharma A.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 – A nonparametric probabilistic forecast model. Journal of Hydrology 239(1-4): 249-258. DOI:10.1016/S0022-1694(00)00348-6, 2000b.

Sharma A., Luk K. C., Cordery I., Lall U.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 — Predictor identification of quarterly rainfall using ocean-atmosphere information. Journal of Hydrology 239(1-4): 240-248. DOI:10.1016/S0022-1694(00)00347-4, 2000.

Stern R. D., Coe R.: A model fitting analysis of daily rainfall data. Journal of Royal Statistical Society Series A 147(Part 1): 1- 34, 1984.

Weeks W. D., Boughton W. C.: Tests of ARMA model forms for rainfall-runoff modeling. Journal of Hydrology 91(1-2): 29- 47, 1987.

Wilks D. S.: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93(3):153–169, 1999.

Wójcik R., Torfs P., Warmerdam P.: Application of Parzen densities to probabilistic rainfall-runoff modeling. Journal of Hydrology and Hydromechanics 51(3): 175-186, 2003.

Figure Captions:

Fig 1: Flowchart showing major steps of Split Markov Process (SMP)

Fig.2. Contour plot of states/sub-state cumulative TPM showing 5%, 50% and 95% probability contours for different stations as shown in title

Fig. 3: Plot of probability level Vs Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)

Fig. 4a: Prediction performance for the period June 1, 1981 to September, 1999 for different stations as shown in title

Fig. 4b: Prediction performance for the period and June 1, 1998 to September, 1999 for different stations as shown in title

Table Caption:

Table 1: Number of occurrences of states and its transitions to different sub-states

Table 2: State/sub-state TPM for the example problem shown in Table 1

Table 3: Cumulative state/sub-state TPM for the example problem shown in Table 1

Table 4: Descriptive statistics of the rainfall data

Table 5: Test for stationarity in mean

Table 6: State/Sub-state Transition Probability Matrix using Split Markov Process

Table 1: Number of occurrences of states and its transitions to different sub-states (ref. section 2.3 for the example problem)

State	Number of occurrences (Total = 99)	Number of observed transitions to sub-state				
		a	b	c	d	e
I	15	5	6	3	0	1*
II	45	15	22[#]	5	3	0
III	18	2	7	6	1	2
IV	12	1	2	5	2	2
V	9	0	1	2	4	2

[*] This cell should be read as there is 1 occurrence of transition from state I to sub-state e

[#] This cell should be read as there are 22 occurrences of transition from state II to sub-state b and other cells should be read in a similar way

Table 2: State/sub-state TPM for the example problem shown in Table 1

State	Sub-states				
	a	b	c	d	e
I	0.333	0.400	0.200	0.000	0.067
II	0.333	0.489	0.111	0.067	0.000
III	0.111	0.389	0.333	0.056	0.111
IV	0.083	0.167	0.417	0.167	0.167
V	0.000	0.111	0.222	0.444	0.222

Table 3: Cumulative state/sub-state TPM for the example problem shown in Table 1

State	Sub-states				
	a	b	c	d	e
I	0.333	0.733	0.933	0.933	1.000
II	0.333	0.822	0.933	1.000	1.000
III	0.111	0.500	0.833	0.889	1.000
IV	0.083	0.250	0.667	0.833	1.000
V	0.000	0.111	0.333	0.778	1.000

Table 4: Descriptive statistics of the rainfall data

Station	Descriptive statistics for daily rainfall data				
	Mean	**Median**	**CV**	**Skewness**	**Kurtosis**
Khandwa	6.22	0	2.69	5.75	52.58
Jabalpur	9.99	1.00	2.21	6.52	111.45
Sambalpur	11.33	1.60	2.09	4.96	52.87
Puri	7.98	0.10	2.46	4.91	41.42

Table 5: Test for stationarity in mean. The p-value (in parentheses) is for the null hypothesis that the mean is equal to the mean for entire period (1901-1999) for that station. The bold face cells indicate that null hypothesis can not be rejected at 5% significance level.

Station	**Mean in mm (p-value)**					
	1901-1999	**1901-1920**	**1921-1940**	**1941-1960**	**1961-1980**	**1981-1999**
Khandwa	6.22	4.97 (0.001)	**6.06 (0.682)**	**7.03 (0.029)**	**6.80 (0.124)**	**6.21 (0.983)**
Jabalpur	9.99	**9.09 (0.060)**	11.30 (0.008)	**10.05 (0.909)**	**9.76 (0.635)**	**9.76 (0.653)**
Sambalpur	11.33	**11.47 (0.787)**	**11.99 (0.209)**	**11.48 (0.777)**	**10.58 (0.147)**	**11.14 (0.724)**
Puri	7.98	**7.62 (0.391)**	**7.88 (0.807)**	**7.88 (0.818)**	**7.79 (0.658)**	**8.78 (0.074)**

Table 6: State/Sub-state Transition Probability Matrix using Split Markov Process

States	**Sub-states**								
	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**k**
1	0.000	0.000	0.000	0.000	0.876	0.088	0.023	0.011	0.002
2	0.000	0.000	0.000	0.001	0.773	0.155	0.048	0.017	0.006
3	0.000	0.000	0.000	0.476	0.291	0.144	0.057	0.020	0.012
4	0.000	0.000	0.000	0.684	0.113	0.112	0.055	0.025	0.011
5	0.000	0.000	0.158	0.648	0.072	0.066	0.033	0.018	0.006
6	0.000	0.000	0.646	0.229	0.044	0.026	0.022	0.026	0.007
7	0.000	0.311	0.500	0.104	0.031	0.018	0.031	0.006	0.000
8	0.000	0.782	0.126	0.058	0.012	0.000	0.000	0.000	0.023
9	0.629	0.258	0.048	0.048	0.000	0.000	0.016	0.000	0.000

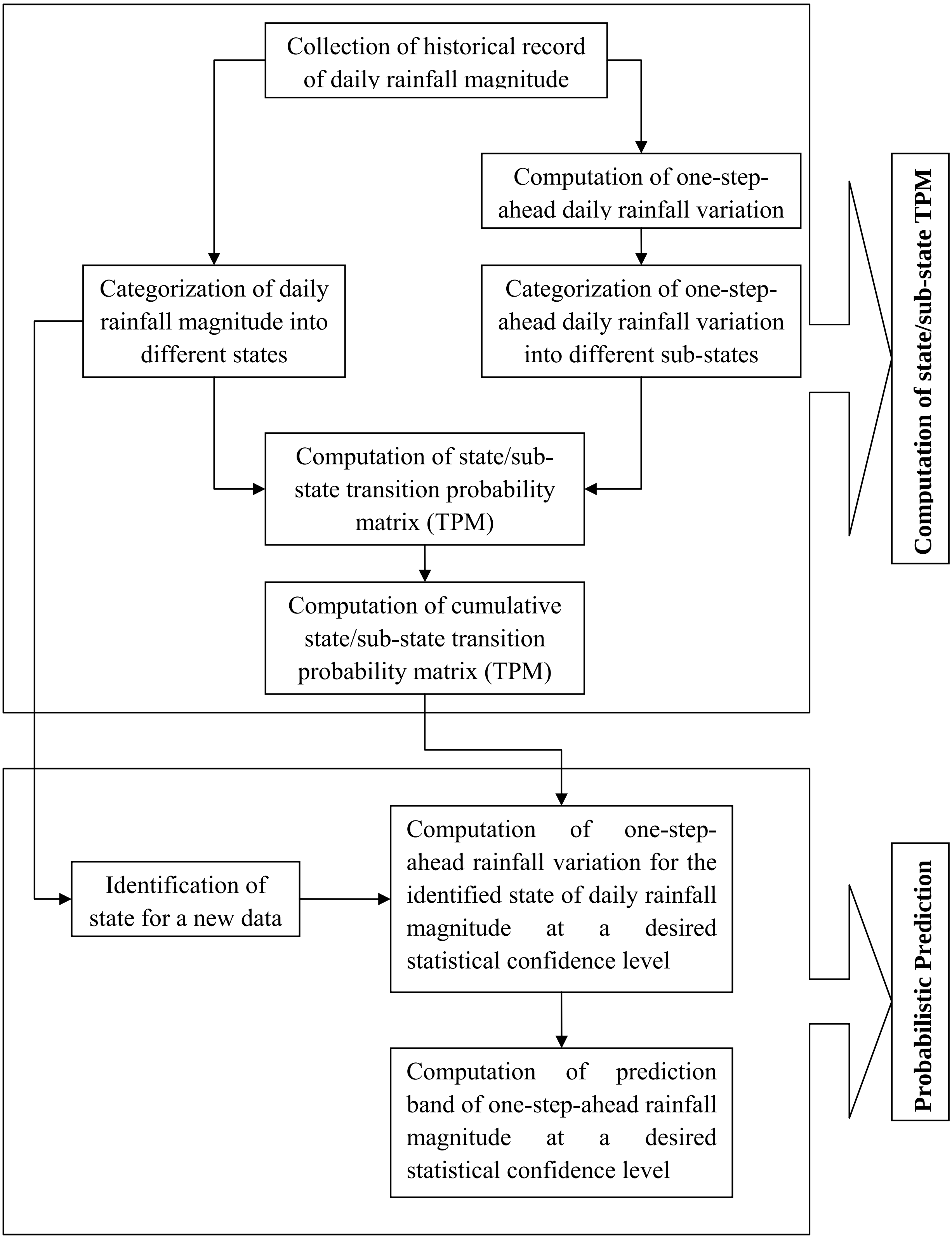

Fig 1: Flowchart showing major steps of Split Markov Process (SMP)

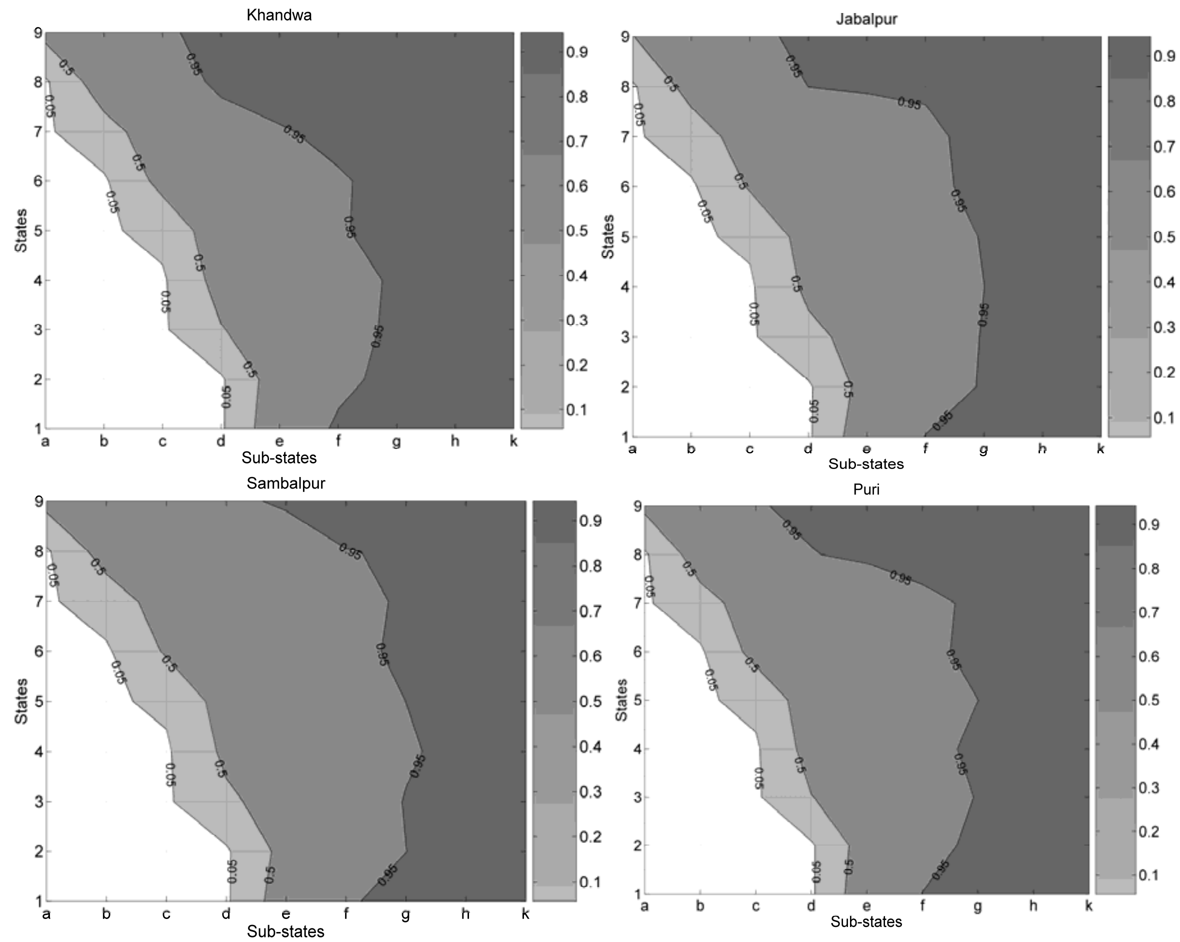

Fig.2. Contour plot of states/sub-state cumulative TPM showing 5%, 50% and 95% probability contours

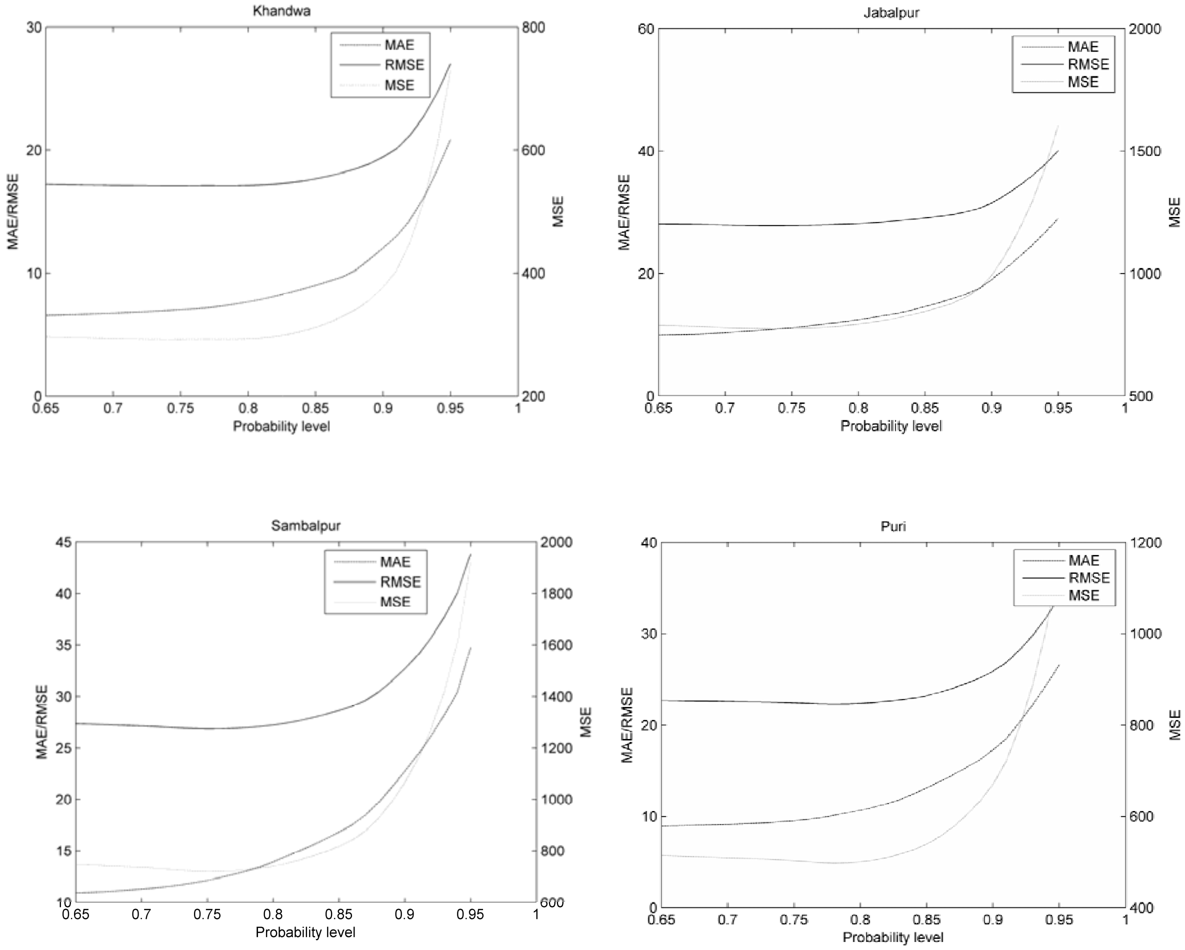

Fig. 3: Plot of probability level Vs Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)

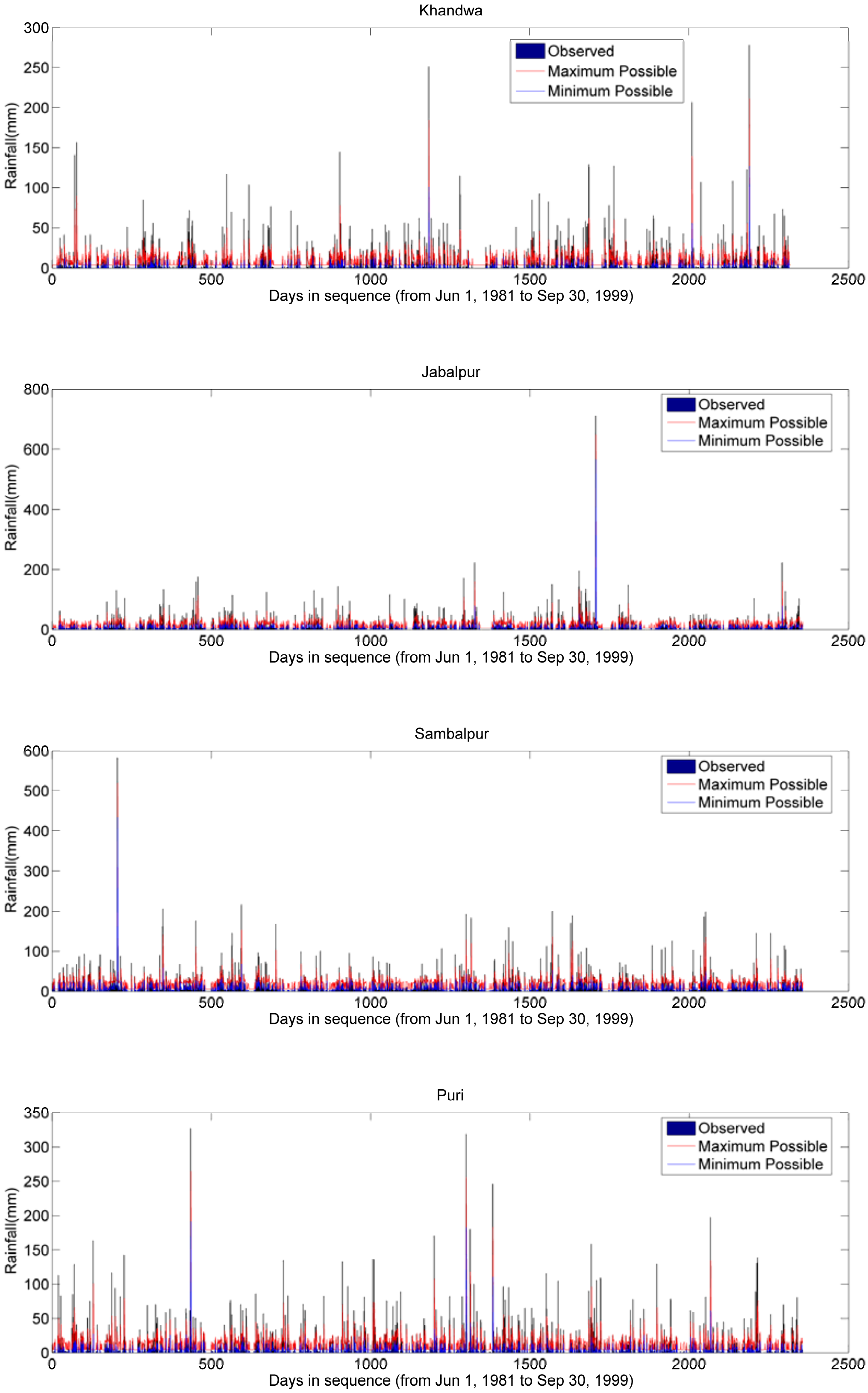

Fig. 4a: Prediction performance for the period June 1, 1981 to September, 1999 for different station as shown in title

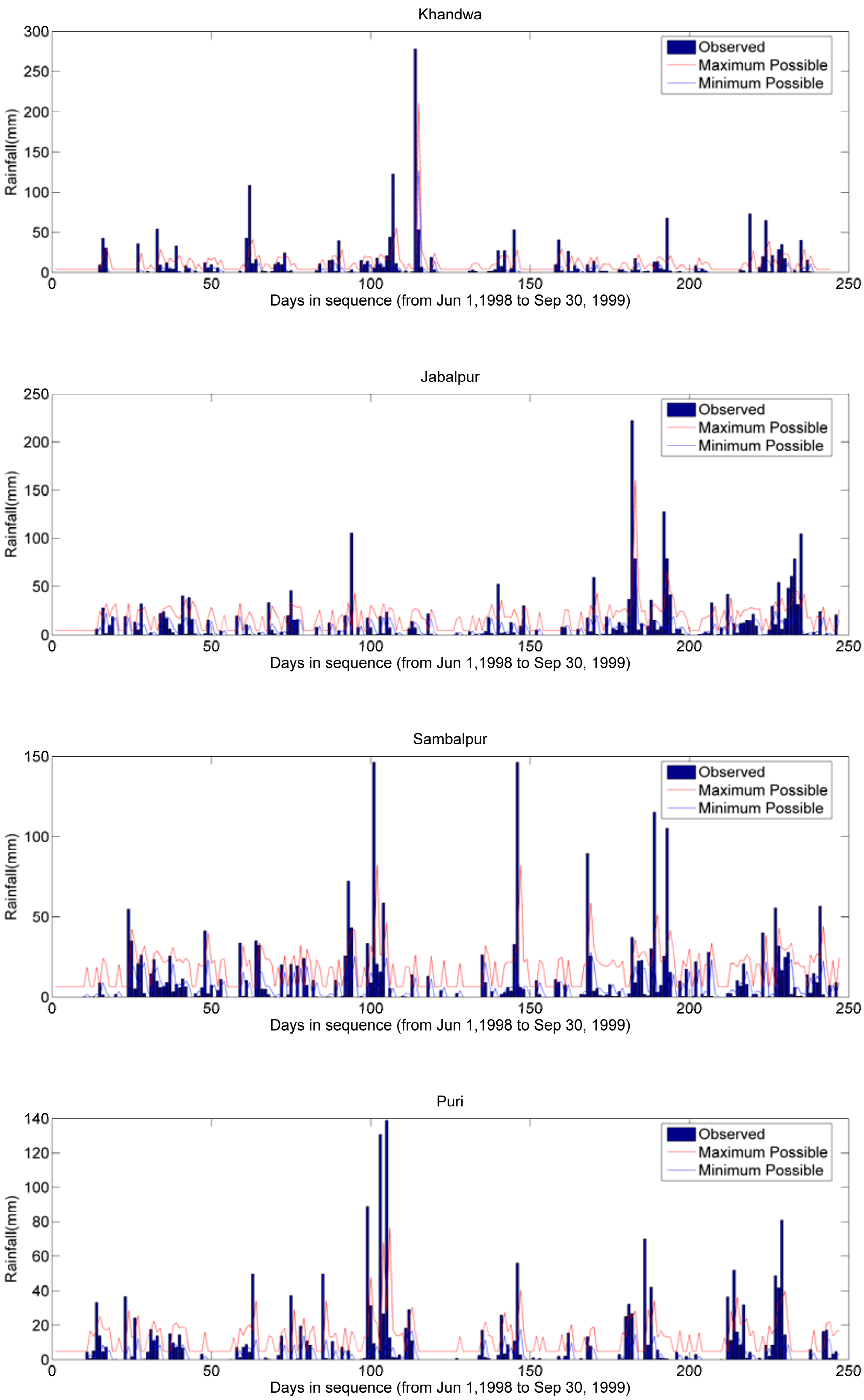

Fig. 4b: Prediction performance for the period and June 1, 1998 to September, 1999 for different station