

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

# Technical note on probabilistic assessment of one-step-ahead rainfall variation by Split Markov Process

R. Maity and D. Prasad

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

Received: 1 December 2010 – Accepted: 3 January 2011 – Published: 12 January 2011

Correspondence to: R. Maity (rajib@civil.iitkgp.ernet.in)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD

8, 189–206, 2011

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 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 of Madhya Pradesh, India. Eighty years of daily monsoon rainfall is used to develop the state/sub-state TPM and twenty years data is used to investigate its performance. It is observed that the predicted range of daily rainfall captures 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.

## 1 Introduction

Rainfall is one of the most complex and difficult components of the hydrologic cycle to model due to the complexity of the atmospheric processes and the wide range of variation both in space and in time. However, prior information of rainfall is essential (both at

**HESSD**

8, 189–206, 2011

### Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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 creates many water-related problems, such as, flood and drought, which seriously affect the crop production. Reasonably accurate rainfall prediction is required to alleviate such problems through 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 26 July 2005 caused 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. Scientists at National Centre for Medium Range Weather Forecasting (NCMRWF), which is a premier institute to provide medium range weather forecast in India, indicate the limitations in the prediction performance for severe weather events, which have a very short life but still cause extensive damage (Bohra et al., 2006). 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.

Probabilistic rainfall prediction has a long history to assess 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, 2000, a, b). The approach is developed for station rainfall data. However, the temporal resolution is seasonal to inter-annual. Application of Markov Process (MP) for short-term rainfall forecast

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



through a probabilistic way is well accepted for a long time. 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 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. The model is based on the estimated transitional probabilities and frequency distributions of rainfall amounts. It is 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 non-parametrically and estimated the transition probabilities through a weighted average of transition by kernel estimator. Baik et al. (2006) developed the transition probabilities of different condition states in Markov chain-based deterioration models for wastewater systems using an ordered probit model and gained greater accuracy in deterioration modeling.

Almost all these approaches follow a general path of creating a set of different states depending on historical record and the probabilities of transition from one state to another is obtained. For rainfall variation study, the change in rainfall magnitude is crucial information as indicated before. However, quantifying these changes, through a single

**Probabilistic  
assessment of  
rainfall variation**

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



set of states, demands large number of defined states. Generally, in the tropical countries, the variation of daily rainfall is very high. 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 study. 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 single event of rainfall (e.g., initiation, growth, peak, decay and vanish) is not of interest. Rather the total depth of rainfall in a day is considered, which is important from water resources point of view. Thus, the transitions through states to sub-states are computed through state/sub-state Transitional Probability Matrix (TPM) for a daily temporal resolution. This state/sub-state TPM is then 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 rain gauge station in the Nimar district of 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 sum of all elements in any row equal to 1 as the transitions from a particular state to all possible states are “collectively exhaustive”.

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

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Process is expressed as  $P[X_t=a_j/X_{t-1}=a_i, X_{t-2}=a_k, \dots, X_{t-m}=a_l]$ . 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

$$P_{ij} = P[X_t = a_j / X_{t-1} = a_i] \quad (1)$$

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. In order to investigate the daily rainfall variation, another set of sub-states is introduced in addition to the set of states. Thus, the states categorize the daily rainfall amount and the sub-states categorize the change in daily rainfall magnitude. The observed rainfall data is classified in different categories depending on its variability and these categories are denoted as different states, say,  $S_1, S_2, \dots, S_n$ ,  $n$  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,  $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_m$ ,  $m$  being the total number of states. The transitions from a particular state to a particular sub-state are computed from historical data and denoted as state/sub-state transition probability. The general  $m$ th order state/sub-state transition probability is expressed as

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

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

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



for the next time step depends on the state of the system at the present time. Thus, a first-order state/sub-state transition probability is expressed as

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

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

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

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

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

The total number of times a particular state is observed and its transitions to different sub-states are 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

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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.

### 3 Application of SMP

The methodology is applied to the daily rainfall at Khandwa rain gauge station located in the Nimar district in Madhya Pradesh, India. The geographical location of the rain gauge station is at latitude  $21^{\circ}$  N and longitude  $79^{\circ}$  E. The daily rainfall data is collected for the period 1901 to 1999 from Indian Meteorological Department (IMD), Pune. The data is used for the monsoon period (June to September) only as most of the annual rainfall (above 80%) occurs in this period only. 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.

#### 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  $\rightarrow R = 0$

State 2  $\rightarrow 0 < R \leq 5$

State 3  $\rightarrow 5 < R \leq 10$

State 4  $\rightarrow 10 < R \leq 20$

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion







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 levels. 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 through 4 at 95% probability level. The third point is that for all the probability lines, for higher 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-state. 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 is 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 and mean square error between observed and the average of upper and lower limits predicted range is prepared (Fig. 3). It is found that the best performance is obtained at 81% probability level. Thus, the prediction performance is obtained at this probability level for the period 1981 to 1999 and shown in Fig. 4. Top panel of Fig. 4 shows the performance for entire validation period and, for clarity, the prediction performance for the period 1998 to 1999 is shown in the bottom panel of Fig. 4. Upper and lower limits of possible next day rainfall are shown in this plot along with the actual observed rainfall. It is found that the observed rainfall lies

**Probabilistic  
assessment of  
rainfall variation**

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## References

- Baik, H. S., Jeong, H. S., and Abraham, D. M.: Estimating transition probabilities in markov chainbased deterioration models for management of wastewater systems, *J. Water Resour. Plan. Manage.*, 132(1), 15–24, 2006.
- 5 Bohra, A. K., Basu, S., Rajagopal, E. N., Iyengar, G. R., Das Gupta, M., Ashrit, R., and Athiyaman, B.: Heavy rainfall episode over Mumbai on 26 July 2005: assessment of NWP guidance, *Curr. Science*, 90(9), 1188–1194, 2006.
- Box, G., Jenkins, G. M., and Reinsel, G.: *Time Series Analysis: Forecasting and Control*, Prentice-Hall, NJ, 1976.
- 10 Fraedrich, K. and Leslie, L. M.: Combining predictive schemes in short-term forecasting, *Mon. Weather Rev.*, 115, 1640–1644, 1987.
- Fraedrich, K. and Mullar, K.: On single station forecasting: sunshine and rainfall Markov Chains, *Beitr. Phys. Atmosph.*, 56(1), 108–134, 1983.
- Haan, C. T.: *Statistical Methods in Hydrology*, 2nd edn., Iowa State Press, Iowa, 2002.
- 15 Haan, C. T., Allen, D. M., and Street, J. O.: A Markov chain model for daily rainfall, *Water Resour. Res.*, 12(3), 443–449, 1976.
- Kaseke, T. N. and Thompson, M. E.: Estimation for rainfall-runoff modeled as a partially observed Markov Process, *Stoch. Hydrol. Hydraul.*, 11(1), 1–16, 1997.
- Rajagopalan, B., Lall, U., and Tarbotan, D. G.: Nonhomogeneous Markov Model for daily precipitation, *J. Hydrol. Engin.*, 1(1), 33–40, 1996.
- 20 Sharma, A.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – a strategy for system predictor identification, *J. Hydrol.*, 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, *J. Hydrol.*, 239(1–4), 249–258, doi:10.1016/S0022-1694(00)00348-6, 2000b.
- 25 Sharma, A., Luk, K. C., Cordery, I., and Lall, U.: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 – predictor identification of quarterly rainfall using ocean-atmosphere information, *J. Hydrol.*, 239(1–4), 240–248, doi:10.1016/S0022-1694(00)00347-4, 2000.
- 30 Stern, R. D. and Coe, R.: A model fitting analysis of daily rainfall data, *J. Royal Stat. Soc. Ser. A*, 147(1), 1–34, 1984.

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Weeks, W. D. and Boughton, W. C.: Tests of ARMA model forms for rainfall-runoff modeling, *J. Hydrol.*, 91(1–2), 29–47, 1987.
- Wójcik, R., Torfs, P., and Warmerdam, P.: Application of Parzen densities to probabilistic rainfall-runoff modeling, *J. Hydrol. Hydromech.*, 51(3), 175–186, 2003.

# HESSD

8, 189–206, 2011

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

**Table 1.** 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

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

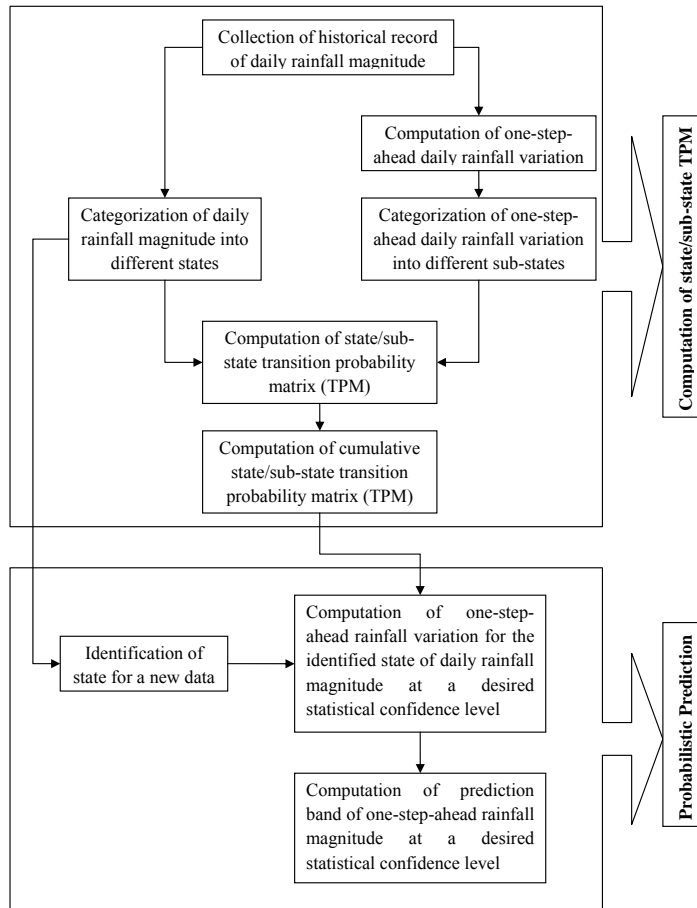
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





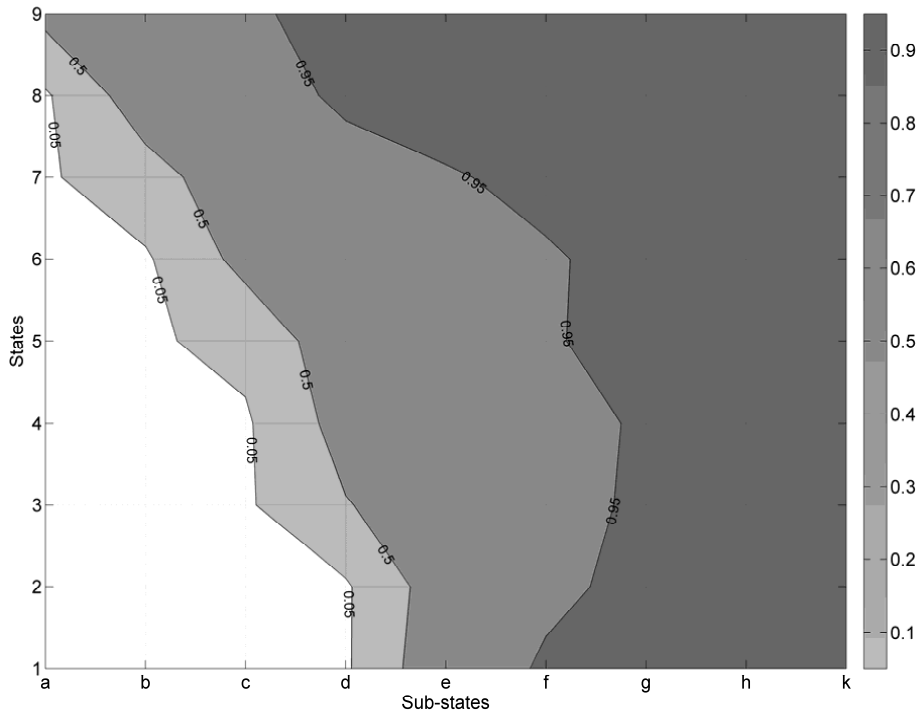
**Fig. 1.** Flowchart showing major steps of Split Markov Process (SMP).

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

<a href="#">Title Page</a>	
<a href="#">Abstract</a>	<a href="#">Introduction</a>
<a href="#">Conclusions</a>	<a href="#">References</a>
<a href="#">Tables</a>	<a href="#">Figures</a>
<a href="#">⏪</a>	<a href="#">⏩</a>
<a href="#">◀</a>	<a href="#">▶</a>
<a href="#">Back</a>	<a href="#">Close</a>
<a href="#">Full Screen / Esc</a>	
<a href="#">Printer-friendly Version</a>	
<a href="#">Interactive Discussion</a>	





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

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

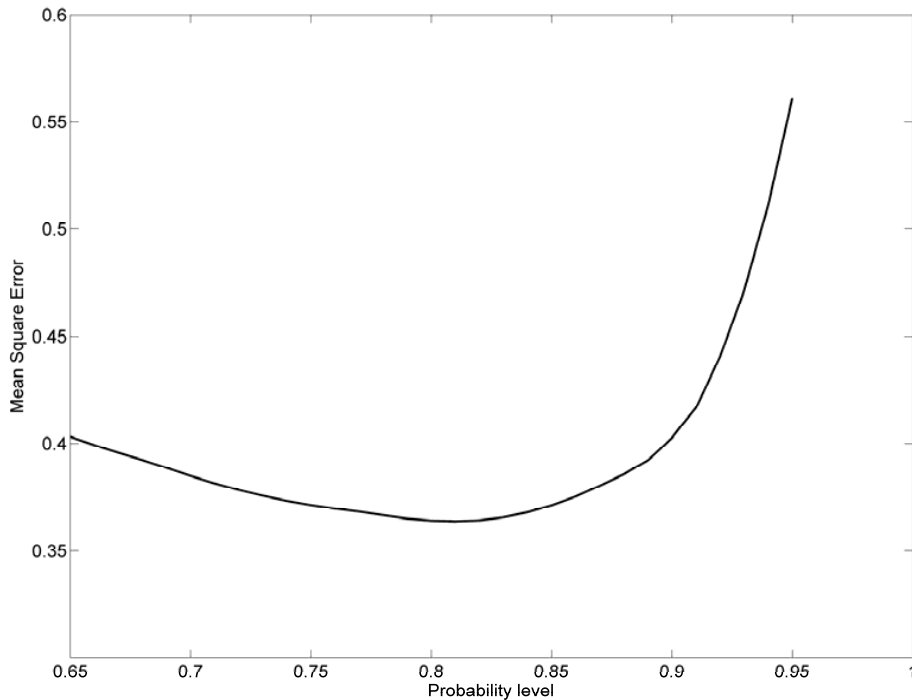
Full Screen / Esc

Printer-friendly Version

Interactive Discussion







**Fig. 3.** Plot between probability level and Mean Square Error (MSE).

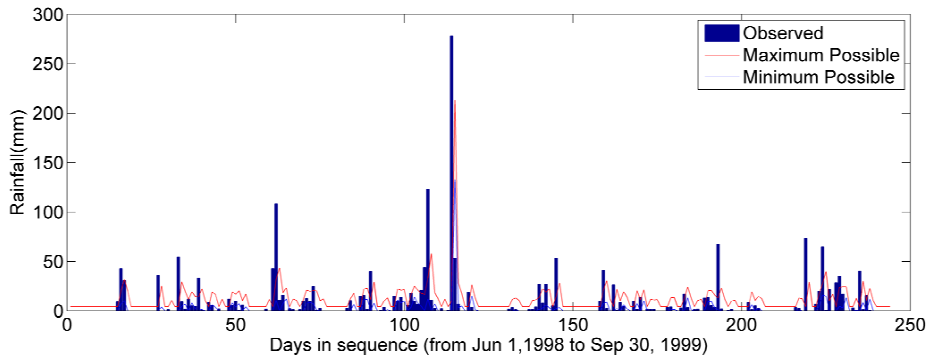
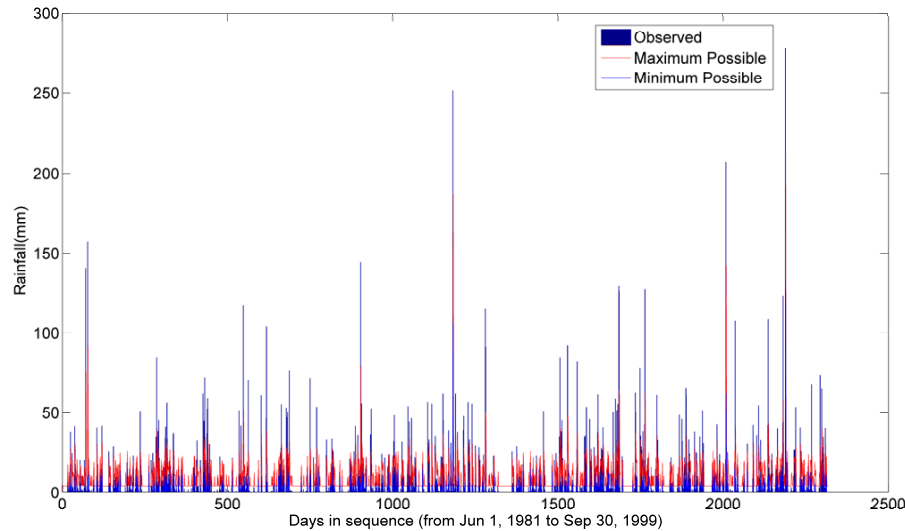
## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	





**Fig. 4.** Prediction performance for the period 1 June 1981 to September 1999 (top) and 1 June 1998 to September 1999 (bottom).

## Probabilistic assessment of rainfall variation

R. Maity and D. Prasad

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

