HOURLY RAINFALL GENERATION BY MARKOV CHAIN RAINFALL GENERATOR

UJJWAL SAHA

[1] Civil Engineering Department, IIEST Shibpur, Howrah-3
E-mail: [1]ujjwalsaha1980@gmail

Abstract- Simulation of hourly rainfall is very important in urban hydrology and small catchments with fast concentration time to model the hydrological processes and prediction of flash flood. Considerable efforts have been given to simulate the daily rainfall but there is not much research on modeling of hourly rainfall. Most hourly rainfall simulation model is based on poisson cluster theory which requires large number of parameters to be estimated and the parameter estimation procedure is complex. To overcome this problem, in this study, a first order two state Markov chain rainfall generator with gamma distribution is used to simulate hourly rainfall of a station in the city of Kolkata. The results indicate that the model is successful to simulate the diurnal cycle of rainfall with an acceptable degree of accuracy but need further improvement.

Keywords- Hourly Rainfall Generator, Markov Chain, Transitional Probabilities, Diurnal Cycle

I. INTRODUCTION

To combat the flooding in urban areas, storm water drains are constructed. The urban storm water systems are designed based on extreme rainfall amount for the return period equal to the expected life of the system and duration of rainfall is equal to the time of concentration of the urban watershed for which the storm water systems are to be designed. As variability of various hydrological, topographical and social parameters is high in an urban area, typically urban storm water facilities are designed for small catchment areas. The time of concentration falls between 45 minutes to 1 hour. Sometimes it can be as low as 15 minutes (Mailhot and Duchesne 2010). Frequency analysis of rainfall data for the appropriate durations has to be performed to get the required design rainfall. But, in developing countries such as India, the availability of short duration rainfall data is scarce. Sometimes, the hourly data will have missing values. Hence, simulation of hourly data is extremely important area but unfortunately not much effort is spent in this area. In this study, an effort has been made to simulate the hourly rainfall of a station in Kolkata city, India.

The simulation of hourly rainfall follows four broad methods. In first set of method, at first daily rainfall is simulated and then the daily rainfall values are disaggregated into hourly values by different techniques such as Random Cascade, scaling theory. Second set of method model the rainfall as Poisson cluster process. The Bartlett-Lewis and Neyman-Scott pulse processes (Rodriguez-Iturbe et al., 1987 and 1998) are the two predominant model variants in this category. Cluster process models provide occurrence and amount in the same model (Koutsoyiannis and Onof, 2001). Both models are based on Poisson process of arrival of storm origins with each origin associated with a random number of cells. In the Neyman Scott process, the positions of the cells are determined from the storm origin assuming it as an independent and identically distributed. On the other hand, in the Barlett-Lewis process, the intervals between consecutive cells are assumed to be independent and identically distributed. In both cluster based models, it is assumed that the placement of cells is governed by an exponential distribution that characterizes either the positioning of cells from the storm origin (NSRP case) or the inter-arrival time between cells (BLRP case). The main problem with cluster process models is that of parameterization. Studies (e.g., Cowpertwait et al., 2002) have noted that cluster process rainfall models break down or reveal lack of fit at certain temporal or spatial scales The third method, non-parametric rainfall generators, have evolved as a simple way to simulate rainfall data while avoiding the statistical assumptions regarding the probability distributions of rainfall. Most of the existing algorithms are extensions of the Young (1994) approach, using a nearest-neighbor resampling scheme in selection of the next time scale weather. The fourth category of method is called chain dependant process and most popular in this category is Markov chain rainfall generators. A Markov chain may be thought as a collection of states of a system. For each time period the state of the Markov chain can either continue in the same state or move into another state (Wilks, 1999). For modeling of the rainfall occurrence, the Markov chain approach is generally limited to two states: rainy or dry. The behavior of a Markov chain is governed by conditional probabilities known as transition probabilities. A transition probability is the probability of going from a given state to the next state in a Markov chain.

There is only one significant work of applying the Markov chain process in simulation of hourly rainfall (Katz, 1995). The authors assumed dependency between successive rainfall amounts by an autoregressive model which made the model complicated. In this work, It is assumed that rainfall

amount of different hours in a day will be independent to each other. The methodology of the simulation process is described in the next section. In the third section, the results of the output of the model is presented and discussed. Finally in the last section, conclusions are drawn.

II. SIMULATION PROCESS

The hourly rainfall is simulated in a two step process. In the first step, the rainfall occurrence i.e sequence of dry and wet hours is simulated and in the next step, the rainfall amount on the wet hours is simulated. For simulating rainfall occurrence, the hourly rainfall process is assumed as an random variable { RSd(h) ; h = 1,2 3 to 24 and d is 1,2… D where D is the total length of the daily time series} is defined as

RSd(h)= 0 if there is no rainfall occurrence on that h th hour of d th day

RSd(h) = 1 if there is rainfall occurrence

Assuming that the random function, R(t) evolves as a function of its previous state and it follows Markovian dependence of first order,

$$P[RS_d(h)=1 \mid RS_d(h-1), RS_d(h-2), ...RS_d(1)]$$
$$= P[RS_d(h)=1 \mid RS_d(h-1)] \quad (1)$$

The generation of the states 0 and 1 of R(t) requires the specification of conditional or transition probabilities. As the random variable R(t) have two state space, it will have a transition probability matrix of 2 × 2 in the form of

$$p_{00} = P[RS_d(h)=0 \mid RS_d(h-1)=0]$$
$$p_{01} = P[RS_d(h)=0 \mid RS_d(h-1)=1]$$
$$p_{10} = P[RS_d(h)=1 \mid RS_d(h-1)=0]$$
$$p_{11} = P[RS_d(h)=1 \mid RS_d(h-1)=1] \quad (2)$$

As there is only two possible state in a particular hour, it follows that the transition probabilities are related by

$$p_{10} = 1 - p_{11}$$
$$p_{01} = 1 - p_{00} \quad (3)$$

Hence, the transition probability matrix can be defined if only two transition probabilities, p01 and p11 is estimated from the observed data. The conditional probabilities can be estimated from the observed conditional frequency of rainfall state transition as follows:

		Current hour		
Previous hour		Dry (0)	Rain (1)	Total
	Dry (0)	N00	N01	Nd
	Rain (1)	N10	N11	Nr

Where N00, N01, N10 and N11 are the observed frequencies:

N00 = Number of dry hours preceded by dry hours
N01 = Number of rain hours preceded by dry hours
N10 = Number of dry hours preceded by rain hours
N11 = Number of rain hours preceded by rain hours
And N00 + N01 = Nd, total number of dry hours
N10 + N11 = Nr, total number of rain hours

Now the maximum likelihood estimator for the transition probabilities are

$$p_{ij} = \frac{Nij}{Nt} \quad \text{for i, j = 1,2} \quad (4)$$

Where Nt = Nd when i=0
= Nr when i=1

Here in order to take seasonal characteristics of rainfall into account, the transitional probabilities are calculated for each month. This improves the intra-annual representation of the random variable RSd(h).

Once the rainfall occurrence process is simulated, next step is to simulate the rainfall amount (Ra) in each wet hour. The rainfall amount generation process is done fitting Gamma distribution to wet hours.

The daily rainfall amount 'x' on wet hour is described by a Gamma distribution with probability function

$$f(x) = \frac{\left(x/\beta\right)^{\alpha-1} \exp\left(-x/\beta\right)}{\beta\Gamma(\alpha)} \quad ; \ \alpha, \beta > 0; \quad x > 0 \quad (5)$$

where α is the shape parameter, and β is the scale parameter, $\Gamma(\alpha)$ is the Gamma function. The parameters of Gamma distributions are estimated by maximum likelihood method .

After the transition probabilities and Gamma distribution parameters for each month are estimated from the observed data, the dry-rain status of each hour needs to be established. For determining dry-rain status of the first hour of the day, a uniform random number is generated; if this random number is greater than 0.5, then the day is taken as dry, otherwise it is considered as a wet. To determine whether the next hour of the sequence is rain or dry, critical value of probability, Pcrit, of rainfall is calculated. It depends on the dry or rain status of the previous hour. If the previous hour is dry, then Pcrit= p01 otherwise Pcrit= p11. Again another uniform random number (u) is generated. The current hour is simulated as dry if u > Pcrit otherwise it is a wet hour. This procedure continues for all remaining days of the year.

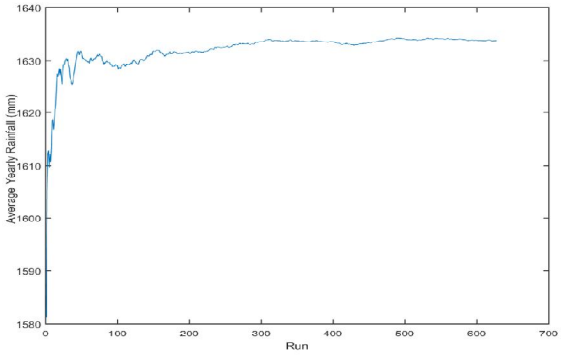
After rainfall status of each hour is determined, next step is to simulate the rainfall amounts in each hour. This is done by monte-carlo simulation. Again, a uniform random number is generated between 1 and 0. This random number is assumed to be CDF of the respective Gamma distribution (equation (5)). The rainfall amount is obtained by setting the CDF equal

to the uniform random number and solving the CDF. Thus the hourly rainfall for each hour is simulated. Since the model uses a random number, every time the model is run it gives a different result. To tackle this issue, the model is run 100 times and the mean of the statistics is taken as simulated. Note that in this study, it is the statistical properties of rainfall which is more important than the exact sequence of rainfall, and a rainfall generator is designed for that purpose.

III. RESULTS AND DISCUSSION

At first it was decided that how many runs of the algorithm are required so that the results converge. In the following figure 1, it can be observed that after at around 100 runs, results seem to be converging. Hence, in this study, it is decided to run the model for 100 times to get the average of the statistical parameters.

At first, the parameters of the models are calculated. In order to investigate further, four types of model have been used in this model. Case1: The model parameters (i.e. Transition probabilities p01 and p11 and gamma distribution parameters α and β) are not allowed to vary diurnally. Case2: The Transition probabilities are not allowed to vary diurnally while gamma distribution parameters are varied diurnally. Case3: All the parameters are allowed to vary diurnally



The calculated parameters of the model in all the cases are given below.

Table1: Model Parameters of Dumdum Station assuming no Diurnal Variation

Month	P01	P11	alpha	beta
Jan	0.004	0.627	0.73	2.02
Feb	0.007	0.516	0.57	4.17
Mar	0.008	0.490	0.54	5.97
Apr	0.011	0.490	0.53	6.80
May	0.021	0.565	0.50	9.05
Jun	0.052	0.580	0.52	6.45
July	0.078	0.570	0.52	5.56
Aug	0.076	0.558	0.54	5.65
Sept	0.056	0.596	0.54	6.07
Oct	0.022	0.614	0.54	6.15
Nov	0.006	0.690	0.57	4.68
Dec	0.002	0.671	0.57	3.34

Table2: Model Parameter (Alpha) of Dumdum Station with Hourly Variation

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Jan	0.72	0.64	1.08	0.57	1.03	0.66	0.69	0.82	1.12	0.53	0.54	0.80	1.16	0.73	0.70	0.64	0.69	1.14	0.88	1.10	0.63	0.82	1.65	1.02
Feb	0.58	0.69	0.55	0.39	0.61	0.53	0.73	0.93	1.08	3.91	0.50	1.03	0.52	0.82	0.70	0.79	0.52	0.83	0.64	0.57	0.65	0.86	0.63	0.50
Mar	0.52	0.48	0.69	0.41	0.66	0.87	0.62	1.25	0.46	0.87	0.40	0.64	0.83	0.51	0.74	0.65	0.54	0.70	0.54	0.61	0.52	0.59	0.63	0.49
Apr	0.45	0.79	0.45	0.37	0.60	0.52	0.58	0.46	1.17	0.61	0.49	1.00	0.55	0.91	0.48	0.66	0.46	0.61	0.52	0.57	0.56	0.70	0.55	1.10
May	0.56	0.66	0.47	0.46	0.45	0.41	0.65	0.50	0.99	0.66	0.42	0.50	0.65	0.50	0.62	0.46	0.40	0.53	0.54	0.52	0.52	0.47	0.53	0.59
Jun	0.47	0.51	0.76	0.59	0.57	0.60	0.54	0.53	0.63	0.62	0.53	0.52	0.58	0.56	0.52	0.55	0.52	0.46	0.50	0.45	0.46	0.42	0.46	0.47
July	0.46	0.54	0.57	0.63	0.54	0.55	0.52	0.48	0.59	0.52	0.53	0.62	0.51	0.61	0.48	0.52	0.52	0.49	0.53	0.57	0.48	0.53	0.46	0.47
Aug	0.51	0.49	0.50	0.51	0.63	0.54	0.57	0.58	0.58	0.67	0.65	0.59	0.53	0.49	0.53	0.57	0.51	0.51	0.47	0.59	0.51	0.56	0.52	0.59
Sept	0.57	0.56	0.55	0.53	0.61	0.64	0.55	0.55	0.61	0.63	0.67	0.55	0.53	0.58	0.50	0.54	0.48	0.53	0.47	0.44	0.49	0.50	0.51	0.54
Oct	0.60	0.47	0.57	0.69	0.60	0.58	0.59	0.45	0.63	0.70	0.56	0.48	0.55	0.68	0.64	0.50	0.45	0.48	0.49	0.49	0.52	0.56	0.61	0.65
Nov	0.52	0.42	0.49	0.51	0.70	0.70	0.56	0.62	0.69	0.62	0.76	0.64	0.70	0.74	0.62	0.66	0.58	0.55	0.94	0.56	0.75	0.69	0.60	0.67
Dec	0.43	0.70	0.33	0.68	0.67	0.99	1.07	0.82	0.58	1.41	1.97	1.62	1.69	1.19	1.84	0.80	0.61	0.52	0.62	0.35	0.36	0.42	0.36	0.45

Table3: Model Parameter (Beta) of Dumdum Station with Hourly Variation

Month	1	2	3	4	5	6	7	8	9	1	11	12	1	14	1	16	17	18	19	20	21	22	23	24
Jan	1.98	2.50	1.22	4.79	1.69	1.84	2.63	2.43	1.20	3.56	3.11	1.03	0.79	1.27	2.87	2.04	1.82	0.66	2.03	1.46	1.90	1.25	0.37	1.43
Feb	5.60	3.63	3.20	4.69	5.73	5.30	1.85	1.20	0.58	0.08	6.38	2.68	5.74	2.95	4.91	2.94	10.28	0.84	3.48	5.78	2.99	1.34	3.22	5.01
Mar	6.89	8.14	1.67	9.99	2.26	2.21	6.30	0.41	6.61	1.56	12.66	2.06	1.74	9.14	5.66	8.49	8.71	2.63	8.78	5.31	7.72	4.14	5.44	6.38
Apr	9.30	2.49	7.35	14.66	2.04	5.12	3.94	6.31	10.16	6.13	3.23	2.16	7.58	2.03	7.76	2.75	10.76	7.54	10.76	6.74	4.41	3.22	5.62	1.93
May	4.98	3.82	6.40	5.91	5.46	13.27	5.35	7.58	2.17	3.58	10.63	9.56	7.78	11.45	6.90	12.23	21.69	12.59	11.78	8.59	7.26	8.48	6.32	3.25
Jun	8.16	5.76	2.23	4.46	5.56	5.24	4.76	6.29	5.17	4.01	7.44	6.12	4.66	5.91	6.58	5.72	7.84	9.43	7.44	10.09	6.94	10.95	5.78	7.84
July	7.76	4.92	3.86	3.79	5.62	4.90	5.95	5.08	3.79	5.25	4.49	3.98	7.05	5.13	8.18	6.33	5.28	7.06	6.29	4.27	6.33	4.21	6.20	5.59
Aug	7.82	7.58	7.67	6.67	4.66	5.42	4.02	3.50	4.61	3.25	3.37	5.22	6.27	7.84	7.69	5.00	6.27	5.40	6.93	3.74	5.16	4.29	5.81	4.25
Sept	4.81	5.22	4.74	6.56	4.45	4.74	5.11	6.06	3.58	4.78	3.54	7.69	9.54	6.77	6.95	4.85	6.80	4.17	7.02	8.58	6.34	5.83	7.42	7.21
Oct	5.52	8.64	5.46	3.11	5.49	5.66	5.12	8.39	4.50	4.93	6.65	10.40	9.20	4.77	5.87	7.06	8.55	6.53	5.06	4.54	4.44	3.57	2.68	3.60
Nov	5.03	11.23	8.77	5.33	2.55	1.90	4.21	4.13	2.36	4.62	5.79	3.67	5.34	3.75	4.58	3.39	3.79	5.40	1.47	5.84	1.43	4.92	2.77	2.35
Dec	6.18	1.68	23.27	2.58	2.36	2.93	1.48	1.04	2.42	1.35	0.39	0.56	0.53	0.72	0.47	1.78	2.43	3.64	1.65	6.33	10.67	6.11	9.71	4.22

Table4: Model Parameter (p11) of Dumdum Station with Hourly Variation

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Jan	0.58	0.58	0.52	0.73	0.82	0.57	0.64	0.56	0.56	0.50	0.75	0.36	0.85	0.90	0.50	0.71	0.37	0.77	0.63	0.91	0.53	0.54	0.55	0.66
Feb	0.61	0.59	0.42	0.50	0.50	0.43	0.45	0.30	0.44	0.60	0.54	0.83	0.42	0.46	0.50	0.46	0.47	0.58	0.42	0.73	0.47	0.47	0.55	0.52
Mar	0.58	0.47	0.50	0.62	0.56	0.61	0.28	0.50	0.53	0.46	0.50	0.62	0.45	0.61	0.50	0.40	0.50	0.40	0.46	0.47	0.51	0.34	0.66	0.34
Apr	0.66	0.46	0.54	0.52	0.31	0.43	0.44	0.07	0.62	0.30	0.50	0.72	0.29	0.55	0.47	0.41	0.46	0.61	0.52	0.46	0.45	0.55	0.40	0.62
May	0.66	0.63	0.68	0.44	0.61	0.51	0.73	0.40	0.63	0.65	0.71	0.67	0.45	0.63	0.64	0.48	0.59	0.59	0.64	0.56	0.52	0.44	0.48	0.36
Jun	0.67	0.62	0.57	0.63	0.61	0.65	0.65	0.52	0.56	0.63	0.59	0.59	0.61	0.56	0.47	0.54	0.56	0.50	0.59	0.56	0.51	0.65	0.58	0.51
July	0.60	0.67	0.61	0.59	0.64	0.65	0.63	0.48	0.61	0.59	0.58	0.55	0.49	0.55	0.49	0.49	0.49	0.56	0.59	0.57	0.55	0.59	0.59	0.54
Aug	0.65	0.66	0.60	0.64	0.62	0.59	0.61	0.54	0.62	0.57	0.54	0.61	0.49	0.50	0.53	0.51	0.46	0.48	0.46	0.47	0.50	0.59	0.63	0.58
Sept	0.64	0.67	0.73	0.70	0.71	0.69	0.62	0.58	0.66	0.63	0.60	0.60	0.58	0.57	0.53	0.45	0.51	0.48	0.49	0.55	0.62	0.58	0.62	0.61
Oct	0.70	0.64	0.69	0.69	0.74	0.60	0.76	0.60	0.55	0.66	0.69	0.60	0.61	0.48	0.58	0.49	0.61	0.52	0.55	0.67	0.62	0.57	0.62	0.63
Nov	0.85	0.83	0.54	0.67	0.60	0.68	0.52	0.52	0.88	0.60	0.83	0.58	0.68	0.78	0.54	0.57	0.72	0.69	0.71	0.72	0.61	0.91	0.93	0.78
Dec	0.58	0.50	0.66	0.66	0.61	0.72	0.66	0.66	0.66	0.66	0.75	0.75	0.66	0.62	0.71	0.57	0.75	0.50	0.71	0.85	0.71	0.71	0.85	0.75

Using these parameters, the model is run for 100 times and the hourly rainfall time series are generated and then the various rainfall statistics are calculated and compared with the observed. The results are shown below.

Case 1:

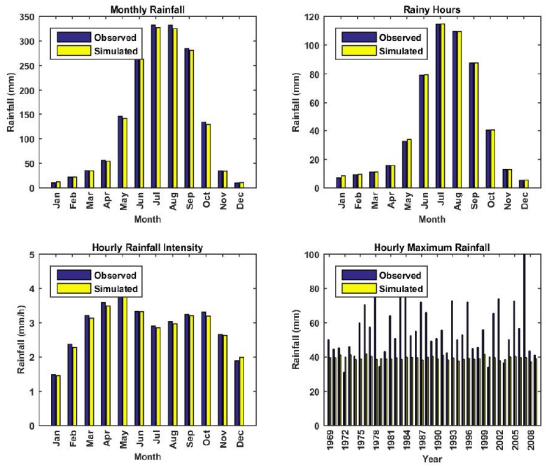


Figure 2: Monthly Rainfall Simulation

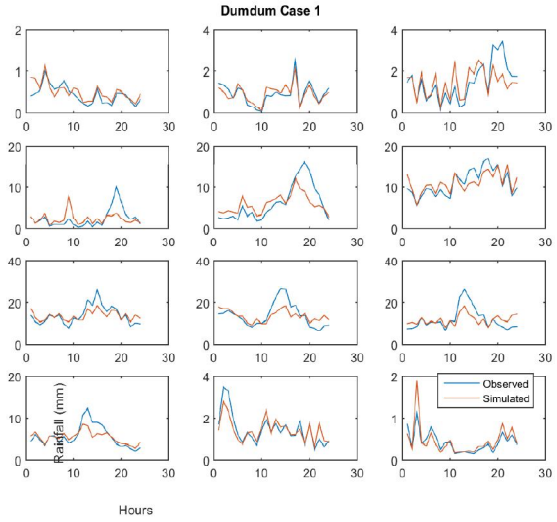


Figure 3: Duirnal Rainfall Simulation

Case 2:

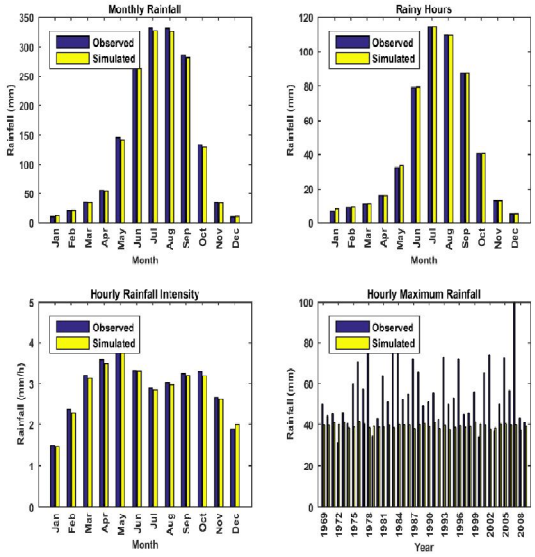


Figure 4: Monthly Rainfall Simulation

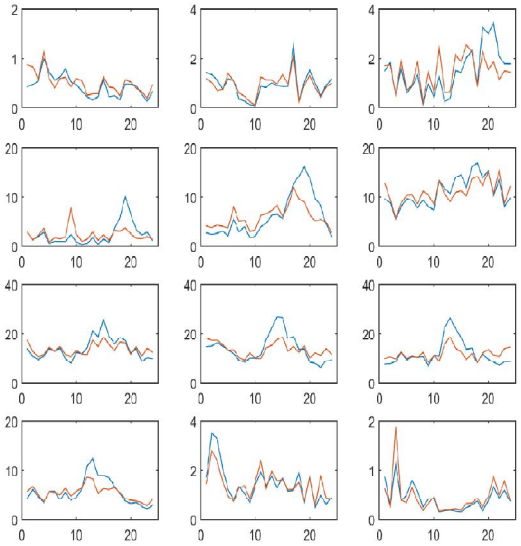


Figure 5: Duirnal Rainfall Simulation

Case 3:

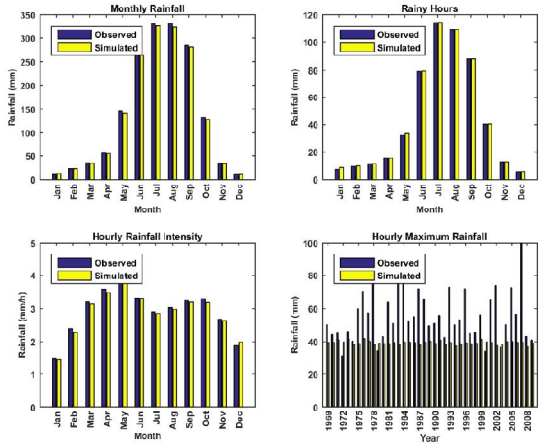


Figure 6: Monthly Rainfall Simulation

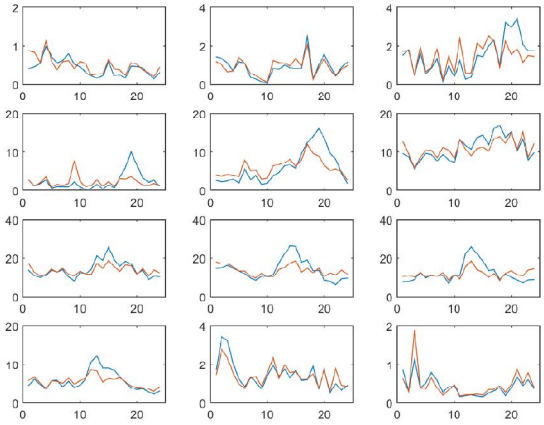


Figure 7: Duirnal Rainfall Simulation

CONCLUSIONS

The hourly rainfall distribution and various statistics of rainfall at different aggregation level are simulated with desired accuracy by this markov chain model but the extreme rainfall statistics is not simulated well. Hence, improvement of this model is necessary

REFERENCES

[1] Cowpertwait, P. S. P., Kilsby, C. G., and O'Connell, P. E. (2002), A spacetime NeymanScott model of rainfall: Empirical analysis of extremes, Water Resources Research, 38(8), 6-11.

[2] Katz, R. W., & Parlange, M. B. (1995). Generalizations of chain-dependent processes: Application to hourly precipitation. Water Resources Research, 31(EFLUM-ARTICLE-1995-007), 1331-1341.

[3] Koutsoyiannis, D., and C. Onof (2001), Rainfall disaggregation using adjusting procedures on a Poisson cluster model, Journal of Hydrology, 246, 109–122.

[4] Mailhot, A., and Duchesne, S. (2010), Design criteria of urban drainage infrastructures under climate change, J. Water Resour. Plann.Manage., 136(2), 201–208

[5] Rodriguez-Iturbe, I., Cox, D. R., Isham, V. (1987), Some models for rainfall based on stochastic point processes, Proc. R. Soc. Lond. A, 410, 269-298.

[6] Rodriguez-Iturbe, I., Cox, D. R., Isham, V. (1988), A point process model for rainfall: Further developments, Proc. R. Soc. Lond. A, 417, 283-298.

[7] Wilks, D. S., and Wilby, R. L. (1999), The weather generator game: a review of stochastic weather models, Prog. Phys. Geogr., 23, 329–358.

[8] Young, K. C. (1994), A multivariate chain model for simulating climatic parameters from daily data, Journal of Applied Meteorology, 33, 661–71.

★★★