

A Rainfall Simulation Model for Agricultural Development in Bangladesh

M. SAYEDUR RAHMAN*

Department of Statistics, Rajshahi University, Rajshahi 6205, Bangladesh

(Received 25 May 1999)

A rainfall simulation model based on a first-order Markov chain has been developed to simulate the annual variation in rainfall amount that is observed in Bangladesh. The model has been tested in the Barind Tract of Bangladesh. Few significant differences were found between the actual and simulated seasonal, annual and average monthly. The distribution of number of success is asymptotic normal distribution. When actual and simulated daily rainfall data were used to drive a crop simulation model, there was no significant difference of rice yield response. The results suggest that the rainfall simulation model perform adequately for many applications.

Keywords: Simulation modelling, Markov chain, Probability distribution, Latin Hypercube Sampling, Agro-technology

INTRODUCTION

The impact of extreme climate can have an importance for economically vulnerable farmer that far outweighs its apparent probability of occurrence. Drought and flood of various intensity occur in almost all parts of Bangladesh. Drought of different intensities occurs both in wet season (Kharif) and dry season (Rabi). During kharif season about 0.574–1.748 million hectares of T. Aman rice crops are severely affected by drought. Rain flood occurs due to high intensity rainfall over Bangladesh.

Due to inadequate drainage capacity, such rain-storm of 3–10 days duration cause localized floods that damage crops and property (Task Force Report, 1991). Simulation modelling techniques are increasingly being used as a means of extrapolating the results of field trials and of designing and assessing agro-technology that can then be passed on to farmers. This paper describes a rainfall simulation model based on a first-order Markov chain model. The results are presented, both in terms of the rainfall simulation model and a crop growth simulation model of rice.

* E-mail: rajucc@citechco.net.

THE RAINFALL MODEL

A two-state Markov chain method involves the calculation of two conditional probabilities: (1) α , the probability of a wet day following a dry day, and (2) β , the probability of a dry day following a wet day. Using a first-order Markov chain model the rainfall simulation model was developed. The model provided as input, monthly probabilities of receiving rainfall if the previous day was dry and monthly probabilities of receiving rainfall if the previous day was wet. The wet-dry probabilities are not available, it can be estimated using the average number of days of rainfall

$$P_j(W/D) = \beta dp_j/dm_j \quad (1)$$

$$P_j(W/W) = 1 - \beta + P_j(W/D) \quad (2)$$

where $P_j(W/D)$ is the probability of rainfall of wet day after dry day in month j , $P_j(W/W)$ is the probability of rainfall of wet day after wet day in month j , dp is the days of rainfall in month j and β is equal to 0.75 for the model. Given the initial wet-dry state, the model determines stochastically if rainfall occurs. A random number (0-1) is generated and compared with the appropriate wet-dry probabilities. If the random number is less than or equal to the wet-dry probability, rainfall occurs on that day. Random number greater than the wet-dry probability give no rainfall. Rainfall is modelled by using censored gamma distribution, restricted below 0.1 mm, to determine daily rainfall amounts on those days that rainfall is experienced.

Daily rainfall amount is generated from skewed normal distribution using the equation (Arnold *et al.*, 1990).

$$R = \mu + 2\sigma/\beta[\{\beta/6(z - \beta/6) + 1\}^3 - 1] \quad (3)$$

where R is the daily rainfall amount, μ is the mean size of the rainfall event given wet day, σ is the standard deviation, β is the skew coefficient and z is the standard normal deviate.

The data used in this study refers to rainfall data in mm of Barind rainfall recording station and

recorded by the Bangladesh Water Development Board. Daily time series data of long period required for climatological studies are not available. However, depending on data availability the study is based on simulation data. In Bangladesh there are three agro-climatic seasons (1) Pre-kharif (summer) which includes the pre-monsoon months of March to May, (2) kharif which includes monsoon months of June to October in which 80% of the total rainfall occurs (Rahman and Alam, 1996), and (3) rabi (winter) season from November to February.

THE MAXIMUM LIKELIHOOD ESTIMATION TECHNIQUE

Let us consider the conditional probabilities, which are denoted by

$$P_0 = \Pr\{W/D\} \quad (4)$$

$$P_1 = \Pr\{W/W\} \quad (5)$$

This sequence is an irreducible Markov chain with two ergodic states. Its stationary probability distribution has a probability of a success

$$P = P_0/1 - (P_1 - P_0) \quad (6)$$

The recurrence time distribution of success has probabilities $P_1, (1 - P_1)P_0, (1 - P_1)(1 - P_0)P_0$, for 1, 2, 3, ... trial up to and including the next success. Hence the mean and variance of recurrence times are given below:

$$\begin{aligned} \mu &= \{1 - (P_1 - P_0)\}/P_0, \\ \sigma^2 &= (1 - P_1)(1 + P_1 - P_0)/P_0^2 \end{aligned} \quad (7)$$

Now, let S be the number of success. It is known that S is asymptotically normally distributed that is

$$S \sim \text{AND}(n/\mu, n\sigma^2/\mu^3) \quad (8)$$

Substituting for and, we know that,

$$E(S) = nP, \quad \text{Var.}(S) = nP(1 - P)(1 + d)/(1 - d) \quad (9)$$

where $d = P_1 - P_0$. This is an asymptotic result which indicates neither the exact distribution for small n , nor the rapidity of approach to normality (Feller, 1957).

The maximum likelihood estimates of the transition probabilities P_{ij} can be estimated by maximizing the probability $\prod_{ij} P_{ij}^{n_{ij}}$ with respect to the P_{ij} , where $P_{ij} \geq 0$ and $\sum_{ij} P_{ij} = 1, i = 1, 2, \dots, m$ when the n_{ij} are the actual observations. It is well-known and easily verified that the maximum likelihood estimates for P_{ij} are $\hat{P}_{ij} = n_{ij}/n_i^*$, where $n_i^* = \sum n_{ij}$ multinomial trials with probabilities P_{ij} ($i, j = 1, 2, 3, \dots, m$).

THE MULTIVARIATE LOGISTIC REGRESSION TECHNIQUE

Multivariate logistic regression is used to predict dichotomous events. Let us consider $P(X_1, X_2, \dots, X_k)$ be the probability that it is raining. Let the logistic response function P is the form

$$P(X_1, X_2, \dots, X_k) = \exp(f) / \{1 + \exp(f)\} \quad (10)$$

where

$$f = \beta_0 + \sum \beta_j X_j. \quad (11)$$

It is a monotonic function of f and its values lie between 0 and 1. Let us define dichotomous variables Y as

$$Y = \begin{cases} 0, & \text{not raining} \\ 1, & \text{raining} \end{cases} \quad (12)$$

Then a set of vectors $X_i = (Y_i, X_{1i}, X_{2i}, X_{ki}), i = 1, 2, \dots, n$. To estimate the β 's, the following procedure can be used

$$P(Y_i = 1) = P_i = \exp(f) / \{1 + \exp(f)\} \quad (13)$$

$$P(Y_i = 0) = 1 - P_i \quad (14)$$

The joint distribution of Y_1, Y_2, \dots, Y_n is given by

$$g(Y_1, Y_2, \dots, Y_n) = \prod P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad (15)$$

The minimize of the negative log of g is equal to the maximize of the joint distribution function. Therefore, the loss function is

Loss

$$= - \left[\sum Y_i \log(P_i) + \sum (1 - Y_i) \log(1 - P_i) \right] \quad (16)$$

A nonlinear procedure is suitable to minimize the loss function (Wilkinson, 1990).

THE CROP GROWTH MODEL

A combined weather generator and crop growth model has been built (Fig. 1). This is based on the EPIC model (Sharpley and Williams, 1990). The crop model has been tested in the Barind Tract in Bangladesh (Rahman and Rushton, 2000). Crop models are just working hypothesis, and it is never possible to prove a hypothesis absolutely correct in science. However, by testing a model under diverse conditions, ample opportunity is provided to identify its areas of weakness (Whisler *et al.*, 1986). This is a simulation model of crop growth and development that operates on a daily step. Photosynthetically active radiation and the dry matter produced

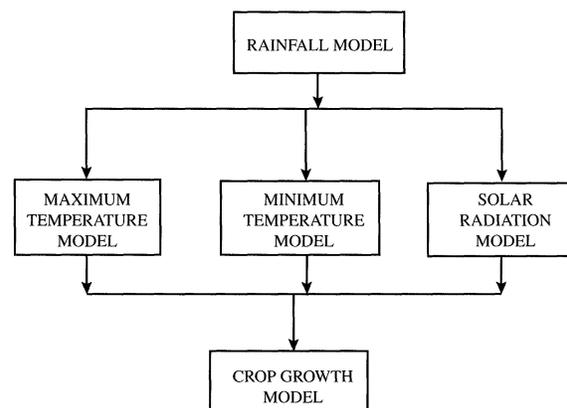


FIGURE 1 A combined weather generator model and crop growth simulation model.

on any day are partitioned between the plant organs that are growing at that time. The phenology of the crop is driven by the accumulation of daily thermal time. Data inputs include daily maximum temperatures, minimum temperatures, rainfall, solar radiation variables, crop and management information and soil data. Temperatures and solar radiation are dependent on rainfall. The model has been used to predict biomass to assess the variability in these prediction estimates, and to identify the key variables that contribute to this prediction imprecision. The structural complexity of the model coupled with a high degree of uncertainty in estimating the values of the input parameters, necessitated the application of the Latin Hypercube Sampling (LHS) uncertainty analysis and the LHS/Partial Rank Correlation Coefficients (PRCC) sensitivity analysis (Iman and Helton, 1988; Hamby, 1994; McKay *et al.*, 1979; Iman and Conover, 1980; Iman *et al.*, 1981).

The program of the stochastic simulation weather model and crop growth simulation model was written in the programming language C for UNIX computers.* Figure 1 shows the flowchart of the model.

THE LATIN HYPERCUBE SAMPLING TECHNIQUE

A LHS techniques are as follows: The range of each input variable is divided into N intervals, and one observation on the input variable is made in each interval using random sampling within each interval. Thus, there are N observations (by stratified sampling) on each of the K input variables. One of the observations on X_1 is randomly selected (each observation is equally likely to be selected), matched with a randomly selected observation on X_2 , and so on through X_k . These collectively constitute X_1^* . One of the remaining observations on X_1^* is then matched at random with one of the remaining observations on X_2 and so on, to get X_2^* . A similar

procedure is followed for $X_3^*, X_4^*, \dots, X_k^*$, exhausts all of the observations and results in a LHS (Iman and Conover, 1980). The N observations of each outcome variable may be used to assess the sensitivity of the outcome variables to the estimation uncertainty in the input parameters. In the LHS scheme, all of the parameters are varied simultaneously and the input.

Parameters are often interdependent; therefore, PRCC can be used to evaluate the statistical relationships. A PRCC indicates the degree of monotonically between a specific input variable and a particular outcome variable (Conover, 1980). The PRCC between the i th input parameter and outcome variable is defined as (Kendall and Stewart, 1979)

$$\Upsilon_{X_j, Y} = b_j * ((1 - R_{X_j}^2)/(1 - R_Y^2))^{1/2} \quad (17)$$

where the value b_j is the standardized rank regression coefficients, the value R_Y^2 is the coefficient of determination from regressing Y on X_1, X_2, X_k and the value $R_{X_j}^2$ is the coefficient of determination from regressing X_j on $Y, X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_k$. Standardized rank regression coefficients are derived from a conditional univariate distribution, while PRCC come from a conditional bivariate distribution. PRCC allow one to judge the unique contribution that an explanatory variable can make. The significance of a nonzero value of $\Upsilon_{X_j, Y}$ is tested by comparing the Kendall PRCC non-parametric statistics (Siegel and Castellan, 1988).

RESULTS AND CONCLUSIONS

Estimate the model parameters from actual records and determine whether the simulated and actual records were in tolerable agreement with respect to important characteristics such as monthly, different agro-climatic seasons, annual rainfall, estimate the conditional probabilities and the probability distribution of the number success. Table I presents the

* C program is available on request.

summary comparison of average monthly rainfall of Barind Tract. The simulated rainfall is greater than the actual annual, seasonal and monthly rainfall but except for month of April, July and August. In kharif season, the total percent of simulated and actual rainfall are 79% and 81% respectively which is closely related to the result (80%) of Rahman and Alam (1996).

The simulated result shows the distribution of number of success in asymptotic normal distribution. The following information was tabulated for

TABLE I Summary comparison of average monthly rainfall of Barind Tract (Nazipur)

	Simulation	Actual
January	2.29	1.52
February	0.88	0.73
March	0	0
April	0.09	0.13
May	5.71	4.97
June	6.34	5.67
July	1.83	2.22
August	2.91	3.64
September	14.08	10.63
October	8.91	8.81
November	0	0
December	0	0
Annual	43.04	38.32
Seasons:		
Rabi	3.17	2.25
%	7	6
Pre-kharif	5.80	5.10
%	13	13
Kharif	34.07	30.97
%	79	81

simulated monthly rainfall: (1) estimated conditional probabilities for Markov chain, (2) probability of success, (3) exact and asymptotic mean and variance of number of success of rainfall. The exact and asymptotic mean and variance of number of success are same for each month of the Barind Tract (Table II). This is especially interesting since the logistic function is approximately linear in f for P values between 0.2 and 0.8 (Neter *et al.*, 1989). The estimated values of P were then placed in the class interval. The simulation results are given in (Table III), which gives estimated values of P against the no-rain/rain values in frequency and the relative frequency of rain. The logistic regression technique allows flexibility in deciding between areas of indicated rain or no rain. Neter *et al.* (1989) suggest a decision rule that no rain is predicted when the estimated value of P is less than 0.5 and rains, when it is greater than 0.5. Table III shows that if estimated P is less than 0.1, it is almost not raining. Basically this type of information is that which gives the advantages of logistic regression over a simple threshold technique.

The rainfall simulation model was tested by using multivariate logistic regression techniques (Rahman, 1999b) and also using maximum likelihood method of Markov chain rainfall model (Rahman, 1999a). The simulated rainfall data was compared with actual rainfall data in different ways. The results of these comparisons indicate that the performance of the model is generally satisfactory.

TABLE II Estimate the conditional probabilities (P_0, P_1) for Markov chain, the probability of success (P), exact and asymptotic mean and variance of the number of success (suggested by daily simulation rainfall data at Barind)

Station parameters	Months											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Nazipur (Potnitala)												
P_0	0.17	0.08	0	0	0.33	0.48	0.23	0.24	0.29	0.28	0	0
P_1	0.42	0.33	0.25	0.25	0.58	0.73	0.48	0.49	0.54	0.53	0.25	0.25
P	0.23	0.11	0	0	0.43	0.65	0.30	0.32	0.39	0.37	0	0
Exact mean	7	3	0	0	13	20	9	10	12	11	0	0
Asymptotic mean	7	3	0	0	13	20	9	10	12	11	0	0
Exact variance	9	4.5	0	0	12.3	11.8	10.5	11.3	12.3	11.6	0	0
Asymptotic variance	9	4.5	0	0	12.3	11.8	10.5	11.3	12.3	11.6	0	0

TABLE III Relationship between the estimated P values using 5 days interval simulation rainfall data and the measured incidence of rain in Nazipur

Estimated values of P	No rain	Rain	Total	Rain/total
0...0.2	201	0	201	0
0.2...0.4	60	15	75	0.2
0.4...0.6	30	20	50	0.4
0.6...0.8	10	15	25	0.6
0.8...1.0	2	8	10	0.8
1.0	0	4	4	1

TABLE IV PRCC calculated from sensitivity analysis

Parameters	PRCC
Photosynthetic active radiation	0.95***
Albedo	0.79***
Water stress factor	0.42***
Leaf area index	-0.33***
Temperature stress factor	-0.14**
Plant water evaporation	-0.11*

The PRCC were used to identify which were the key variables in contributing to the imprecision in predicting the outcomes. The magnitude of the PRCC (Table IV) indicates the importance of uncertainty in estimating the value of the specific variable in contributing to the imprecision. The sign of the PRCC indicates the qualitative relationship between the input variable and the output variable. The order of ranking of the variables is important which is to be considered as the key variables.

The PRCC of four of the seven parameters are statistically significant ($P < 0.001$) and of the remaining two, one is significant at 5% level and another at 1% level (Siegel and Castellan, 1988). The values of the PRCC (Table IV) suggest that the uncertainties in estimating the values of four parameters are the most critical in affecting the prediction imprecision of the above ground biomass production; these four parameters are the photosynthetic active radiation, albedo, water stress factor and leaf area index. The uncertainties in estimating the values of the remaining two parameters (Table IV) are statistically significant, but are lesser importance ($PRCC < 0.33$) in contributing to the prediction for the above ground

biomass production. The sign of the PRCC identifies the specific qualitative relationship between the input and the outcome variables. The positive value of PRCC of the variables implies that when the value of the input variable increases, the above ground biomass will also increase. The above ground biomass decreases as the leaf area index, temperature stress factor and plant water evaporation decreases. The possibilities for improving yields of rice in Bangladesh are through increased sowing density in order to increase leaf area index and capture more solar radiation and through raising harvest index.

This model can easily be used for simulating other cereal crops like wheat, barley, maize, etc. by modifying the values for crop specific parameters in the crop parameter file. The model was also modified for the tropical climate of Bangladesh to analyze some management problems of rice cultivation prevailing there. The results suggested that the rainfall simulation model perform adequately for much application.

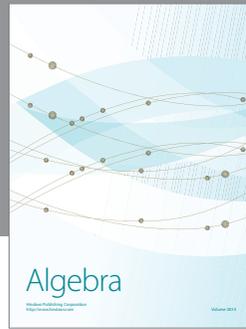
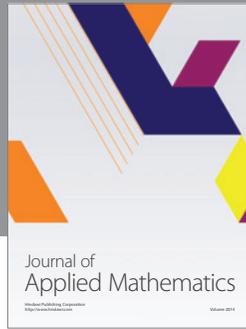
Acknowledgment

The authors are very grateful to Dr. S.P. Rushton, Dr. R. Sanderson, Dr. Mark Shirley and J.J. Barr University of Newcastle upon Tyne, UK, for their invaluable advice and help during suggestions on all aspects of the research. A special regard is for the CLUWRR, University of Newcastle upon Tyne, UK, modelling groups who provide generous support to enrich the research work. He also grateful acknowledges the comments and suggestions of the anonymous reviewer for reviewing the manuscript.

References

- Arnold, J.G., Williams, J.R., Nicks, A.D. and Sammons, N.B. (1990), SWRRB (Simulator for Water Resources in Rural Basins): A Basin Scale Simulation Model for Soil and Water Resources Management, Texas A & M University Press, College Station, pp. 19–23.
- Conover, W.J. (1980), *Practical Nonparametric Statistics*, Second edn., John Wiley & Sons, New York.
- Feller, W. (1957), *An Introduction to Probability Theory and Its Application*, Vol. 1, 3rd edn., John Wiley and Sons, New York.

- Hamby, D.M. (1994), A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, **32**, 135–154.
- Iman, R.L. and Conover, W.J. (1980), Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Journal of Communication Statistics: Theory and Methods*, **A9(17)**, 1749–1842.
- Iman, R.L. and Helton, J.C. (1988), An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Analysis*, **8**, 71–90.
- Iman, R.L., Helton, J.C. and Campbell, J.E. (1981), An approach to sensitivity analysis of computer models. Part I – Introduction input variable selection and preliminary variable assessment. *Journal of Quality Technology*, **13**, 174–183.
- Kendall, M. and Stewart, A. (1979), *The Advanced Theory of Statistics*. Vol., Hafner Publishing Company, New York.
- McKay, M.D., Conover, W.J. and Beckman, R.J. (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.
- Neter, J., Wasserman, W. and Kutner, M.H. (1989), *Applied Linear Regression Models*. Irwin.
- Rahman, M.S. (1999a), A stochastic simulated first-order Markov chain model for daily rainfall at Barind, Bangladesh. *Journal of Interdisciplinary Mathematics*, **2(1)**, 7–32.
- Rahman, M.S. (1999b), Logistic regression estimation of a simulated Markov chain model for daily rainfall in Bangladesh. *Journal of Interdisciplinary Mathematics*, **2(1)**, 33–40.
- Rahman, M.S. and Alam, M.S. (1996), Patterns of rainfall variability and trend in the high Barind region, The Rajshahi University Studies, *Journal of Science, Part-B*, **22–23**, 135–148.
- Rahman, M.S. and Rushton, S.P. (2000), Validation, uncertainty and sensitivity analysis of crop growth simulation model for Barind in Bangladesh. *Journal of Interdisciplinary Mathematics* (in press).
- Sharpley, A.N. and Williams, J.R. (1990) (Eds.), EPIC – Erosion/Productivity Impact Calculator: I. Model Documentation. USDA Technical Bulletin No. 1768, pp. 3–68.
- Siegel, S., and Castellan Jr., N.J. (1988), *Nonparametric Statistics for the Behavioral Sciences*, Second edn., McGraw-Hill Book Company New York.
- Task Force Report (1991), Report of the Task Forces in Bangladesh Development Strategies for the 1990's. Environment Policy, Environment and Development, Vol. 4, Part-III, pp. 63–70.
- Whisler, F.D., Acock, B., Lambert, D.N., Lemmon, H.E., McKinion, J.M. and Reddy, V. (1986), Crop simulation models in agronomic systems. *Advances in Agronomy*, **40**, 141–208.
- Wilkinson, L. (1990), *SYSTAT: The System for Statistics*, SYSTAT Inc.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

