MODELLING THE DAILY RAINFALL OCCURRENCE: A STOCHASTIC APPROACH

Prepared by-

Class roll- SH 101

Session 2002-2003

B.Sc. (Honours) in Applied Statistics.

MODELLING THE DAILY RAINFALL OCCURRENCE: A STOCHASTIC APPROACH

Weather is a complex phenomenon. One of the most important components of weather is Rainfall, which is influenced by many factors. But here the object is to identify the simplest Stochastic model.

Objectives of the Study

- Objective 1. To identify suitable length of period to fit various probability models for daily rainfall data.

- Objective 2. To assess the validity of the assumptions of the various probability models for daily rainfall data.

- Objective 3. To fit various probability models for daily rainfall data and identify the best model by comparison.

Stochastic Models

In the study, two Stochastic models (both involving two parameters) were considered:

Model.1. The Markov Chain model and

Model.2. The Alternating Renewal Model.

Source of Data

The data is provided by Department of Meteorology, Government of People's Republic of Bangladesh. The data used for the current study is of the daily rainfall occurrence data for 19 years (1984-2002) for Khulna stations.

Necessary Definitions

Wet Spell

A wet spell is defined as a sequence of wet days preceded and followed by a dry day. Thus, a wet spell is defined as a continuous run of k days, each of which is credited with a measurable amount of precipitation.

Dry Spell

A dry spell is a sequence of dry days preceded and followed by wet days. Thus, a dry spell is defined as a continuous run of k days, none of which is credited with a measurable amount of precipitation, (k= 1,2, …).

Weather Cycle

A Weather Cycle is defined as the combination of a wet spell with the immediate successive dry spell or a dry spell with the immediate successive wet spell.

Table 1: Summary characteristics of the spells and cycles at Khulna in rainy seasons (1984-2002)

Month	No. of days		No. of Spells	Probability (relative frequency)		True correlation is equals to 0 test	
	Wet	Dry		(wet) p'	p (dry)	Dry-wet pairs	Wet-dry pairs
June	343	227	93	.75331	.623481	0.1388	0.1786
July	445	144	90	.80686	.427673	0.613	0.9754
August	413	176	97	.76225	.448863	0.735	0.1991

Assumptions

Assumption.1. The daily rainfall data follows a first order dependence process.

Likelihood Ratio (LR) c^2 test statistic:

$$\sum_{i_2 \ldots i_m i_{m+1}} \left\{2\log(l_{i_2 \ldots i_m i_{m+1}})\right\} = 2 \sum_{i_1 i_2 \ldots i_m i_{m+1}} n_{i_1 i_2 \ldots i_m i_{m+1}} \log\left(\frac{\hat{P}_{i_1 i_2 \ldots i_m i_{m+1}}}{\hat{P}_{i_2 \ldots i_m i_{m+1}}}\right) \text{ as } l_{i_2 \ldots i_m i_{m+1}} = \prod_{i_1 i_2 \ldots i_m i_{m+1}} \left(\frac{\hat{P}_{i_1 i_2 \ldots i_m i_{m+1}}}{\hat{P}_{i_2 \ldots i_m i_{m+1}}}\right)^{n_{i_1 i_2 \ldots i_m i_{m+1}}} \text{ and } \hat{P}_{jk} = \frac{\sum_{k=1}^{m} n_{ijk}}{\sum_{i=1}^{m}\sum_{k=1}^{m} n_{ijk}} = \frac{\sum_{t=2}^{T} n_{jk}(t)}{\sum_{t=2}^{T} n_j(t-1)}$$

Score c^2 Statistic: $$c^2 = \sum_{j=1}^{m} c_j^2 = \sum_{i,j,k} n_{ij}^* \frac{(\hat{P}_{ijk} - \hat{P}_{jk})^2}{\hat{P}_{jk}}$$

$$\text{As } \hat{P}_{i_1 i_2 \ldots i_m i_{m+1}} = \frac{n_{i_1 i_2 \ldots i_m i_{m+1}}}{n^*_{i_1 i_2 \ldots i_m}} \text{ for } n_{i_1 i_2 \ldots i_m i_{m+1}} = \sum_{t=r}^{T} n_{i_1 i_2 \ldots i_m i_{m+1}}(t) \text{ and } n^*_{i_1 i_2 \ldots i_m} = \sum_{i_{m+1}}^{s} n_{i_1 i_2 \ldots i_m i_{m+1}}$$

AIC and BIC procedure: $$l_{k,r} = l_{k,k+1}\, l_{k+1,k+2} \ldots l_{r-1,r} \text{ for } l_{r-1,r} = \frac{L(\hat{P}'_{i_1 i_2 \ldots i_m i_{m+1}})}{L(\hat{P}_{i_1 i_2 \ldots i_m i_{m+1}})} = \frac{L_0(\max)}{L_1(\max)}$$

$$\Rightarrow {}_k h_r = -2\ln l_{k,r} = -2\ln l_{k,k+1} - 2\ln l_{k+1,k+2} - \ldots - 2\ln l_{r-1,r}$$

$$\text{AIC(k)} = {}_k h_r - 2(s^r - s^k)(s-1) \quad \text{and AIC } (\hat{k}_{AIC}) = \min_{0 \le k < r} AIC(k)$$

$$\text{BIC(k)} = {}_k h_r - (s^r - s^k)(s-1)\ln n, \text{ and BIC } (\hat{k}_{AIC}) = \min_{0 \le k < r} BIC(k)$$

The order of the Markov Chain was identified (by Chi-square Likelihood Ratio test, Chi-square Score test, AIC, BIC Procedure) and it was found that during the rainy season, only each of the June, July and August month follow a first order dependent process, others being higher ordered for data under consideration. Therefore, in the present study, only the months June, July and August were considered.

Table 2: Identification of Order of Markov Chain model for Khulna (1984-2002) for each month of the whole year by various methods

	Likelihood ratio Test	Score test	AIC	BIC
January	1	2	1	1
February	1	-	1	1
March	1	1	1	1
April	1	4	3	1
May	2	2	2	2
June	1	1	2	1
July	1	1	1	1
August	1	1	1	1
September	2	2	2	2
October	2	3	2	2
November	2	-	-	-
December	1	-	-	-

The order of the Markov Chain was identified (by Chi-square Likelihood Ratio test, Chi-square Score test, AIC, BIC Procedure) and it was found that during the rainy season, only each of the June, July and August month follow a first order dependent process, others being higher ordered for data under consideration. Therefore, in the present study, only the months June, July and August were considered.

Table 2: Identification of Order of Markov Chain model for Khulna (1984-2002) for each month of the whole year by various methods

	Likelihood ratio Test	Score test	AIC	BIC
January	1	2	1	1
February	1	-	1	1
March	1	1	1	1
April	1	4	3	1
May	2	2	2	2
June	1	1	2	1
July	1	1	1	1
August	1	1	1	1
September	2	2	2	2
October	2	3	2	2
November	2	-	-	-
December	1	-	-	-

One quick guess may suggest that the combined analysis with only these three months may reveal again a first order Markov Chain process. However, this was checked in the current work, and it was found that if the months are combined, then the daily rainfall data follows a second order Markov Chain Process (confirmed by both AIC and BIC procedures).

Table 3: Values of ${}_k h_r$, AIC, BIC for various order of Markov Chain for Rainy season (Combined period of June, July and August).

k	${}_k h_r$	AIC(k)	BIC(k)
0	153.1494283	27.14943	-317.2229
1	29.1134787	-94.88652	-433.7926
2	8.4076093	-111.5923 #	-439.5660 #
3	2.9167488	-109.08325	-415.1920
4	0.4836618	-95.51634	-357.8953
5	0.4721511	-63.52785	-238.4471

Values with # are respective column minimums

One quick guess may suggest that the combined analysis with only these three months may reveal again a first order Markov Chain process. However, this was checked in the current work, and it was found that if the months are combined, then the daily rainfall data follows a second order Markov Chain Process (confirmed by both AIC and BIC procedures).

Table 3: Values of ${}_k h_r$ **,** AIC, BIC for various order of Markov Chain for Rainy season (Combined period of June, July and August).

k	${}_k h_r$	AIC(k)	BIC(k)
0	153.1494283	27.14943	-317.2229
1	29.1134787	-94.88652	-433.7926
2	8.4076093	-111.5923 #	-439.5660 #
3	2.9167488	-109.08325	-415.1920
4	0.4836618	-95.51634	-357.8953
5	0.4721511	-63.52785	-238.4471

Values with # are respective column minimums

Assumption.2. The spells (especially dry spells) follow a geometric model. In this study, it was shown that all wet spells and dry spells follow geometric distribution well (according to a Chi-square goodness of fit test) for all the months under consideration. The probability of a wet spell to last exactly k days is $(1-p')p'^{k-1} = q'p'^{k-1}$

Figure 1: Distributions of length of Dry and Wet Spell for June, July months

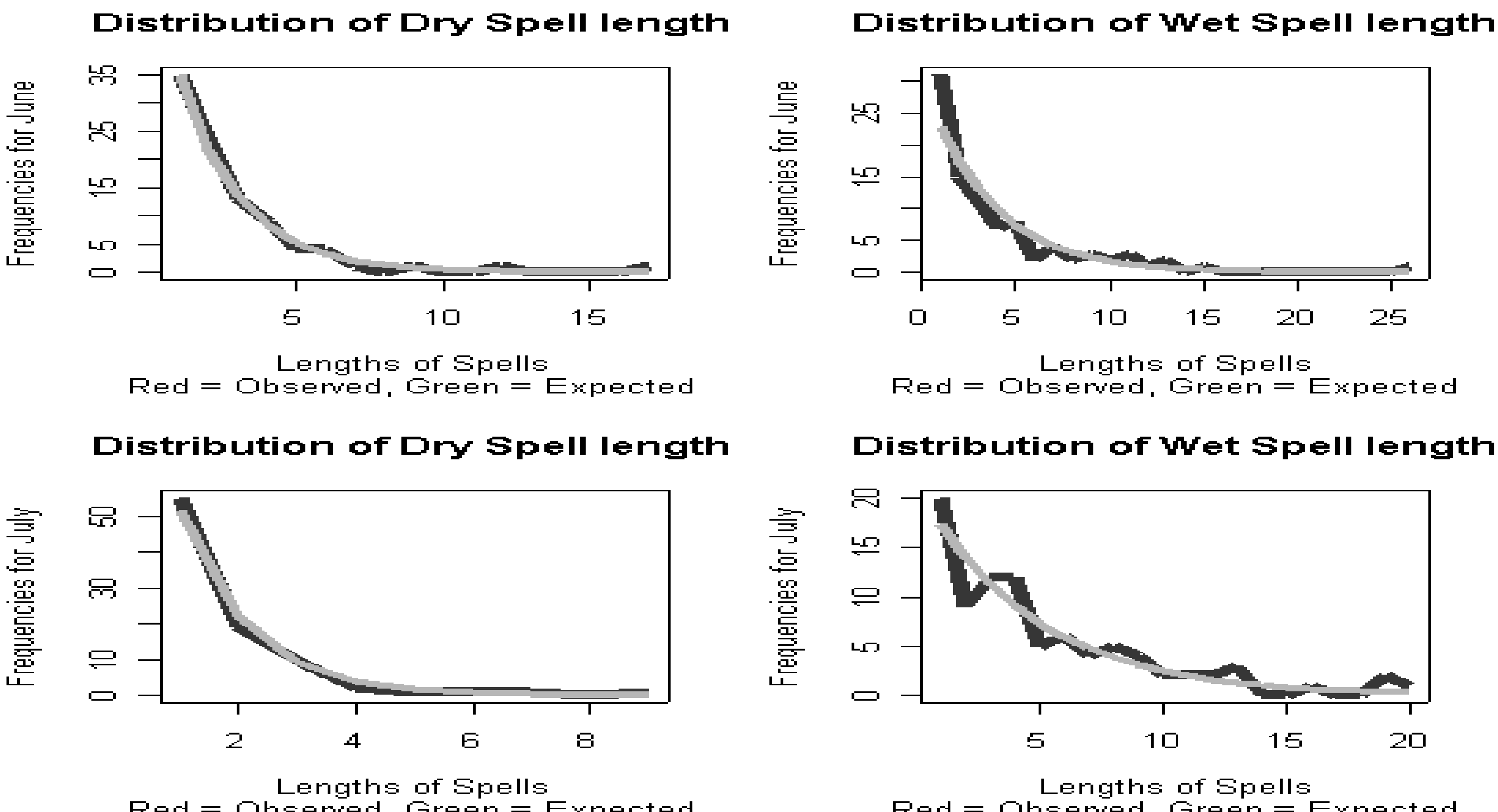


Figure 2: Distributions of length of Dry and Wet Spell for August month

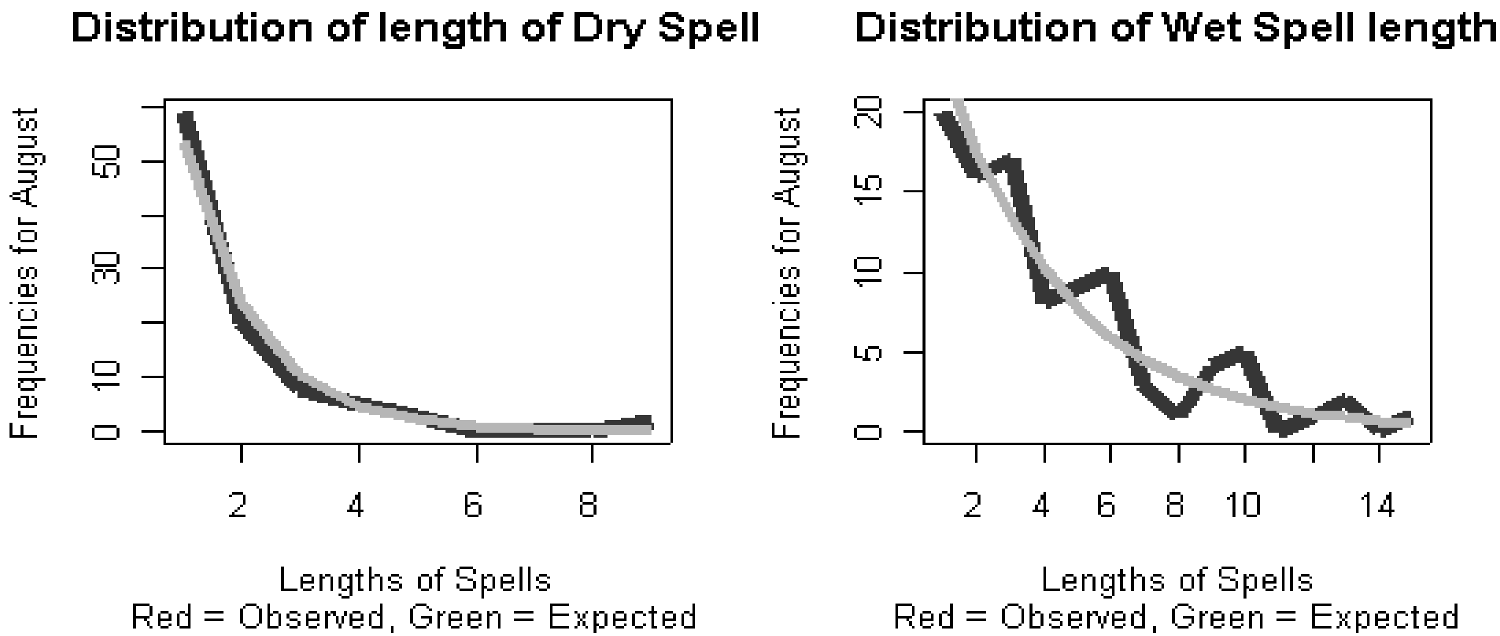


Table 4: P-values for different month of goodness-of-fit with suggested theoretical Geometric distribution

Months	Dry Spells	Wet Spells
June	0.966797	0.524563
July	0.842113	0.834162
August	0.682423	0.500732

Assumption.3. The series is Stationary.

In this study, it was shown that the data were not stationary if the whole June, July, August month data were combined as a rainy season total data. In was also found that the data for separate months also are non-stationary (concluded by AIC, BIC Procedure). Thus, the assumption of Stationarity was violated.

Table 5: Stationarity test by AIC, BIC procedure for Rainy season months (June, July and August) of Khulna (1984-1992 and 1994-2002 periods)

	June	July	August
AIC1	-98.19226	-117.144	-117.8774
AIC2	-115.9061-101.7768 = -217.6829	-121.0005-108.0677 = -229.0682	-118.0293-110.4152 = -228.4445
Conclusion by AIC procedure	**Non-stationary**	**Non-stationary**	**Non-stationary**
BIC1	-317.7081	-385.2542	-385.5634
BIC2	-299.1674-282.1578 = -581.3252	-348.349-333.2028 = -681.5518	-343.1644-330.4135 = -673.5779
Conclusion by BIC procedure	**Non-stationary**	**Non-stationary**	**Non-stationary**

Markov Chain Model

The transition probabilities guarantee a trend in the data from month to month. Therefore, analyses with separate month data instead of combined data are justified.

Table 6: Wet days and dry days classified by rainfall occurrence on preceding day by month; and estimate of conditional probabilities of rainfall occurrence

	Preceding Day	Actual Day			Estimate of Probability	
		Dry	Wet	Total	p_{01}	p_{11}
June	Dry	**135**	**92**	**227**	**0.4052863**	
	Wet	**92**	**250**	**342**		**0.7309942**
July	Dry	**58**	**86**	**144**	**0.5972222**	
	Wet	**86**	**358**	**444**		**0.8063063**
August	Dry	**80**	**96**	**176**	**0.5454545**	
	Wet	**96**	**316**	**412**		**0.7669903**

Markov Chain Model

The transition probabilities guarantee a trend in the data from month to month. Therefore, analyses with separate month data instead of combined data are justified.

Table 6: Wet days and dry days classified by rainfall occurrence on preceding day by month; and estimate of conditional probabilities of rainfall occurrence

	Preceding Day	Actual Day			Estimate of Probability	
		Dry	Wet	Total	p_{01}	p_{11}
June	Dry	**135**	**92**	**227**	**0.4052863**	
	Wet	**92**	**250**	**342**		**0.7309942**
July	Dry	**58**	**86**	**144**	**0.5972222**	
	Wet	**86**	**358**	**444**		**0.8063063**
August	Dry	**80**	**96**	**176**	**0.5454545**	
	Wet	**96**	**316**	**412**		**0.7669903**

Markov Chain Model

The transition probabilities guarantee a trend in the data from month to month. Therefore, analyses with separate month data instead of combined data are justified.

Table 6: Wet days and dry days classified by rainfall occurrence on preceding day by month; and estimate of conditional probabilities of rainfall occurrence

	Preceding Day	Actual Day			Estimate of Probability	
		Dry	Wet	Total	p_{01}	p_{11}
June	Dry	135	92	227	0.4052863	
	Wet	92	250	342		0.7309942
July	Dry	58	86	144	0.5972222	
	Wet	86	358	444		0.8063063
August	Dry	80	96	176	0.5454545	
	Wet	96	316	412		0.7669903

p00 = Prob[dry day | previous day was dry] = Prob[$X_k = 0 \mid X_k\text{-}1 = 0$]

p 01 = Prob[wet day | previous day was dry] = Prob[$X_k = 1 \mid X_k\text{-}1 = 0$]

p 10 = Prob[dry day | previous day was wet] = Prob[$X_k = 0 \mid X_k\text{-}1 = 1$]

p 11 = Prob[wet day | previous day was wet] = Prob[$X_k = 1 \mid X_k\text{-}1 = 1$]

for **P** =

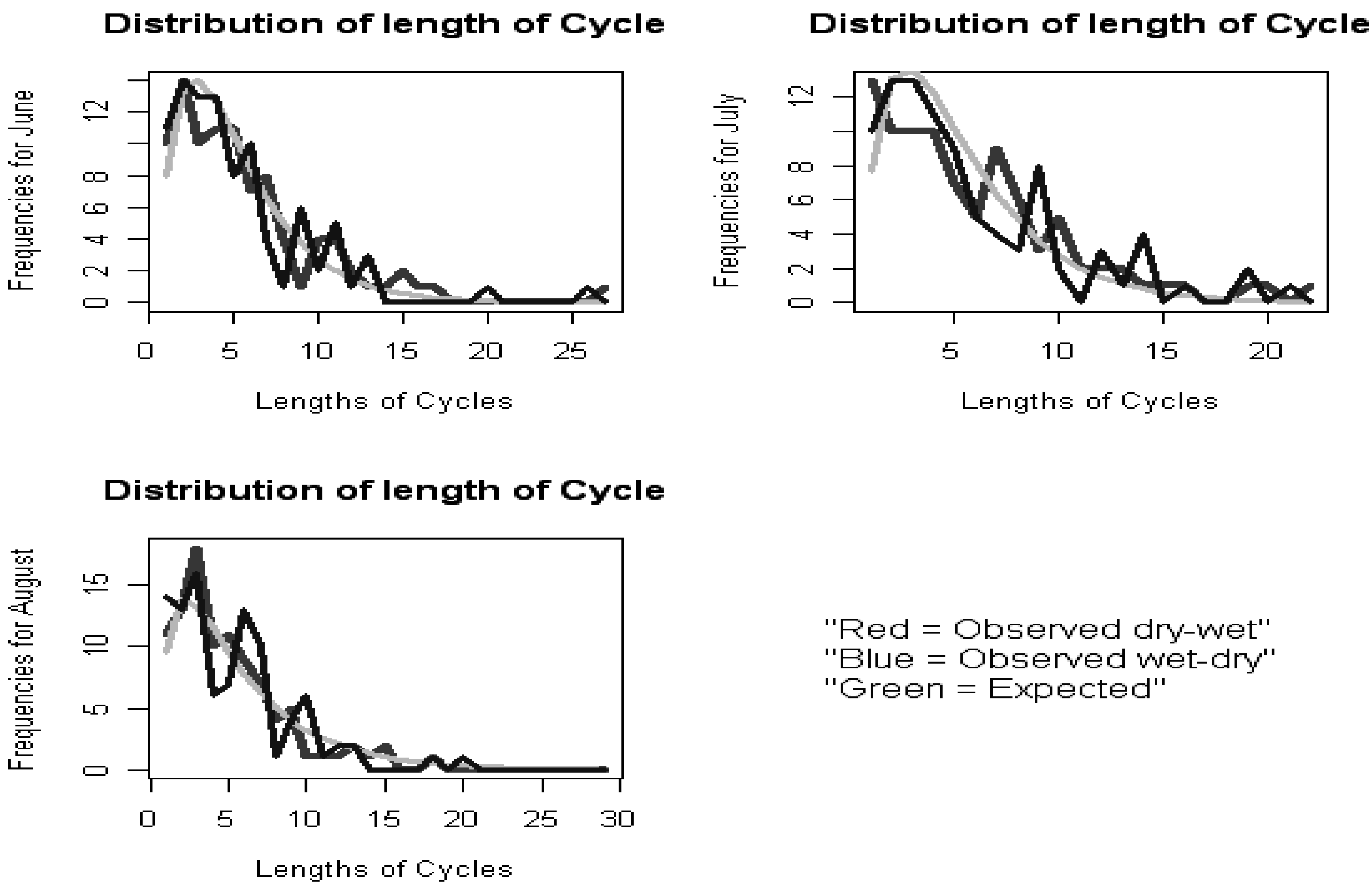
Actual Day

$$\begin{array}{cc|cc} & & 0 & 1 \\ \textit{Previous Day} & 0 & p_{00} & p_{01} \\ & 1 & p_{10} & p_{11} \end{array}$$

The lengths of successive spells are readily seen to be independent and the *probability of a weather cycles of n days* is:

$$p_{01}(1-p_{11})\frac{(1-p_{01})^{n-1}-p_{11}^{n-1}}{1-p_{01}-p_{11}}$$

Figure 3: Length distribution of observed (dry-wet and wet-dry) and expected weather cycles according the Markov Chain Model for June, July and August



Markov Chain model was assessed and it is observed that all weather cycles (both wet-dry and dry-wet) all months fit well to the Markov Chain model.

Table 7: P-values for different month of goodness-of-fit for Markov Chain model for Cycle lengths

Months	Dry-wet Cycle	Wet-dry Cycle
June	0.02647427	0.1267707
July	0.6321408	0.3004013
August	0.9395444	0.0565228

Therefore, we find that the Markov Chain Model fits the weather cycles well for the data under consideration.

Alternating Renewal Model

$\hat{a}$ = -ln{1-(W|D)}

$$\hat{b} = \frac{\hat{a}\{1-(W \mid W)\}}{(W \mid D)\{(W \mid W)-(W \mid D)\}}$$

$$P_n = \frac{DW^{n+1}}{DW^n} = (W \mid DW^n) = \frac{1}{1-Q_n} - b = 1 - b + \frac{(b-1)}{P_{n-1}}$$

$$Q_n = \frac{(W^{n+1})}{W^n} = (W \mid W^n) = 1 - b + \frac{b-a}{Q_{n-1}}$$

$$P_0 = (W \mid DW^0) = (W \mid D) \qquad Q_0 = (W \mid W^0) = W$$

$$W^{n+1} = W^n - (DW^n)$$

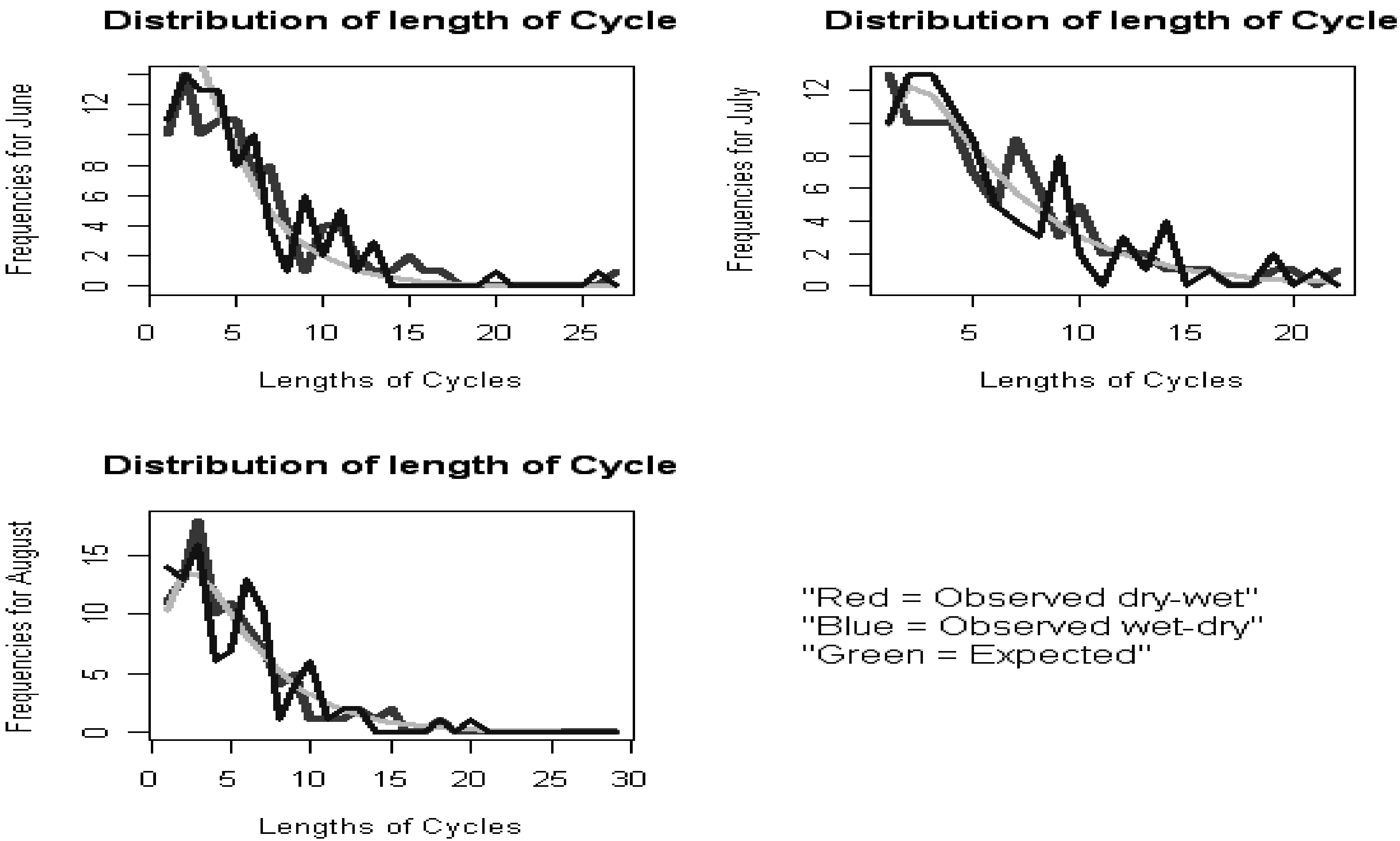
a = $(D \mid d)p_{21}$; b = $p_{12} + p_{21} + (D \mid d) - 1$

$$DW^{n+1} = a(W^n) - b(DW^n)$$

$$(D^{n-1}W \mid WD) = (D \mid d)^{n-1}\{1-(D \mid d)\}$$

$$(W^{n-1}D \mid DW) = (W^{n-1} \mid DW) - (W^n \mid DW) = \frac{\{(DW^n) - (DW^{n+1})\}}{(DW)}$$

Figure 4: Length distribution of observed (dry-wet and wet-dry) and expected weather cycles according the Alternating Renewal Model for June, July and August month



Alternating Renewal model has been assessed and it was observed that the Alternating Renewal model gives good fit to observed data.

Table 8: P-values for different month of goodness-of-fit for Alternating Renewal model for Cycle lengths

Months	Dry-wet Cycle	Wet-dry Cycle
June	**0.487**	**0.511**
July	**0.14661**	**0.05065**
August	**0.66880**	**0.02959**

Therefore, we find that the Alternating Renewal Model fits the weather cycles well for the data under consideration.

Comparison of the Models

Table 9: Comparison between p-values of goodness of fit tests for cycle length distributions of June, July, August months (1984-2002) of Khulna stations

		June	July	August
Markov Chain Model	**Dry-wet Cycle**	**0.026**	**0.63214**	**0.87638**
	Wet-dry Cycle	**0.127**	**0.30040**	**0.06608**
Alternating Renewal Model	**Dry-wet Cycle**	**0.487**	**0.14661**	**0.66880**
	Wet-dry Cycle	**0.511**	**0.05065**	**0.02959**

Comparison of the Models

Table 9: Comparison between p-values of goodness of fit tests for cycle length distributions of June, July, August months (1984-2002) of Khulna stations

		June	July	August
Markov Chain Model	**Dry-wet Cycle**	0.026	0.63214	0.87638
	Wet-dry Cycle	0.127	0.30040	0.06608
Alternating Renewal Model	**Dry-wet Cycle**	0.487	0.14661	0.66880
	Wet-dry Cycle	0.511	0.05065	0.02959

Conclusion

It is concluded that Markov Chain model gives better fit to the data of July and August months, while for June month, Alternating Renewal model is more appropriate (based on higher p-values) for fitting weather cycles for the data under consideration.