A NOTE ON THE USE OF EGGENBERGER–POLYADISTRIBUTION AS AN ALTERNATIVE MODELFOR DRY SPELL ANALYSIS IN NORTHERN ODISA, INDIA

ABSTRACT

In this paper, Eggenberger–Polya distribution has been preferred as an alternative to a 2-state Markov Chain model for describing persistence behaviour of the sequence of dry days (dry spell) for Northern Odisha, India during south-west monsoon season. Attempts have been made to compare performance of the two competing models with reference to their strength to fit the empirical distribution of dry spell length under two goodness-of-fit criteria.

***Key Words*:**Chi-Squared test, dry spell, Eggenberger–Polyamodel, goodness-of-fit, Kolmogorov-Smirnov test, Markov Chain model.

INTRODUCTION

Rainfall is the most concerned climatic factor and has multiple effects on the society, environment, agriculture, and other water-related issues. Wet and dry spells (or runs) being two main physical characteristics of the rainfall occurrences,deficiency in the total seasonal rainfall volume in a location depends mainly on the distribution of dry spells (sequences of dry days between two wet days).The intensity and duration of drought conditions are also directly related to the number of dry spells. Long dry spells can significantly influence not only agriculture and hydroelectric power generation but also other sectors such as hydrology, meteorology, fisheries, health, ecology, environment etc. Dry periods do not persist in a regular manner but vary randomly over space and time.Thus, model-based scientific investigationsof the dry spell characteristics (length and frequency) provide useful information to solve various water management problems including crop failure due to deficiency of rainfall.

Several studies have recognized Markov Chain (MC)model as an appropriate model to explore stochastic behaviours of the distributions of wet or dry spells starting from the pioneering work of Gabriel and Neumann (1962). However, some authors[*cf*., Feyerherm and Bark (1967), Chang *et al*. (1984), Foufoula-Georgiou and Georgakakos (1986), Wantuch *et al*. (2000) among others] have established that these models only approximately replicate some statistical properties of wet and dry spells of the daily precipitation but usually overestimate very short dry sequences and underestimate very long dry sequences, and do not reproduce long term persistence.Wiser (1965), Gates and Tong (1976), and Green (1964, 1967 and 1970) have also confirmed that there are several data sets which are not properly fitted by the MC model. These negative aspects of the MC model have encouraged researchers to use some alternative probabilistic models to explore variability of dry spell characteristics including the persistence and the pattern of dry periods.Eggenberger–Polya(EP) distribution is one among such models considered by Berger and Goossens (1983), Goossens and Berger (1984), Giuseppe *et al*. (2005), Kamar and Rao (2004), Epifani *et al*. (2004), Deka *et al*. (2010), Sukla*et al*. (2012)along with others. In these papers, although theperformance of the EP model over others was evaluated under different climatic conditions, has not yet been received noticeable attention in the literature.

North Odisha is comprised of five districts of Odisha. But, for analysing theiragro climatic features, the districts are considered under three agro climatic zones: North Western Plateau, North Central Plateau and North Eastern Coastal Plain. The three zones are more or less similar in respect of mean annual rainfall, mean maximum summer and winter temperature, and cropping pattern, crop planning and production. Both the plateau zones are identical in connection with their climatic and topographic features, ecologies, soil type, and irrigation facilities whereas on these grounds North Eastern Coastal Plain is slightly different. North Odisha receives rainfall from the south-west monsoon with an average annual rainfall of 1580 mm and the total rainy day in a year ranging from 50 to 60 days. The most pre-dominant crop in this region is paddy covering about 80% of the total cultivated area. Besides paddy, some other cash crops and short duration cropsare also grown in the substantial areas during the monsoon season. Irregularities in the distribution of rainfall and its variation on spatiotemporal basisarethe sources of great tension to the farming activities as the agriculture is mostly rain fed with non-availability of assured irrigation facilities. In some years due todeficient rainfall, drought causes the major abiotic stress that affects the crop production.Hence, an appropriate modelling of the occurrence pattern of the sequence of dry days during monsoon period is therefore extremely useful for planning agricultural activities and managing the associated water supply systems at various locations of the study domain.

The objective of the present study is to appraiseEP model as an alternative to a 2-state MC model to represent the frequency distribution of dry spell length for the Northern Odisha. We evaluated and compared fitting power of the two alternative models working with daily rainfall data for 35 years during monsoon season with the aid of two goodness-of-fit (GOF) tests: Chi-Squared (CS) test and Kolmogorov-Smirnov (KS) test.According to the Indian Meteorological Department, the criterion of less than one millimetre rainfall was used for tabulating a dry day.

DATA AND METHODOLOGY

Source and Nature of Data

Quantitative data on the daily rainfall volume (in mm) at all three meteorological stations: Balasore, Baripada and Kendujhar of the Northern Odisha are utilized for this study. The relevant data for 35 years (1984-2018), were collected from the Meteorological Centre, Bhubaneswar, Odisha. The south-western monsoon season *i.e.*, *kharif* season (June – September) was taken as the study period because the study region normally receives about 80% of its total annual rainfall during this season which also coincides with the growth season of paddy crop, the most important cash crop. Hence, we analysed daily rainfall data from 1st June to 30th September (122 days). Figures on the daily rainfall amount for the whole study area were decidedon the average of such figures of the three recording stations in order to find proper representative figures with reduced errors that arising due to random causes [*cf.*, Mooley and Parthasarathy (1984)].

Markov Chain Probability Model

The basic assumption behind the considered 2-state MCmodel is that the occurrence of a dry or a wet day depends on the weather condition of the previous day. The parameters of this probability model are two conditional probabilities defined by $p_o =$ the probability of a wet day given that the previous day was dry, and $1-p_1$ =the probability of a dry day given that the previous day was wet. This means that$p_1 = \Pr\{W/W\}$, $1-p_1 = \Pr\{D/W\}$, $p_0 = \Pr\{W/D\}$ and $1-p_0 = \Pr\{D/D\}$, where W and D stand for wet and dry days respectively.

It is well known that under the Markovian preconditions, the random variable $X =$ the dry spell length has a geometric distribution. The probability of a dry spell of length x is therefore given by

$$P(X = x) = p_0(1-p_0)^{x-1}, x = 0, 1, 2, \dots\dots$$

The expected length of dry spell is given as $E(X) = \frac{1}{p_0}$,and the maximum likelihood estimate of p_0 is given by

$$\hat{p}_0 = \frac{n\{W/D\}}{n\{W/D\}+n\{D/D\}},$$

where $n\{W/D\}$ and $n\{D/D\}$ are the observed frequencies of the respective conditional events [see, for example, Cox and Miller (1967)].

Eggenberger–Polya Probability Model

According to Berger and Goossens (1983), the probability of a dry spell of length x under this model is given by

$$P(X = x) = \frac{h+(x-2)d}{(x-1)(1+d)} P(x-1), x = 2, 3, \dots\dots$$

such that $P(x = 1) = (1+d)^{-\frac{h}{d}}$ and $E(X) = h = \overline{x} - 1$,where$\overline{x}$ is mean length of dry spells, andthe parameter d represents the degree of influence of an event on the following event computed by $d = \frac{s^2}{h} - 1$ withs^2 as the variance of dryspell.

Tests for Model Selection

The CS and KS GOF test procedures used to assess the performance of the models are outlined as follows:

The statistic used for the CS test of GOF is defined by

$$\chi^2 = \sum_x \frac{(O_x - E_x)^2}{E_x},$$

where O_x and E_x are respectively the observed and expected (predicted) frequencies corresponding to the spell length x. On the other hand, for the KS test of GOF, the test statistic defined by

$$D_{max} = \max_x |OR_x - ER_x|,$$

is used, where OR_x and ER_x are respectively the observed and expected relative cumulative frequencies corresponding to the spell length x.

DATA ANALYSIS, RESULTS AND DISCUSSIONS

On the basis of rainfall quantity, all 122 days from the 1st June to the 30th September for 35 years are classified as dry or wet. This classification shows the occurrences of alternative sequences of dry and wet days and accordinglyobserved (empirical) frequency distribution of spell lengthsareconstructed.

Findings on Descriptive Statistics

Table 1 presentsprobabilities of occurrences of dry and wet days, number of dry and wet spells along with some important statistical descriptors of their length distributions,andmaximum spell lengths as computed from the empirical data set. In the light of the tabular results, some major findings are as follows:

Table 1: Statistical Descriptors of Spell Lengths

Day	Number	Probability of Occurrence	Spell Length Descriptors					
			Spell	Number	Mean	SD	CV	Maximum Length
Dry	2397	0.5614	Dry	787	2.7	2.3	85.19	21
Wet	1873	0.4386	Wet	807	2.3	1.8	78.26	15

(i) More chance of occurrence of a dry day than a wet day confirms that the study area is likely to enjoy more dry days than wet days.

(ii) Comparisons of mean and maximum spell lengths show that dry spells have longer duration than wet spells causing higher persistence level of dry conditions.

(iii) Figures on the standard deviation (SD)and coefficient variation (CV=$\frac{\text{SD}}{\text{mean}}$×100) clarify that occurrence of dry spells is more fluctuating and more inconsistent than wet spells.

These findings clearly indicate that dry spell has definitely adverse impacts on the water management system of the Northern Odisha region during monsoon season.

Model Comparison

Previously said 122 days are further classified into four classes according to the occurrence of four conditional eventsW/D , D/W , W/D and D/Dsuch that 1st June is classified basing on the weather condition of 31st May. After counting class frequencies of different classes, the parameters of the MC model are estimated as p_o =0.3408,$1-p_1$ =0.4348 and $E(X) = 2.9345$. Then, the parametersof the EP model *i.e.*, $d, P(x = 1)$and $h = E(X)$are estimated as 0.3024, 0.6226 and 2.7354 respectively from the observed frequency distribution of the dry spell length.

Table 2: Observed and Expected Frequency Distributions of Dry Spell Lengths

Spell Length (x)	OF	EF		Spell Length (x)	OF	EF	
		MC	EP			MC	EP
1	1236	1073	1250	12	13	11	10
2	688	707	685	13	7	7	7
3	449	466	422	14	12	5	5
4	265	307	270	15	3	3	3

5	166	203	175	16	2	2	2
6	117	134	115	17	1	1	1
7	76	88	76	18	0	1	1
8	55	58	51	19	0	1	1
9	24	38	34	20	1	0	1
10	22	25	23	21	1	0	0
11	9	17	15	Total	3147	3147	3147

Observed frequencies (OF) and corresponding expected frequencies (EF) of the dry spell length varying from 1 to 21 days, according to the competing models,are provided in Table 2. Table 3 shows calculated values of the CS and KS test statistics as well as their 5% critical values for rejecting/accepting the null hypothesis of no disagreement between the theoretical (expected) and observed frequency distributions of dry spell length.

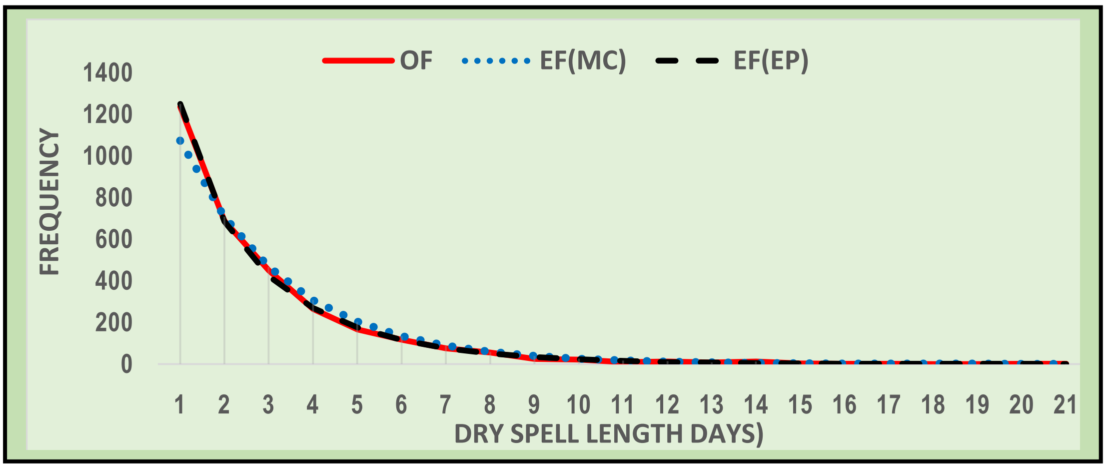
Table 3: GOF Statistics Values and 5% Critical Values

Model	CS Statistic		KS Statistic	
	Calculated Value	Critical Value	Calculated Value	Critical Value
MC	61.776	23.685	0.0518	0.0242
EP	18.997		0.0051	

From the entries of Table 3, we see that application of MC model to the counts of sequences of dry days gives significant CS and KS statistics values. Hence, our analysis shows that the MC model has failed to provide a satisfactory fit to the empirical distribution of dry spell length. On the other hand, EP model yields very insignificant CS and KS statistics values. This means that this model is rather successful in giving a very close representation of the frequency of sequences of dry days reported here. The results of our analysis further confirm that EP model is more efficient (has more fitting ability) than the MC model in replicating dry spell for the period of 35 years.

In Figure 1, the curves of the observed and expected frequencies of dry spell lengths, in respect of the candidate models, are drawn to illustrate the agreement between them graphically. It appears that EF curve of the EP model gains appreciable closeness to the observed counterpart compared to the MC model. This graphical finding supports our empirical finding in the sense that the EP model provides a better fit to our rainfall data.

Figure 1: Observed and Expected Frequency Curves



CONCLUSIONS

Our analysis for the consideration ofEP model as an alternative to the2-state MC model shows that the later model had not only proved to be unsatisfactory but also had proved to be inferior to the former one in relation to the data at our disposal.Both empirical and graphical findings signifythat the distribution of dry spell length of Northern Odisha region during monsoon season can proficiently be approximated by anEP model with an average of 2.7340 ≅ 3 days.This convincing finding may be useful for predicting probability of occurrence of a particular dry spell or a critical dry spell. Prior information on this probability can be very helpful to the crop planers or agronomists for taking precautionary measures during the growth period of the paddy and to develop a system to monitor drought or in planning of water resource programs in the study region.

In this study, although MC model statistically fails to fit our dataset under CS and KS GOF test procedures, the model cannot be rightly regarded as an inappropriate one. Because, from the graph it seems thatonly few EF points corresponding to spell lengths less than 6 days to have slight divergence from theOF curve whereas considerable amount of such points fall either along or nearest the said curve.The probable reason for this is that the MC model under the circumstance overestimates very short dry spells and underestimates very long dry spells as it is location specific and temporal specific [*cf*., Mathugama and Peiris (2011)]. On the other hand, from the discovery of EP distribution as an ideal model, it seems apparent that other probability models may also be more successful. Hence, in view of the saidconsiderations, we stress that more detailed and exhaustive investigations may be carried out with the help of other models using other GOF test procedures.

REFERENCES

Berger, A. and Goossens, C. (1983). Persistence of wet and dry spell at Uccle (Belgium). *Journal of Climatology*, **3**, 21-34.

Chang, T. J., Kavvas, M. L. and Delleur, J. W. (1984). Modelling of sequences of wet and dry days by binary discrete autoregressive moving average processes. *Journal of Climate and Applied Meteorology*, **23**, 1367-1378.

Cox, D. R. and Miller, H. D. (1967). *The Theory of Stochastic Process*. Wiley: New York.

Deka, S., Borah, M. and Kakaty, S.C. (2010). Statistical modelling of wet and dry spell frequencies over North-East India. *Journal of Applied and Natural Science*, **2(1)**, 42-47.

Epifani C., Esposito S. and Vento D. (2004). Persistence of wet and dry spells in Italy. First results in Milano from 1858 to 2000. *Proceedings from the 14th International Conference on Clouds and Precipitation 2004*, Bolobna, 18-24.

Feyerherm, A. M. and Berk, L. D. (1967). Goodness-of-fit of a Markov Chain model for sequences of wet and dry days. *Journal of Applied Meteorology*, **6**, 770-773.

Foufoula-Georgion, E. and Lettemaier, D. P. (1986). Continuous-time versus discrete-time point process models for rainfall occurrence series. *Water Resources Research*, **22**, 531-542.

Gabriel, K. R. and Neumann, J. (1962). A Markov Chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of Royal Meteorological Society*, **88**, 90-95.

Gates, P. and Tong, H. (1976). On Markov Chain modelling to some weather data. *Journal of Applied Meteorology*, **15**, 1145-1151.

Giuseppe, E. D., Vento, D. Epifani, C. and Esposito, S. (2005). Analysis of dry and wet spells from 1870 to 2000 in four Italian sites. *Geophysical Research Abstracts*, Vol. **7**, 07712. SRef-ID: 1607-7962/gra/EGU05-A-07712, © European Geosciences Union 2005.

Goossens, C. and Berger, A. (1984). Persistence of dry and wet spells in Belgium.*Archives for Meteorology, Geophysics, and Bioclimatology*, **B34**, 243–256.

Green, J. R. (1964). A model for rainfall occurrence. *Journal of the Royal Statistical Society*, **B26**, 345-353.

Green, J. R. (1967). A modified model for rainfall occurrence. *Journal of the Royall Statistical Society*, **B29**, 151-153.

Green, J. R. (1970). A generalized probability model for sequences of wet and dry days. *Monthly Weather Review*, **98**, 238-241.

Kamar, K. and Rao, T. V. (2004). Dry and wet spells at Campina Granade. *Revista Brasileira de Meteorologia*, **20**, 71-74.

Mathugama, S. C. and Peiris, T. S. G. (2011). Critical evaluation of dry spell research. *International Journal of Basic and Applied Sciences*, **11**, 153-160.

Mooley, D.A., Parthasarathy, B. (1984). Fluctuations in All-India summer monsoon rainfall during 1871-1978. *Climatic Change*, **6**, 287-301.

Sukla, M. K., Mangaraj, A. K., Sahoo, L. N. and Sethy, K. M. (2012). A comparative study of three models for the distribution of wet and dry spells in the Mahanadi Delta. *New York Science Journal*, **5 (11)**, 54-61.

Wantuch, I. D., Mika, J. and Szeidi, L. (2000). Modelling wet and dry spells with mixture distributions. *Meteorology and Atmospheric Physics*, **73**, 1436-5065.

Wiser, E. H. (1965). Modified Markov probability models of sequences of precipitation events. *Monthly Weather Review*, **93**, 511-516.