Using of Eggenberger–Polya Distribution as an Alternative Model for Dry Spell Analysis in Northern Odisha, India

Abstract

In this paper, Eggenberger–Polya distribution has been preferred as an alternative to a 2-state Markov Chain model for characterizing persistence behaviour of dry spell for Northern Odisha, India during south-west monsoon season. An attempt has 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–Polya model, goodness-of-fit, Kolmogorov-Smirnov test, Markov Chain model.

1 Introduction

Rainfall is the most concerned climatic factor and has multiple effects on the society, environment, agriculture, and other water-related issues. It is well known that the deficiency in the seasonal rainfall totals at any location mainly influenced by the dissipation of dry spells or runs (periods of consecutive dry days). The intensity and duration of drought conditions are directly related to the number of days without rainfall. Longer persistence of dry periods significantly influence agriculture, hydroelectric power generation, environment and many other essential socio-economic sectors of human activities. For more details on various causes of dry spells and their disastrous effects on diversified areas, we may refer to [1], [2], [3], [4], [5] and [6] among others. The impact of climate change and variability in rainfall patterns always lead to irregularities in the occurrence framework of dry spell durations on spatial and temporal scales. Thus, model-based scientific investigations of 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 and dry spell lengths after the classic paper of [7]. In this context we may also refer to [6], [8], [9], [10], [11], [12], [13] and [14]. Some researchers are in the opinion that these rainfall models, although statistically fit to the observed (actual or empirical) frequency distributions of spell lengths, have a relatively short memory so that they may limit the model's ability to reproduce adequately long/short dry spells as well as inter-annual variability, see for example, [2], [15], [16], [17], [18] and [19]. These limitations have encouraged researchers to use some efficient alternatives to the MC probabilistic models to analyse 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 [20], [21], [22], [23], [24], [25] and [26] along with others. In these papers, performance of the EP model over others was evaluated under different climatic conditions. But, there is a lack of comprehensive investigation on short-term to long-term variations in the duration of dry spells using historical climatic data.

North Odisha is comprised of five districts of Odisha. But, for analysing their agro 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. The total annual rainfall in the North Odisha is about 1580 mm and the total rainy (wet) days enjoyed by the region annually ranges from 50 to 60 days. Here the most important crop is paddy that covers about 80% of the total cultivated areas. Besides paddy, some other cash crops and short duration crops are also grown in the substantial areas during the monsoon season. Irregularities in the distribution of rainfall and its variation on spatiotemporal basis are the 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, especially within the crop growing period, due to rainfall deficiency drought causes the major abiotic stress that drastically reduces the crop production. Hence, model-based exploration of the variability of dry spells and frequency of their occurrence is extremely useful for effective crop planning, water resources management and subsequent drought monitoring at the study site.

The objective of the present study is to appraise EP model as an alternative to a 2-state MC model to study stochastic behaviour of the distributional system of dry spell length for 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 Indian Meteorological Department, the criterion of less than one millimetre rainfall was used for tabulating a dry day.

2. Methodology

2.1 Source and Nature of Data

Climatic data on daily rainfall volume (in mm) for 35 years (1984-2018) at all three meteorological stations: Balasore, Baripada and Kendujhar of Northern Odisha are utilized for this study. The data were picked up 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. This is also agreed with the growing season of paddy crop, the predominated cash crop in the tract. Hence, we analysed daily rainfall data from 1st June to 30th September (122 days). Daily rainfall amounts for the overall study domain were decided on the average of such amounts of the three recording stations in order to find proper representative figures with reduced errors that arising due to random causes [27].

2.2 Markov Chain Probability Model

As desired, assuming that the weather condition of a particular day depends only on the weather condition of the just preceding day, the two parameters of the 2-state MC model are defined by $p_o =$ conditional probability of a wet day following a dry day, and $1 - p_1 =$ conditional probability of a dry day following a wet day. This means that $p_0 = \Pr\{W/D\}$, $p_1 = \Pr\{W/W\}, 1 - p_0 = \Pr\{D/D\}$ and $1 - p_1 = \Pr\{D/W\}$, 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 so that occurrence probability of a dry spell of length x is

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

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 frequencies corresponding to the conditional events $\{W/D\}$ and $\{D/D\}$, see for example, [28].

2.3 Eggenberger–Polya Probability Model

Following Berger and Goossens (1983), under EP probability model we have

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

such that $P(x = 1) = (1 + d)^{-\frac{h}{d}}$, $E(X) = h = \overline{x} - 1$ and $d = \frac{s^2}{h} - 1$, where $\overline{x}$ and s^2 are respectively mean and variance of the dry spell length.

2.4 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 given by

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

where O_x and E_x are defined as the observed and expected (predicted) frequencies in favour of the dry spell of length x. On the other hand, the KS test statistic is given by

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

where OR_x and ER_x are respectively observed and expected relative cumulative frequencies for the dry spell of length x.

3. Results and Discussions

After identifying dry and wet days, and their sequences according to our concepts, observed or actual frequency (OF) distributions of spell lengths are constructed.

3.1 Findings on Descriptive Statistics

Occurrence probabilities of dry and wet days, and number of dry and wet spells along with some descriptive statistics of their lengths, as computed from the empirical data set, are presented in Table 1. In the light of the tabular results, some major findings are as follows:

Table 1. Descriptive statistics 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) Less chance of experiencing a wet day than a dry 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=$\frac{\text{SD}}{\text{mean}}\times 100$) clarify that occurrence of dry spells is more fluctuating and more inconsistent than wet spells.

From the said outcomes, it seems that dry spells can have negative impacts on the various socio- economic activities including water security and food production in Northern Odisha during monsoon season.

3.2 Model Comparison

Previously said 122 days are categorized into four classes on the happening of the conditional events $\{W/D\}, \{D/W\}, \{W/D\}$ and $\{D/D\}$. But, decision on the 1st June has been taken according to the weather status of the 31st May. After tabulating class frequencies, the MC model parameters are estimated as $p_o = 0.3408$, $1 - p_1 = 0.4348$ and $E(X) = 2.9345$. Then, the parameters of the EP model *i.e.*, $d, P(x = 1)$ and $h = E(X)$ are estimated as 0.3024, 0.6226 and 2.7354 respectively on the basis of the actual frequency distribution of the dry spell length.

OFs and respective expected frequencies (EFs) for varying dry spell length, as claimed by the candidate models, are given in Table 2. Table 3 includes evaluated values of the test statistics besides their 5% significant values for rejecting/accepting the null hypothesis of no disagreement between the expected (theoretical) and observed frequency distributions of 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

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	

The calculated significant values of the CS and KS test statistics in favour of the MC model (Table 3) illustrates that the model has failed to serve as a well fitted model for the actual (observed) distribution of dry spell length. From the point of view of very insignificant test statistics values for the alternating EP model, it seems that the model necessarily provides a very good fit to the said distribution. Findings of the conducted analysis also demonstrate that EP model is more precise (has more fitting ability) than the MC model in replicating dry spell for the period of 35 years relating to our daily rainfall dataset.

OF and EF curves in respect of the competing models are drawn in Figure 1 to illustrate the agreement between them graphically. It appears that EF curve of the EP model gains noticeable nearness to its observed counterpart in comparison with that of the MC model. This means the graphical result is in agreement with empirical (statistical) result that the EP model contributes a finer fit to our rainfall data.

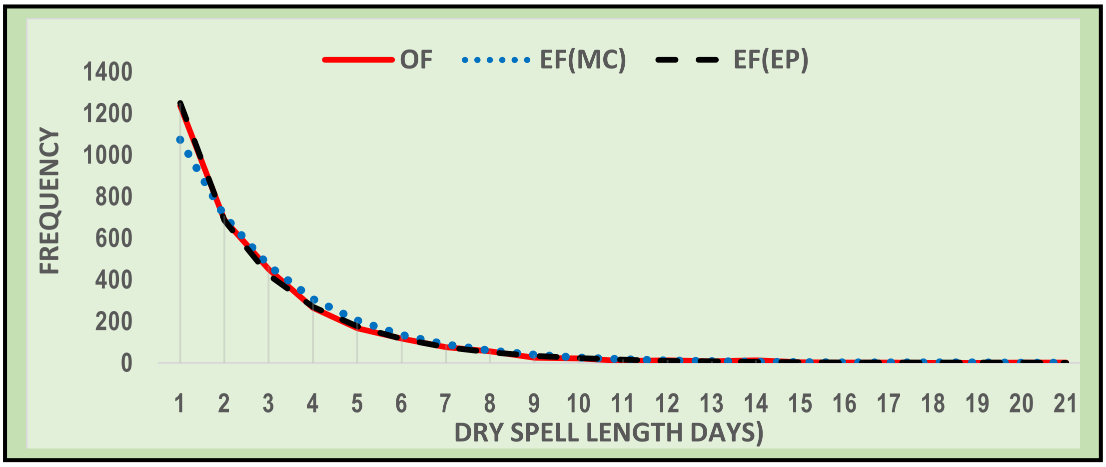

Fig. 1. Observed and expected frequency curves

4. Conclusions

Our analysis for the consideration of EP model as an alternative to the 2-state MC model shows that the later model is unsatisfactory and less efficient than the former one relating to the data at our disposal. Both empirical and graphical findings signify that the distribution of dry spell length of Northern Odisha region during monsoon season can proficiently be approximated by an EP model with an average of $2.7340 \cong 3$ days. This convincing finding can be skilfully used to assess the predictability of dry spell patterns. Hence, it is possible to develop predictable strategies in the study region for acquiring preventive initiatives during the monsoon season or growing season of crops for monitoring and managing droughts and dry spell regimes, and solving various problems related to water resources.

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 that only few EF points corresponding to spell lengths less than 6 days to have slight divergence from the OF 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 [29]. On the other hand, discovery of EP model as a good fitted model gives an indication for the identification of other alternative models in the present context. Hence, in view of the said considerations, we recommend for further analysis using other models and other GOF test procedures.

Acknowledgement

Authors are grateful to the Editor and anonymous Reviewers for their useful comments that have helped to improve the paper.

References

[1] Sanchi ID, Alhassan YJ, Sabo YA, Hamid BJ. Critical review of the causes and effects of dry spell in 2021 rainy season in Danko Wasagu Local Government, Kebbi State, Nigeria. Cross Current International Journal of Agriculture and Veterinary Sciences. 2021; 3(8): 66-75. Available: 10.36344/ccijavs.2021.v03i08.001.

[2] Agbazo MN, Adechinan JA, N'gobi GK, Bessou J. Analysis and predictability of dry spell lengths observed in Synoptic Stations of Benin Republic (West Africa). American Journal of Climate Change. 2021; 10: 597-618. Available: https://www.scirp.org/journal/ajcc.

[3] Usman AA, Aji Mamman MB. Spatio-temporal analysis of dry spell for agricultural decision support in North-Central Nigeria. Environmental and Earth Sciences Research Journal. 2022; 9(1): 1-7. Available: https://doi.org/10.18280/eesrj.090101.

[4] Wainwright CM, Allan RP, Black E. Consistent trends in dry spell length in recent observations and future projections. Geophysical Research Letters. 2022; 49(12). Available: https://doi. org/10.1029/2021GL097231.

[5] Chimimba EG, Ngongondo C, Li C, Minoungou B, Monjerezi M, Eneya L. Characterisation of dry spells for agricultural applications in Malawi. SN Applied Sciences. 2023; 5: 199. Available: https://doi.org/10.1007/s42452-023-05413-9.

[6] Daniel S, Mengistu MG, Olivier C, Clulow AD. Analysis of dry-spells in the western maize-growing areas of South Africa. Water. 2023; 15: 1056. Available: https://doi.org/10.3390/w15061056.

[7] Gabriel KR, Neumann J. A Markov Chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of Royal Meteorological Society. 1962; 88: 90-95.

[8] Fischer B, Mul M, Savenije HHG. Determining spatial variability of dry spells: A Markov-based method applied to the Makanya catchment, Tanzania. Hydrology and Earth System Sciences. 2013; 17(6): 2161-2170. Available: https://doi. org/10.5194/hess-17-2161-2013.

[9] Dabral PP, Purkayastha K, Aram A. Dry and wet-spell probability by Markov Chain model—A case study of North Lakhimpur (Assam), India. International Journal of Agricultural and Biological Engineering. 2014); 7: 8–13.

[10] Dabral PP, Dada M, Odi H. Dry and wet spell probability analysis by Markov Chain model for Kohima (Nagaland), India. AgricEngInt: CIGR Journal Open Access. 2019; 21(4): 43-47. Available: http://www.cigrjournal.org.

[11] Ray M, Biswasi S, Sahoo KC, Patro HA. A Markov Chain approach for wet and dry-spell and probability analysis. International Journal of Current Microbiology and Applied Sciences. 2018; 6: 1005–1013. Available: http://www.ijcmas.com.

[12] Tyubee BT, Iwan MT. A Markov Chain analysis of wet and dry spell for agricultural crop planning in the Middle Belt Region of Nigeria. Journal of Agriculture and Environmental Sciences. 2019; 8 (2): 132-147. Available: https://doi.org/10.15640/jaes.v8n2a16.

[13] Pawar PS, Khodke UM, Waikar AU. Dry and wet spell probability by Markov Chain model for agricultural planning at Parbhani. International Journal of Bio-resource and Stress Management. 2019; 10(3): 233-24. Available: https://doi.org/10.54302/mausam.v70i3.282.

[14] Basse J, Camara M, Diba I Diedhiou A. Probability of dry and wet spells over West Africa during the summer monsoon season. Scientific Research and Essays. 2021; 16(3): 20-35. Available: https://doi.org/10.5897/SRE2021.6718.

[15] Wantuch ID, Mika. J, Szeidi L. Modelling wet and dry spells with mixture distributions. Meteorology and Atmospheric Physics. 2000; 73: 1436-5065.

[16] Cahill AT. Significance of AIC differences for precipitation intensity distributions. Advances in Water Resources. 2003; 26: 457-464.

[17] Hui W, Xuebin Z, Elaine MB. Stochastic modelling of daily precipitation for Canada. Atmosphere-Ocean. 2005; 43: 23-32.

[18] Lana X, Burgueño A, Serra C, Martínez MD. Multifractality and autoregressive processes of dry spell lengths in Europe: An approach to their complexity and predictability. Theoretical and Applied Climatology. 2017; 127(1-2): 285-303. Available: 10.1007/s00704-015-1638-0.

[19] Sirangelo B, Caloiero T, Coscarelli R, Ferrari E. A stochastic approach for the analysis of long dry spells with different threshold values in Southern Italy. Water. 2019; 11: 26. Available: https://doi.org/10.3390/w11102026.

[20] Berger A, Goossens C. Persistence of wet and dry spell at Uccle (Belgium). Journal of Climatology. 1963; 3: 21-34.

[21] Goossens C, Berger A. Persistence of dry and wet spells in Belgium. Archives for Meteorology, Geophysics and Bioclimatology.1984; B34: 243–256.

[22] Kamar K, Rao TV. Dry and wet spells at Campina Granade. Revista Brasileira de Meteorologia. 2004; 20: 71-74.

[23] Epifani C, Esposito S, Vento D. 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. 2004; 18-24.

[24] Giuseppe ED, Vento D, Epifani C, Esposito S. Analysis of dry and wet spells from 1870 to 2000 in four Italian sites. Geophysical Research Abstracts. 2005; 7: 07712. Available: 1607-7962/gra/EGU05-A-07712.

[25] Deka S, Borah M, Kakaty SC. Statistical modelling of wet and dry spell frequencies over North-East India. Journal of Applied and Natural Science. 2010; 2(1): 42-47.

[26] Sukla MK, Mangaraj A K, Sahoo LN, Sethy KM. A comparative study of three models for the distribution of wet and dry spells in the Mahanadi Delta. New York Science Journal. 2012; 5 (11): 54-61. Available: http://www.sciencepub.net/newyork.

[27] Mooley DA, Parthasarathy B. Fluctuations in All-India summer monsoon rainfall during 1871-1978. Climatic Change. 1984; 6: 287-301.

[28] Cox DR, and Miller HD. The Theory of Stochastic Process. New York: Wiley; 1967.

[29] Mathugama SC, Peiris TSG. Critical evaluation of dry spell research. International Journal of Basic and Applied Sciences. 2011; 11: 153-160. Available: 10.4236/vp.2023.93008.