A Review of Rainfall Modelling for Rainfall Occurrence and Amount

Muhammad Az-zuhri Azman[a] , Roslinazairimah Zakaria[b] and Noor Fadhilah Ahmad Radi[c]

[a,b]*Faculty of Industrial Sciences & Technology, University Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia.*

[c]*Institute of Engineering Mathematics, Universiti Malaysia Perlis, Taman Bukit Kubu Jaya, Jalan Seraw, 02000 Kuala Perlis, Perlis IK, Malaysia.*

Abstract. Rainfall modelling is very important in water management and is frequently used especially in the field of agriculture, hydrology, climatology and industries. The lack of proper water management will results in negative consequences. There are two different types of rainfall modelling on any given time scales in which are based on the rainfall occurrence and rainfall amount. However, many researchers believe that using separate models for rainfall occurrence and rainfall amount will results in losing certain information of the rainfall features. Therefore, many researchers have tried to solve these weaknesses by consolidating the models of rainfall occurrence and rainfall amount simultaneously. Thus, this paper will review various types of rainfall models which have been used in modelling rainfall occurrence, rainfall amount or combination of both rainfall occurrence and rainfall amount.

Keywords: rainfall occurrence, rainfall amount, rainfall model.

INTRODUCTION

Rainfall is the primary source of water for most part of the world especially in the areas of agriculture, industries, hydrology and climatology. All the agriculture activities and crop management can be well planned by understanding the rainfall model [1]. Rainfall components can be categorized as both discrete and continuous. The discrete rainfall component is based on the occurrence of rain event. The rain event occurs when the amount of rain is greater than or equal to certain threshold amount and the day is defined as wet. Otherwise, the day is considered as dry day (no rain event). Continuous component of rainfall is based on the rainfall amounts. Hence, it is vital to have precise knowledge of the rainfall component characteristic in order to provide the rainfall utilization in various areas of study. However, rainfall modelling also includes the modelling of both rainfall components (discrete and continuous) simultaneously. This paper will review different types of rainfall models based on rainfall amount, rainfall occurrence and combination of both.

TYPES OF RAINFALL MODEL

Modelling Rainfall Occurrence

Markov chain model has been used widely for modeling rainfall occurrences. Roldan and Woolhiser [2] found that Markov chain is superior than alternating renewal process (ARP) in modelling rainfall occurrence at four stations in the United States. According to Gabriel and Neumann [3] and Elseed [4], the rainfall occurrence on a day depends on the occurrence of rainfall on the previous day using the first-order or simple Markov chain models.

Katz [5] and Deni et al. [6] used a higher-order Markov chains to model rainfall occurrence which assumed that the occurrence of rainfall on a day depends on the occurrence of rainfall on two or more days earlier. Deni et al. [6] also claimed that using higher-order Markov chain models performed marginally better even though the model are more complex. Wilks [7] used hybrid Markov chains by considering different orders for wet and dry days while Robertson et al. [8] demonstrated using hidden Markov chains by considering some hidden states in northeast

Brazil. However, Chandler [9] used logistic regression to model dry or wet day as a function of a number of predictors in North Atlantic.

Modelling Rainfall Amounts

Linear regression model is the simplest models used for modelling monthly rainfall amounts [10–12]. Boer et al. [13] used linear regression to model the amount of seasonal and annual rainfall in New South Wales, Australia. Chowdury and Sharma [14] applied linear regression model to measure the effect of El Niño based on the southern oscillation index (SOI) on the amount of monthly rainfall. Using the data from Kota region, Bhakar et al. [15] used additive time series decomposition models to model monthly rainfall amounts. In Myanmar, Zaw and Naing [16] used polynomial regression to model the amount of monthly rainfall by considering nonlinear effects of some covariates. All of these models are not suitable in the cases where the time scales of rainfall amounts are highly skewed to the right since the basic assumptions for these models need to be normally distributed and have constant variances. Mostly the rainfall amounts for some stations on the larger timescales for example annually approximately follow normal distribution. However, for the smaller time scales of rainfall amount such as monthly, weekly and daily do not follow normal distribution and usually positively skewed to the right [17,18]. For positively skewed data, the transformation of data is needed in order to be able to model rainfall amounts using regression methods.

Meng et al. [19] have successfully transformed the monthly rainfall to normality using logarithmic transformation. In contrast, Mooley [20] shows that the poor performance is resulted when the logarithmic transformation is applied to 39 rainfall stations in Asia during the summer monsoon. Nicks and Lane [21] have used the skewed normal distribution that generalizes the normal distribution to allow for non-zero skewness while Zucchini and Adamson [22] used Weibull distribution. Moreover, Katz and Parlange [23] have used power transformation to achieve normality. In another study, some of the researchers have converted non-normal rainfall amounts to rainfall anomaly [24] or standardized precipitation index [25,26] for fitting linear regression models to the converted indices. However, these transformation methods have problem since some information regarding the rainfall data are lost. Therefore, the smaller timescales need alternative distributions for modeling the rainfall amounts.

In general, the distribution of rainfall amount is positively skewed to the right and their main interest is about the amount of rainfall in different time scales. Therefore, the gamma distribution has formed the basis of the alternative distributions for modelling the rainfall amounts which are highly skewed to the right [27–33]. Nevertheless, there are many other theoretical distributions have also been used as stated by Hasan and Dunn [1][1] for analyzing daily rainfall amounts such as Kappa [34], generalized log-normal [35], exponential [36] and mixed exponential [7,37–39] distributions. Jamaludin and Jemain [40] used four different types of distribution (exponential, gamma, mixed exponential and mixed gamma) to model daily rainfall amounts in Malaysia. Based on the Akaike Information Criteria (AIC) it is concluded that the mixture distributions are better to illustrate the daily rainfall amounts rather than using the single distribution.

Modelling Rainfall Occurrence and Amount Simultaneously

Many studies have highlighted that using separate models for modelling the rainfall occurrence and rainfall amount have resulted in losing certain information of the rainfall characteristics. Therefore, researchers believe that both rainfall components (occurrence and amount) need to be modelled in a single model to overcome the weaknesses.

Makhnin and McAllister [41] demonstrated that the daily rain gauge data sets for three areas in Continental US can be modelled for both occurrences and amounts simultaneously using truncated and power-transformed normal distribution with the spatial-temporal dependence represented by multivariate auto-regression. These distributions are able to overcome the spatial interval problem and able to give realistic special behavior by integrating both the occurrence and amount components.

In 1985, Alexandersson [42] used the compound Poisson-exponential (Cpe) distribution which is a continuous stochastic process to model the monthly rainfall occurrence and amount simultaneously using data from southern peninsula of Sweeden. The continuous stochastic process is integrated between a Poisson process and an exponential distribution. The two parameters from this model are the occurrence parameter (the mean number of independent precipitation events) and amount parameter (the mean amount at each event). Some advantages of using Cpe are that it does not require to be extended and modified for zero precipitation and the parameters can be estimated from the series of monthly rainfall amounts.

In 2004, Dunn [43] used Poisson-gamma distribution to model the monthly rainfall in Charleville, Queensland (1882-1994) and both daily and monthly rainfall in Melbourne (1981-1990), Australia. Poisson-gamma is in the family of Tweedie model. The Tweedie family of distributions fit in the class of exponential dispersion models (EDM) which are vital in generalized linear models since they represent the response distribution [44]. Withers and Nadarajah [45] used the compound Poisson-gamma distribution to illustrate the annual maximum daily rainfall for 14 locations in west central Florida from the year 1907 to 2000. The parameter estimation methods applied are unconditional maximum likelihood estimation, conditional maximum likelihood estimation and moments estimation. The result shown that unconditional maximum likelihood estimation gives the best fit.

Dunn [43] points out some vital features about the Tweedie family distribution which are very suitable to be used in modelling rainfall simultaneously:

- Total precipitation is considered as a sum of precipitation on smaller time scales.
- This family distributions is belong to the exponential family of distributions [44], upon which generalized linear models (GLMs) are based. Consequently, there is a framework already in place for fitting models based on the Tweedie distributions, and for diagnostic testing. In addition, covariates can be incorporated into the modelling procedure.
- They provide mechanism for understanding the fine-scale structure in coarse-scale data.

In the recent studies, Hasan and Dunn [1,17] have used two forms of Poisson-gamma models. The first form of Poisson-gamma does not involve predictor which fit separate model for each month whereas the second form of Poisson-gamma use a single model for station involving sine and cosine forms. Hence, they concluded the second form produces a simpler Poisson-gamma model. Their results show the flexibility of Poisson-gamma models which can be used to model monthly rainfall data very well by fitting separate models for each month without predictor [1] and using a single model for each station (not one for each month) using the sine and cosine forms which produced the simpler models [17].

In another study, Seigel [46] used non-central chi-squared distribution with zero degrees of freedom to model the January snowfall in Seattle, Washington, U.S. from 1906 to 1960 upon which the data is continuous with variety of distributions that contain exact zero values. The empirical cumulative distribution indicates the closeness of fit for the estimated parameter using the non-central chi-squared distribution.

Statistical Model for Modelling Rainfall

The Markov chain model used for rainfall occurrence has assumed the probability of rainfall on any day depends only on whether the previous day was wet (rainfall occur) or dry (rainfall not occur). Probability of rainfall is assumed as independent with the given events on the previous day and events on the next days. Parameters of two conditional probabilities referred as Markov chain [3] are given by

$$p_1 = P_r\{wet\ day \mid previous\ day\ wet\}$$

$$p_0 = P_r\{wet\ day \mid previous\ day\ dry\}$$

Therefore the probabilities of rainfall i days after a wet or a dry day are:

$$P + (1 - P)d^i$$

where

$$d = p_1 - p_0$$

$$P = \frac{p_0}{1 - d}$$

The gamma distribution is used to fit non-zero rainfall data. The gamma probability distribution function is given by

$$f(x) = \frac{(x/\beta)^{\alpha-1} e^{\frac{-x}{\beta}}}{\beta\,\Gamma(\alpha)}$$

where

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt$$

and x represents as rainfall amount, a and b are the shape and scale parameters, respectively [47].

Poisson-gamma is used to model both rainfall occurrence and amount. Modelling for both rainfall occurrence and amount simultaneously has assumed any rainfall event i produces an amount of rainfall, R_i and that each R_i comes from a gamma distribution Gam(α, γ) (mean $\alpha\gamma$ and variance $\alpha\gamma^2$). Assume the number of rainfall events in any one month is N, where N has a Poisson distribution with mean λ; that is $N \sim Pois(\lambda)$. This implies months with no rainfall when $N = 0$. The total monthly rainfall Y is retained has [43]:

$$\sum Y = R_1 + R_2 + \ldots + R_N$$

If $N = 0$ then $Y = 0$. The probability function can be simplified by writing it in logarithmic form:

$$\log f_p(y; \mu, \phi) = \begin{cases} \lambda & for\ x \,|= 0, \\ -\frac{y}{\gamma} - \lambda - \log y + \log W(y, \phi, p) & for\ y > 0 \end{cases}$$

where

$$\gamma = \phi(p-1)\mu^{p-1},\ \lambda = \frac{\mu^{2-p}}{[\phi(2-p)]}$$

and W is the Wright's generalized Bessel function given by

$$W(y, \phi, p) = \sum_{j=1}^{\infty} \frac{y^{-j\alpha}(p-1)^{j\alpha}}{\phi^{j(1-\alpha)}(2-p)^j\, j!\,\Gamma(-j\alpha)}$$

with $\alpha = \dfrac{2-p}{1-p}$. The mean of the distribution is m and the variance is $\mathrm{var}[Y] = \phi\mu^{p}$. The type of Poisson-gamma distribution used is determined by the value of index p where $1 < p < 2$. Therefore, the probability of zero precipitation is given by

$$\Pr(Y = 0) = \exp(-\lambda) = \exp\left[\frac{\mu^{2-p}}{\phi(2-p)}\right]$$

The summary of rainfall models and their properties are shown in TABLE 1.

CONCLUSION

This paper reviews varieties of the probability models which have been used for modelling rainfall occurrence, rainfall amount and combination of both components. Most models which deal in rainfall amount only consider when the rain event occurred. Certain information of the rainfall features will be lost due to modelling rainfall occurrence and amount separately. However, Poisson-gamma takes into account both components when there is rain event and no rain event. Therefore, modeling the rainfall using the Poisson-gamma distribution from the Tweedie family has a high potential of their intuitive appeal.

TABLE 1. Types of rainfall models

	Model	Properties	Limitations
Occurrence	Markov: simple; higher-order; hidden; hybrid	• for wet or dry spell • define: {0} = dry, {1} = wet • better than renewal process • good for parameter estimation	• higher order not parsimonious • not suitable for multisite • do not include higher order correlation
	Logistic Regression	• for rainfall occurrence	• cannot fully capture the inter-site variability • incapable to handle spatial data
Amount	Linear regression; additive time series decomposition; polynomial regression	• normally distributed • have constant variances	• not suitable for smaller timescales with highly • skewed to the right
	Normal based distribution	• to achieve normality, data is transformed using logarithmic transformation or power transformation	• transformation method used has caused lost • of data information
	Gamma	• use nonzero rainfall amounts	• underestimate extreme rainfall amounts
Occurrence & amount	Poisson-gamma	• total precipitation is considered as a sum of precipitation on smaller time scales • belong to the exponential family of distributions based on GLMs • framework already in place for fitting models based on the Tweedie distributions and for diagnostic testing • covariates can be incorporated into the modelling procedure • provide mechanism for understanding the fine scale structure in coarse-scale data	• the variance for the observed data is always greater than the predicted data (over dispersion) • the density function cannot be written in closed form
	Truncated and power-transformed normal distribution	• able to overcome the spatial interval problem computationally expensive Gibbs sampler • easy to use for the derivation of an extreme value distribution • use generator to integrate the occurrence and amount • exogenous variables can be incorporated easily	• parameter values should be estimated using computationally expensive Gibbs sampler
	Compound Poisson-exponential(Cpe)	• able to accommodate zero and non-zero precipitation	