Daily Rainfall Disaggregation Using HYETOS Model for Peninsular Malaysia

Ibrahim Suliman Hanaish, Kamarulzaman Ibrahim, Abdul Aziz Jemain

Abstract— In this paper, we have examined the applicability of single site disaggregation model (HYETOS) based on the Poisson cluster model to disaggregate daily rainfall to hourly data using proportional adjusting procedure. In this study, the modified Bartlett Lewis Model (MBL) is fitted to the hourly rainfall depth from 1970 to 2008 available at the rain gauge station in Petaling Jaya. In addition, the synthetic hourly rainfall is generated by inputting the estimated parameters found based on MBL into the Hyetos model. The nature of occurrence of rainfall in Peninsular Malaysia comprising of very heavy rain during a short period of time contribute to the discrepancy between the synthetic data and the observed and expected data. When the disaggregated synthetic rainfall data is compared with the observed and expected data, it is found that the mean values for the three types of depths are quite closed; however, the synthetic data are quite different from the observed and expected depths when comparison is base on autocorrelation and standard deviation. The model is also validated by considering statistical property that was not used in the fitting procedure such as the extreme values. A comparison is also made between the extreme values based on the disaggregated model and the observed data. A bad fit in the extremes is found at all time scales considered, which are 1, 6 and 12 hour levels of aggregation. An underestimation of the disaggregated values is evident at all time scales.

Keywords— Disaggregation, Poisson cluster processes, Hyetos.

I. Introduction

THE rainfall in Malaysia is usually received using daily rain gauge that are available at the meteorological stations throughout the country. However, rainfall data are often required at a finer scale such as hourly rather than daily. Accordingly, disaggregation has recently become a major technique for hydrologic modeling of rainfall time series data. It is an important step in the process to obtain lower-level time scale data (e.g. hourly rainfall data) from higher time scales (e.g. daily data). There are many approaches and methodologies available to disaggregate rainfall data. Models based on stochastic point processes have been used for disaggregation, particularly, on Poisson cluster models. The Bartlett-Lewis model is based on stochastic point processes, is pertinent to the problem of spatial-temporal disaggregation. Many efforts have been put into disaggregating rainfall amounts temporally such as daily rainfall totals into hourly precipitation at a single site [1] used a modified Bartlett-Lewis model to disaggregate daily rainfall for a site in Italy while [2] proposed a methodology based on Bartlett-Lewis model for a data set in the US such that the hourly rainfall sequence added up consistently to the daily totals. The work by [3] combined a modified Bartlett-Lewis model with a regionalized hybrid model for the purpose to disaggregate Australian daily rainfall to hourly data. In Ethiopia, a method to redistribute the outputs of disaggregated hourly rainfall of the modified Bartlett-Lewis model is suggested [4]. However, rainfall disaggregation models based on stochastic point processes have not been widely applied for Peninsular Malaysia. The objective of this paper described here is to fit a spatial-temporal stochastic model to the observed hourly rainfall data taken from Petaling rain gauge station and to use the fitted stochastic model to disaggregate the daily data to hourly data using HYETOS model.

Manuscript received May 26, 2011: Revised version received xxx.

Ibrahim Suliman Hanaish is with the School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM), 43600, Bangi, Selangor, MALAYSIA (corresponding author to provide phone: 0060173948730; e-mail: henaish@yahoo.com)

Kamarulzaman Ibrahim is with the School of Mathematical Sciences, Faculty of Science and Technology and also with the Solar Research Institute Universiti Kebangsaan Malaysia, 43600 UKM Bangi, and SERI Selangor, MALAYSIA.

Abdul Aziz Jemain is with the School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, MALAYSIA.

II. Single-site temporal disaggregation model

A. Model description

The Modified Bartlett-Lewis rectangular pulse model (**MBL**) was considered in many works due to its wide applicability for describing various different climates. The diagrammatic explanation of the Modified Bartlett-Lewis rectangular pulse model (**MBL**) is depicted in Fig. (1) and the assumptions for the model are as follows. The storm origins are assumed to follow a Poisson process with rate λ and the cell origins follow a Poisson process with rate β. Cell arrivals terminate after a particular time, and this length of period is exponentially distributed with parameter γ. Each cell has a duration exponentially distributed with parameter η. The distribution of the uniform intensity is typically assumed exponential with a parameter μ_x. For each storm, the parameter η is randomly varied from storm to storm with a gamma distribution with shape parameter α scale parameter, i.e. $E(\eta) = \alpha/v$ and $var(\eta) = \alpha/v^2$. Subsequently, parameters β and γ also vary in a manner that the ratios $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$ be constant. Therefore, a **6** parameter model **MBL** is described by the set of parameters $(\lambda, \mu_x, \alpha, v, \kappa, \phi)$ as shown in Figure 1. The equations of the BL model, in its original or the modified (random parameter) configuration, may be found in the appropriate references such as Rodriguez-Iturbe et al.[5]. These equations relate the statistical properties of the rainfall process in discrete time in the entire time domain, to the model parameters and serve as the basis for model fitting using these statistical properties.

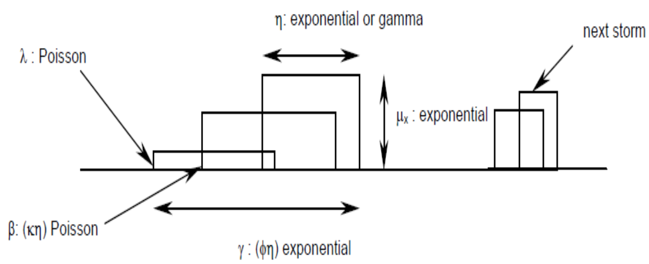

Fig.1. Explanatory sketch for the Bartlett-Lewis rectangular pulses model

B. Fitting of the MBL

The model fitting to empirical data is difficult for models based on Poisson cluster processes because the parameters are related rather indirectly to observable properties of rainfall sequences. Fitting techniques may be classified broadly into moment, likelihood and Bayesian methods. The latter two approaches allow for a formal comparison between models, but require the formulation of a likelihood function. Since the likelihood functions for models based on Poisson cluster processes cannot be obtained in a closed form, method of generalized moments as suggested by [6] is used for parameter estimation of the MBL. This method involves choosing parameters that achieve as close as possible a match, according to a weighted least-squares criterion, between the observed and expected values. The resulting set of non-linear equations could be solved simultaneously to derive parameters for the model. Specifically, let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ be the parameter vector for the model, $\boldsymbol{T} = (T_1, T_2, \ldots, T_K)'$ be a vector of summary statistics computed from the data and $\boldsymbol{\tau}(\boldsymbol{\theta}) = (\tau_1(\theta), \tau_2(\theta), \ldots, \tau_K(\theta))'$ denotes a vector of the expected value of **T** under the model. The idea behind the method of generalized moments is to choose $\boldsymbol{\theta}$ to minimize a quadratic form given by

$$S(\boldsymbol{\theta}) = [\mathbf{T} - \tau(\boldsymbol{\theta})]' \mathbf{W} [\mathbf{T} - \tau(\boldsymbol{\theta})] \tag{2.2.1}$$

where $\mathbf{W}$ is a $k \times k$ matrix of 'weights' which is determined based on historical data. A special case of (2.2.1) can be given by

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{k} w_i [\mathrm{T}_i - \tau_i(\boldsymbol{\theta})]^2 \tag{2.2.2}$$

where $(w_i : i = 1, 2, \ldots, k)$ is a collection of positive weights. This is a special case of (2.2.1), in which the matrix $\mathbf{W}$ is diagonal with ith diagonal element w_i. Therefore, this objective function is minimized in order to reduce the error between the calculated statistics (T_i) and the expected statistics (τ_i). Minimization of the objective function (1.2.2) is done using the Nelder-Mead optimization with diagonal element w_i.

III. The HYETOS Model

Once the MBL parameters are obtained, they can be used as an input to the single site disaggregation HYETOS model [7]. The Hyetos software disaggregates the daily rainfall at a single site into hourly data using Bartlett Lewis model as a background stochastic model for rainfall generation. The daily series is divided into clusters of wet days and several runs from the Bartlett Lewis model are performed separately for each cluster of wet days. The runs continue until the sequence of synthetic daily depths matches the sequence of daily totals with a tolerance distance, d, defined as

$$d = \left[\sum_{i=1}^{L} ln\left(\frac{Z_i + c}{\tilde{Z}_i + c}\right)^2\right]^{\frac{1}{2}}, \tag{3.1}$$

where Z_i and $\tilde{Z}_i$ are respectively, the original and simulated daily totals at the rain gauge station, L is the sequence of wet days, c is a small constant (set to 0.1mm). A correction procedure referred to as proportional adjustment to make the generated hourly series fully consistent with given daily totals is applied based on [2]. The proportional adjusting procedure modifies the initially generated values $\tilde{X}$ to get the modified values X according to

$$X_s = \tilde{X}_s\left(\frac{Z}{\sum_{j=1}^{24} \tilde{X}_j}\right), s = 1, 2, \ldots, 24, \tag{3.2}$$

where Z is the daily depth to be disaggregated.

IV. Study Area and Input Data

Peninsular Malaysia lies entirely in the equatorial zone, situated between 1° and 6° in the northern latitude and between 100° to 103° in the eastern longitude. It experiences rainfall that varies seasonally with respect to the occurrence of the monsoon winds. This seasonal variation is mainly influenced by the southwest monsoon which occurs between May and August and the northeast which blows from the month of November and February. During the northeast monsoon, many areas in the east coast of the Peninsula are expected to receive heavy rainfall. On the other hand, the areas that are sheltered by the mountain ranges on the west coast are more or less free from the influence of the north east monsoon. In addition, the transition period between the monsoons, i.e. the inter-monsoon period, occurs in the months of March to April and September to October. In this study, hourly rainfall data were obtained from the rain gauge station in Petaling Jaya, which is located in the western area and typically of midlands. We choose this station because it is very much influenced by the monsoons characterized by frontal and convective rainfall as seen from Figures 2 .The hourly data ranging from the period of 1970 to

ISBN: 978-1-61804-016-9

2008 can be considered as having a good quality since more than 98% of the data are available. Since the hourly data are for the period of 38 years and the proportion of missing data is small, it is therefore possible to study the model based on various scales of aggregated data. The rainfall station is being maintained by Malaysian Meteorological Service (MMS). The most frequently used sampled moments to determine model parameters are 1 hour mean (Mean1), 1 hour variance (Var1), 6 hours variance (Var6), 24 hours variance (Var24),1 hour autocorrelation of lag-1 (ACF1(1)), 24 hours autocorrelation of lag-1 (ACF1(24)), 1 hour probability of wet (Pdry1), 24 hours probability of wet (Pdry24) as [5].

V. Results and discussions

A. Evaluation of model performance

The accuracy of the model was assessed using goodness of fit statistics. The performance of single site disaggregation model stations is made with the allowance for the different monsoon periods. This implicitly takes into accounts of the monthly variation. As suggested by [8], the disaggregated time series cannot be compared to the measured data on an hour by hour basis because of the uncertainty in the start times of storms in the disaggregated data. The four important statistics that should be matched with the disaggregated data are the mean, probability of wet, variance and the autocorrelation lag-1 of hourly rainfall. The expected statistics were calculated under the fitted MBL model and compared with values obtained by applying the disaggregation procedure to the daily data from the rain gauge station. The expected statistics reproduced by the model are closed to the statistics determined using the observed data, which is a clear indication of the appropriateness of the model in the study area. Graphically, Figure 2 gives a comparison of observed, MBL fitted and disaggregated values of some hourly summary statistics for Petaling Jaya rain gauge station. It can be observed that a close agreement between the observed and the disaggregated series is obtained for the mean. Larsen standard deviation in the observed statistics for the whole year is however noted, partly due to the large variance compared to the disaggregated statistics. An overestimation of wet probability is observed for all months. The autocorrelation for the disaggregated series highly overestimated those for the observed series for all months of the year. Quantitatively, the accuracy of the model predictions was assesses using goodness of fit statistics included the bias $\bar{e}$, Eq. (5.1.1); the standard error of the difference S_e, Eq. (5.1.2); the modified standard error of the difference S_{em}, Eq. (5.1.3). These above measures are given by

$$\bar{e} = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i - Y_i) \tag{5.1.1}$$

$$S_e = \sqrt{\frac{1}{v}\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2} \tag{5.1.2}$$

$$S_{em} = \sqrt{\frac{1}{v}\sum_{i=1}^{N}(\hat{Y}_i - \bar{e} - Y_i)^2} \tag{5.1.3}$$

where N is the total number of hours during which weather data were observed and generated, v is the degree of freedom $(v = N - 1)$. Y and $\hat{Y}$ are observed and predicted hourly rainfall, respectively. The goodness of fit statistics can also be standardized to yield dimensionless indices such as the relative bias R_b, Eq. (5.1.4); the relative standard error R_s, Eq. (5.1.5); the relative difference between observed and predicted hourly standard deviations ΔS, Eq. (5.1.6); and significance of difference test Eq. (5.1.7), which are given by

$$R_b = \frac{\bar{e}}{\bar{Y}} \tag{5.1.4}$$

$$R_s = \frac{S_e}{S_y} \tag{5.1.5}$$

$$\Delta S = \frac{S_y - \hat{S}_y}{S_y} \tag{5.1.6}$$

$$Test = \frac{|\bar{Y} - \bar{\hat{Y}}|}{2S_e} \tag{5.1.7}$$

where $\bar{Y}$ and $\bar{\hat{Y}}$ are the means of observed and predicted hourly rainfall respectively, S_y and $\hat{S}_y$ are the standard deviations of observed and predicted hourly rainfall respectively. Values of R_b, R_s, and ΔS greater than one indicate reasonable differences between measured and model predicted values. The statistical significance of the differences was evaluated employing a criterion of $2 \times S_e$, which roughly corresponds to a two sided test at 5% significance level [9]. That is, if the value of the Test is larger than 1.0, then the difference is statistical significance. Table 5 shows the goodness of fit statistics computed for each measure for all months. Results from Table 5 indicate that the Hyetos model did not reproduce the disaggregated rainfall very well. The high correlation coefficients between disaggregated and observed hourly rainfall data $(r = 0.57)$ at July, low values for $\Delta S = -0.2918$ and $R_s = 0.8262 < 1.0$, and statistically non significant differences $(Test < 0.001 \ll 1.0)$. The HYETOS model is effective in reproducing the mean hourly rainfall since there is a moderate correlation between the disaggregated and observed hourly rainfall data.

B. Effect of single site disaggregation

To assess the effect of using HYETOS model, at least 20 runs are made for each level for disaggregation. The results show an overall bad fit between the disaggregated and observed series, particularly, the reproduction of the skewness is not encouraging for the performance of the extreme value. According to [10], high probability of zero values combined with the proportional adjusting procedure with repetition or not, may introduce the bias in the variation and skewness of the process, as shown in Figure 2. This phenomenon is thought to be related to the fact that the model is not very accurate at reproducing the skewness of the time series. The HYETOS model tends to produce the time series with a significantly lower skewness than the observed data; therefore, indicating the lack of extreme values in the simulated series.

Table 5 Summary of the goodness of fit statistics at Petaling Jaya rain gauge station

statistics \ months	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
CORRELATION	0.304	0.203	0.202	0.125	0.134	0.162	0.578	0.240	0.245	0.181	0.220	0.277
Bias	-0.003	0.000	0.000	-0.008	0.000	-0.007	0.000	0.000	0.000	0.000	0.009	0.000
S_e	10.642	12.800	17.181	20.395	13.217	12.813	9.531	10.389	13.142	16.503	17.26	14.392
S_{em}	10.642	12.800	17.181	20.395	13.217	12.813	9.531	10.389	13.142	16.503	17.26	14.392
R_b	-0.003	0.000	0.000	-0.004	0.000	-0.006	0.000	0.000	0.000	0.000	0.003	0.000
R_s	1.017	1.006	1.069	1.082	1.037	1.018	0.826	0.984	0.992	1.029	1.004	0.992
TEST	0.019	0.000	0.000	0.090	0.000	0.050	0.000	0.000	0.000	0.000	0.079	0.000
ΔS	-0.339	-0.564	-0.367	-0.442	-0.557	-0.586	-0.291	-0.597	-0.542	-0.514	-0.541	-0.473

S_e, S_{em}, R_b, R_s, and are the standard error, modified standard error, relative bias and relative standard error respectively

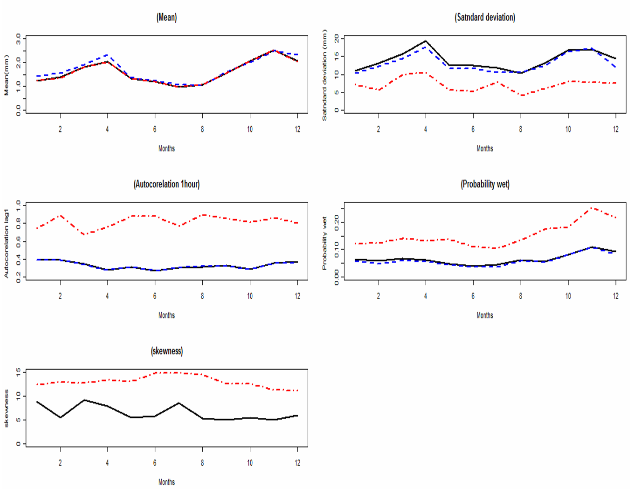

Fig. 2. The fitted HYETOS single-site disaggregation model on properties of hourly rainfall for Petaling Jaya rain gauge station. In each plot, the Observed (solid line); Expected (dashed line); and disaggregated (dashed-dotted line) statistics

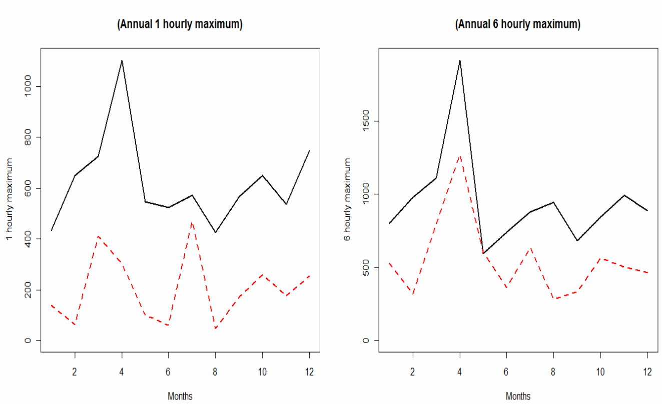

Fig. 3. The fitted HYETOS single-site disaggregation model on properties of annual maximum rainfall for Petaling Jaya rain gauge station. In each plot, the Observed (solid line) and disaggregated (dashed-dotted line) statistics.

VI. Conclusion

A single site disaggregation model (HYETOS Model) has been proposed, which enables single site rainfall sequences to be generated at a subdaily time scale. Hourly rainfall data are then obtained using HYETOS model based on the Poisson cluster models with repetition techniques and proportional adjusting procedure to obtain a subdaily temporal profile Petaling Jaya rain gauge station. The performance of the proposed disaggregation procedure has been assessed in terms of the reproduction of a set of standard statistics of levels of aggregation. Results indicate a good performance of the methodology in preserving the mean, however, we observe a large bias of the disaggregated series with respect to observed series which must be attributed to the reasons explained in 5.2. In addition, the model was not able to closely match a range of important statistical property that was not used in the fitting procedure, such as extreme values. For both the hourly and 6 hourly time scales, the model underestimated the extreme values, as it had not generated enough extreme rainfall events within the simulated period, as compared to observed data. This phenomenon is thought to relate to the fact that the model is not very accurate at reproducing the skewness of the time series. Further works may involve exploring the application of the MuDRain model (Multivariate Disaggregation Rainfall Model) to improve the reproduction of dry periods and standard deviation.

References

[1] Bo, Z., S. Islam, and E. Eltahir, *Aggregation–disaggregation properties of a stochastic rainfall model.* Water Resources Research, 1994. **30**(12): p. 3423–3435.

[2] Koutsoyiannis, D. and C. Onof, *Rainfall disaggregation using adjusting procedure on a Poisson cluster model.* hydrology 2001. **246**: p. 108-122.

[3] Gyasi-Agyei, Y., *Stochastic disaggregation of daily rainfall into one-hour time scale.* Journal of Hydrology, 2005. **309**: p. 178–190.

[4] Agizew, N.E. and E. Michel, *Characterization and disaggregation of daily rainfall in the Upper Blue Nile Basin in Ethiopia.* Journal of Hydrology, 2011. **399**: p. 226-234.

[5] Rodriguez-Iturbe, I., D.R. Cox, and V. Isham, *Some models for rainfall based on stochastic point processes.* Proc. R. Soc. Lond, 1987a. **A 410**: p. 269-288.

[6] Rodriguez-Iturbe, I., D. Febres, and J.B. Valdes, *Rectangular pulses point process models for rainfall: analysis of empirical data.* J. Geophys. Res, 1987b. **92**: p. 9645-9656.

[7] Koutsoyiannis, D. and C. Onof, *HYETOS a computer program for stochastic disaggregation of fine scale rainfall*. 2000, Available from http://www.itia.ntua.gr/e/softinfo/3/>.

[8] Gutierrez-Mangness, A.L. and R.H. McCuen, *Accuracy evaluation of rainfall disaggregation methods.* Journal of Hydrologic Engineering, 2004. **9**(2): p. 71-78.

[9] Buishand, T. and T. Brandsma, *Multi-site simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling.* Water Resour. Res., 2001. **37**: p. 2761-2776.

[10] Onof, C. and H.S. Wheater, *Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model.* Journal of Hydrology, 1993. **149**(1-4): p. 67-95.