Multivariate Rainfall Disaggregation Using MuDRain Model: Malaysia Experience

Ibrahim Suliman Hanaish

Department of Statistics, Faculty of Science, Misurata University, Misurata, Libya

Abstract— *Disaggregation daily rainfall using stochastic models formulated based on multivariate approach (MuDRain) is discussed in this paper. Seven rain gauge stations are considered in this study for different distances from the referred station starting from 4 km to 160 km in Peninsular Malaysia. The hourly rainfall data used are covered the period from 1973 to 2008 and July and November months are considered as an example of dry and wet periods. The cross-correlation among the rain gauges is considered for the available hourly rainfall information at the neighboring stations or not. This paper discussed the applicability of the MuDRain model for disaggregation daily rainfall to hourly rainfall for both sources of cross-correlation.The goodness of fit of the model was based on the reproduction of fitting statistics like the means, variances,coefficientsof skewness, lag zero cross-correlation of coefficients and the lag one autocorrelation of coefficients. It is found the correlation coefficients based on extracted correlations that was based on daily are slightly higher than correlations based on available hourly rainfall especially for neighboring stations not more than 28 km. The results showed also the MuDRain model didnot reproducestatistics very well.In addition,a bad reproduction of the actual hyetographs comparing to the synthetic hourly rainfall data.Mean while, it is showed a good fit between the distribution function of the historical and synthetic hourly rainfall. These discrepancies are unavoidable because of the lowest cross correlation of hourly rainfall. The overall performance indicated that the MuDRain model would not be appropriate choice for disaggregation daily rainfall.*

Keywords— Hourly rainfall, MuDRain Model, Rainfall Disaggregation.

I. INTRODUCTION

The availabilityof rainfall data at fine scale such as hourly datais very limited in the most area. For many hydrological applications, hourly data are often required especially, in flood studies andmodeling of storm runoff in an urban setting.

Limited availability of such data leads to the option of disaggregation daily rainfall in to hourly data. They are several techniques have been considered for disaggregation such as linear disaggregation models studied by many authors starting from the work by Valencia and Schaake(1972; 1973)following with Mejia and Rousselle(1976),Tao and Delleur(1976), Hoshi and Burges (1979), Lane (1979, 1982), Salas et al. (1980), Todini(1980), Stedinger and Vogel(1984), Pereira et al.(1984), Stedinger et al.(1985) and Salas(1993).

This linear approach has been applied for univariate and multivariate applications such as disaggregation annual to monthly rainfall. However,Valencia and Schaake(1972)advised that this kind of disaggregation is not suitable for disaggregation of rainfall less than monthly due to some properties for smaller timescalesrainfall such as skewed distributions and intermittent nature of the rainfall process.Schaake et al. (1972)developed a Markov chain model for the disaggregation of monthly rainfall into daily.Woolhiser and Osborn(1985)who first studied disaggregation rainfallat smaller timescales. Based on anondimensionalized Markov process they presented a scheme for disaggregation of individual storm depths into shorter periods.

Following this study, many studies have been conducted for disaggregation rainfall approach to finer timescales such as daily rainfall to hourly. Some of these works were studied by Koutsoyiannis and Xanthopoulos(1990), Koutsoyiannis(1994) and Koutsoyiannis and Onof(2000, 2001) for the univariate approach. All fine time scale rainfall disaggregation studied mention above have a common characteristic: they are single-site approach (univariate approach).The problem of multivariate fine-timescale rainfall disaggregation problem has not been studied so far in the rainfall modeling literature.This multi site approach presents significant differences from that of the univariate disaggregation approach, including increased mathematical complexity. The spatial correlation between different sites must be maintainedin the general multivariate spatial-temporal rainfall disaggregation problem, i.e. the simultaneous rainfall disaggregation at several sites.In other words, the spatial correlation is turned to advantage since, in combination with the available single-site hourly rainfall information, it enables more realistic generation of the synthesized hyetographs.

Multivariate rainfall disaggregation is of significant practical interest even in problems that are traditionally regarded as univariate aspect, i.e. performing temporal disaggregation at one location only. In all univariate disaggregation rainfall models, the synthetic hourly series arefully consistent with the known daily series, and simultaneously, statistically consistent with the actual hourly rainfall series.

Obviously, however, these synthetic series resemble the actual ones only in a stochastic sense, i.e. they represent a likely realization of the process and do not coincide with the actual one.Thus, for example, the location of a rainfall event within a day and the maximum intensity would not be arbitrary, as in the case of univariate disaggregation, but resemble their actual values.In this study, we investigate the possibility of generating a spatially and temporally consistent hourly rainfall series at a raingaugeof interest by using the daily rainfall data available at a neighboring raingaugewhen the cross-correlation between the two stations is significant.

Amounts of contributions on disaggregation methods give support to solve this data limitation problem.A first attempt to incorporate more than one site into rainfall disaggregation was done by Socolofskyet al.(2001)who disaggregated daily rainfall to hourly.Koutsoyiannis et al.(2003) have studied a comparison between multivariate approach and the univariate approach to disaggregate daily rainfall to hourly for several stations usingthe Brue catchment in South-Western Englandas the referee station. Subsequently Fytilas(2002)to the Tiber river basin in central Italy.More recently,Bekele et al. (2007)based on two referee stations was performed disaggregation of daily rainfalls into hourly values and results were compared with various approaches (uniform distribution and stochastic models) of disaggregation daily rainfall into hourly data in the Cedar Creek watershed, TX, USA.

In this study, we examine the applicability of the multivariate disaggregation model to disaggregate daily rainfallto hourly rainfall in Peninsular Malaysia.

I. DATA SET

Peninsular Malaysia is located in the equatorial zone, situated between 1° and 6° in the northern latitude and between 100° and 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, areas on the west coast that are sheltered by the mountain ranges are more or less free from the influence of the northeast 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.

Peninsular Malaysia has many rainfall gauge stations of hourly data which only seven stations located at west, are used in this study for their better data quality and larger number of records and specifically those of Genting, Gombak,Ampang, Petaling Jaya, Seremban, Kangsar and Sitiawan.

The hourly data used are covered the period from 1973 to 2008. July and November months are considered as an example of dry and wet periods. Two key rain gauge stations, Genting and Gombak were used as sources of hourly. The other rain gauge stations are used for tests and comparisons with the simulated series obtained in this disaggregation framework to allow the effectiveness of the methodology to be evaluated. Details of these particular stations and their corresponding neighbouring stations are given in Tables 1, and their location is shown in Figure 1.

Table 1 Geographical location and distance from the target station in bold and the neighboring stations at the west coast of Peninsular Malaysia used in this study

No.	Station name	Latitude	Longitude	Euclidean Distance (km)
2	Genting	3.23	101.75	0.000 (0.00)
3	Gombak	3.16	101.75	0.050 (4.01)
1	Ampang	3.27	101.72	0.070 (28.12)
4	Petaling	3.1	101.65	0.164 (48.75)
5	Seremban	2.73	101.95	0.539 (60.31)
6	Kangsar	3.9	102.43	0.955 (105.30)
7	Sitiawan	4.22	100.7	1.443 (159.57)

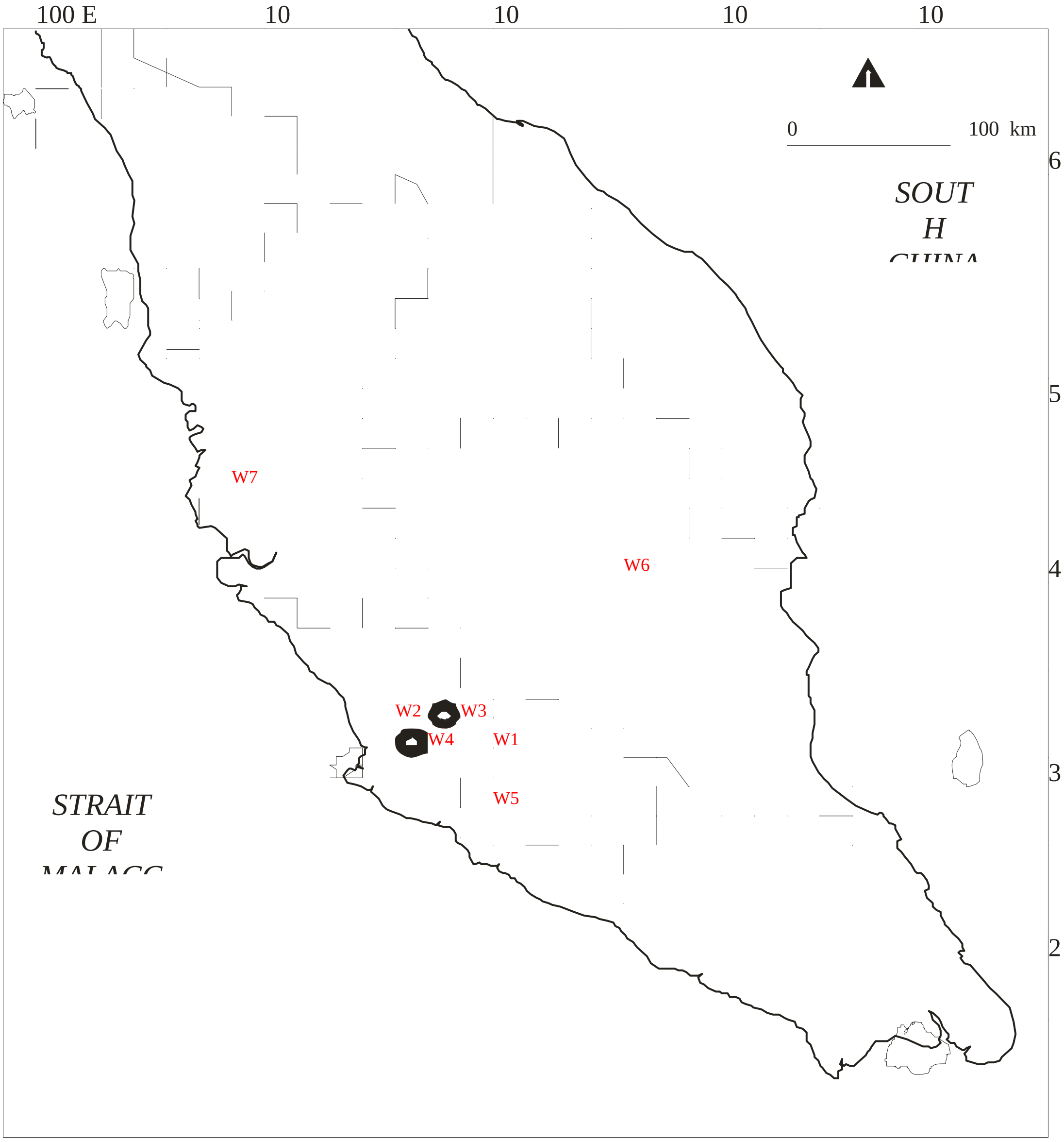


Fig.1: The physical map showing the location of seven rainfall stations considered in this study.

are $(n \times n)$ matrices of parameters and $V_s = (s = ..., 0, 1, 2, ...)$ is an independent identically distributed (iid) sequence of size n vectors of random variables which are both spatially and temporally independent.Alternatively, This model can be expressed in terms of some nonlinear transformation $\mathbf{X}_s^*$ of the hourly depths $\mathbf{X}_s$ (e.g., a power transformation) in which case (1) is replaced by

$$\mathbf{X}_s^* = \mathbf{a}\mathbf{X}_{s-1}^* + \mathbf{b}\mathbf{V}_s \, . \tag{2}$$

This model can also preserve the essential statistics of the multivariate rainfall process considered here which are the means, variances, coefficients of skewness, lagzero

II. MULTIVARIATE DISAGGREGATION RAINFALL MODEL (MUDRAIN)

Two multivariate disaggregation rainfall models have developed by (Koutsoyiannis and A. 1996) and (Koutsoyiannis 2001), are a simplified multivariate rainfall model and a transformation model that provide the required hourly rainfall series.

3.1 Simplified multivariate rainfall models

For n locations, we may assume the simplified multivariate rainfall model of the hourly rainfall distributions follow an AR(1) process which is given by

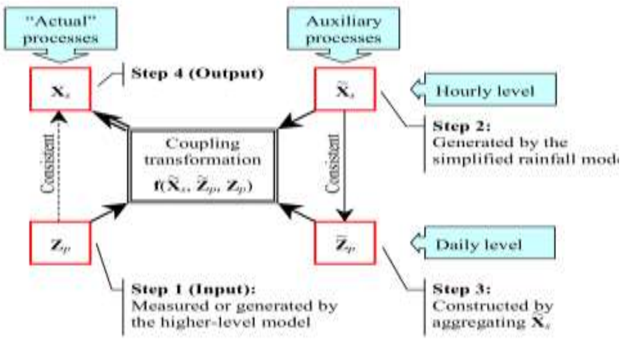
$$\mathbf{X}_s = \mathbf{a}\mathbf{X}_{s-1} + \mathbf{b}\mathbf{V}_s \tag{1}$$

Where $\mathbf{X}_s := [X_s^1, X_s^2, ..., X_s^n]^T$ represents the hourly rainfall at time (hour) s at n locations, $\mathbf{a}$ and $\mathbf{b}$

a specific stochastic model is available for each involved time scale. This frame work is depicted in Figure2 (adapted from Koutsoyiannis et al. (2003)), where $\mathbf{X}_s$ and $\mathbf{Z}_p$ represent the actual hourly and daily level processes, given by (3) and $\tilde{\mathbf{X}}_p$ and $\tilde{\mathbf{Z}}_p$ denote some auxiliary processes, represented by thesimplified rainfall model in our case, which also satisfy a relationship identical to (3). A linear transformation, $f\ (\tilde{\mathbf{X}}_s, \tilde{\mathbf{Z}}_p, \mathbf{Z}_p)$ whose outcome is a process identical to $\mathbf{X}_s$ and consistent to $\mathbf{Z}_p$ is calledcoupling transformation
(Koutsoyiannis 2001) is given by

$$\mathbf{X}_p^* = \tilde{\mathbf{X}}_p^* + \mathbf{h}(\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*) \qquad (4)$$

where



$$\mathbf{h} = \mathrm{Cov}[\mathbf{X}_p^*, \mathbf{Z}_p^*]\{\mathrm{Cov}[\mathbf{Z}_p^*, \mathbf{Z}_p^*]\}^{-1} \qquad (5)$$

Fig. 2: Schematic representation of actual and auxiliary processes, their links, and the steps followed to construct the actual lower-level process from the actual higher-level process

The quantity $\mathbf{h}(\mathbf{Z}_p^* - \tilde{\mathbf{Z}}_p^*)$ in (4) represents the correction applied to $\tilde{\mathbf{X}}$ toobtain $\mathbf{X}$. Whatever the value of this correction is, the coupling transformation will ensure preservation of first and second order properties of variables (means and variance-covariance matrix) and linear relationships among them (in our case the additive property (3)). However, it is desirable to have this correction as small as possible in order for the transformation not toaffect seriously other properties of the simulated processes (e.g., the skewness). It is possible to make the correction small enough, if we keep repeating the generation process for the variables of each period (rather than performing a single generation only) until a measure of the correction becomes lower that an accepted limit. This measure can be defined as cross-correlation of coefficients and the lag one autocorrelation of coefficients.

Equations to estimate the model parameters $\mathbf{a}$ and $\mathbf{b}$ and the moments of $\mathbf{V}_s$ directly from the statistics to be preserved are given for instance by(Koutsoyiannis 1999)for the most general case.

3.2 The Problem Formation

Given a time series $\mathbf{z}_p$ of the actual process $\mathbf{Z}_p$ has been generated using model (2) (or any other, linear or nonlinear, parametric or nonparametric, model, or even, it has been acquired from measurements).To this aim, we first generate another (auxiliary) time series $\tilde{\mathbf{x}}_s$ using the simplified rainfall process $\tilde{\mathbf{X}}_s$. The latter time series is generated independently of $\mathbf{z}_p$ and, therefore, $\tilde{\mathbf{x}}_s$ do not add up to the corresponding $\mathbf{z}_p$, as required by the additive property (3), but to some other quantities, denoted as $\tilde{\mathbf{z}}_s$. Thus, in a subsequent step, we modify the series $\tilde{\mathbf{x}}_s$ thus producing the series $\mathbf{x}_s$ consistent with $\mathbf{z}_p$ (in the sense that $\mathbf{x}_s$ and $\mathbf{z}_p$ obey $\sum_{s=(p-1)k+1}^{pk} \mathbf{X}_s = \mathbf{Z}_p$ (3), where k is the number of fine scale time steps within each coarse-scale time step (24 for the current application) without affecting the stochastic structure of $\tilde{\mathbf{x}}_s$. For this modification we use a so-called coupling transformation.

3.3 Coupling transformation

In the case of univariate problems, transformation techniques commonly known as adjusting procedures,are able to modify a series generated by any stochastic process tosatisfy some additive property (e.g. to make the sum of the values of a number of consecutive variables be equal to a given amount), without affecting the first and second order properties of the process, studied by (Koutsoyiannis 1994) and (Koutsoyiannis and A. 1996), although the adjusting procedures can be applied to multivariate problems as well, but in a repetition framework.In addition, a true multivariate transformation was studied by(Koutsoyiannis 2001) which is proposed a generalized framework for coupling stochastic models at different time scales, i.e. it is applied simultaneously to all the variables of all locations involved in the problem examined, rather than adjusting the variables of each location separately. Furthermore, the methodology proposed can be applied not only to the simple AR(1) model but to any type of stationary or seasonal stochastic model for any time scale, under the only requirement that relationship for computing hourly cross-correlations for the rest of the six daily stations is given by

$$r_{ij}^{h} = \left(r_{ij}^{d}\right)^{m} \tag{7}$$

Where:

r_{ij}^{h} is the cross-correlation coefficient between rain gages i and j at the hourly time scale, r_{ij}^{d} is the cross-correlation coefficient between rain gages i and j at the daily time scale and m is an exponent that can be estimated by regression using the known cross-correlation coefficients at the hourly and daily time scale or, in case no hourly data is available, its value can be assumed approximately in the range 2 to 3Fytilas(2002).To run the MuDRain model, we prepare the data as follows:

1. Calculate the cross-correlation between the hourly and daily rainfall data for each two rain gauge key stations, Genting and Gombak based on monthly separately.
2. Fitting between hourly and daily cross-correlations for each month based on data at each two rain gauge key stations and the values of m are given in Table2.
3. Calculate the cross-correlation between daily rainfall data at Genting station with other six daily surrounding stations each.
4. Calculate the cross-correlation between each key two stations and other six surrounding stations based on the values of m under step 2 and daily cross-correlations under step 3.

IV. ANALYSIS: CASE STUDY

In this framework we apply two approaches in the case whether the hourly data from neighbouring stations are available or not. The latter we apply the empirical

$$\Delta = \left\| \mathbf{h}(\mathbf{Z}_{p}^{*} - \mathbf{Z}_{p}^{*}) \right\| / (m\sigma_{X}) \tag{6}$$

where m is the common size of $\mathbf{X}_{p}^{*}$ and $\mathbf{X}_{p}^{*}$, σ_X is standard deviation of hourly depth (common for all locations due to stationarity assumption) and $\|.\|$ denotes the Euclidian norm.

III. METHODOLOGY IMPLEMENTATION

The methodology was implemented in a computer program with the name MuDRain (abbreviating Multivariate Disaggregation of Rainfall). The program automates most tasks of parameter estimation, performs the disaggregation and provides tabulated and graphical comparisons of historical and simulated statistics of hourly rainfall. In the parameter estimation phase, the program estimates all statistics to be preserved that are mention above.

To run the MuDRain model, we need two rainfall data sets: (1) daily rainfall data from the stations to be disaggregated into hourly data, and (2) hourly rainfall data from a reference weather station. In addition, cross-correlation between hourly rainfall data is also required to run the model. Computation of cross-correlations, and thus disaggregation runs, are recommended on monthly bases to maintain seasonality in rainfall distributions(Koutsoyiannis 2001). Hourly rainfall data of Genting and Gombak areconsidered based on monthly and cross correlations between hourly and daily rainfall data are also computedfor the two stations. The empirical relationship to find the unknown cross-correlation coefficients at hourly levelestimated indirectly by using (7) and the values of m for all months are given in Table 3. The values of m were ranged between 1 and 2 different of what was reported in the study by Fytilas (2002), Koutsoyiannis(2003) and recently Bekele(2007).

Table 2 represents the values of m obtained from the hourly and daily cross correlations for each month based on data at Genting and Gombak stations

months	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
m	2.02	1.62	1.82	1.44	1.57	1.73	1.65	1.75	1.61	1.69	1.75	1.98

Table 3and Table4show the lag-zerocross-correlation coefficientsin the case when the hourly rainfall are available for all stations, and also the cross-correlation coefficientsin the case when estimated indirectly with different values of m for the months July, November respectively.The extracted cross-correlation coefficientsdepend on daily rainfall show higher values for close stations than the correlations depend on hourly.

The MuDRain model synthetic hourly statistics were given for the seven stations, for stations 1 and 2 are being identical to the historical values.Table 5 and Table 6 show the goodness of fit statistics, such as proportion of dryness, mean, standard deviation, skewness and lag-1 autocorrelationamong the historical, the theoretical and simulated statistics. These results show the MuDRain model reproduced proportion of dryness well for close stationsespecially whenextracted correlations are used. For the mean values are matched excellent for the two correlation methods. On the other hand, the model underestimates the standard deviation and skewness compared to historical and theoretical values for far stations for July and more significant difference notice for November. For lag-1 autocorrelation the model underestimated the correlations andwas showed also similar values for the two disaggregated rainfall for close stations.lag-zero cross-correlation coefficients and the extracted lag-zero cross-correlation coefficients for the seven stations of hourly data for July and November are also given in Table 3 and Table 4, we observed that correlation extracted are slightly higher than other those correlation based on hourly for November than July especially for rain gauges not so far from the referred station like Ampang station, is 28 km.

A graphical comparison of the distribution function of historical, synthetic, synthetic extracted hourly rainfall during wet days are given in Figure3 and Figure 4 for stations 3 and 5 for the months July and November respectively. The distribution function showed a good agreement between the synthetic and synthetic extracted distribution function corresponding the historical distribution function.Figure 5 and Figure 6depicta comparison in terms of autocorrelation function for higher lags, up to lag 10 for stations 3 and 5 for July and November. It is found the autocorrelation function for the MuDRain model corresponding to synthetic and synthetic extracted departs significantly from the historical for Julyand November at lag-1 for both stations 3 and 5. An additional graphical comparison, some hyetographs are given in Figure 7and Figure 8. We found that the MuDRain model did not reproduce well the actual hyetographs for July and November for stations 3 and 5. This discrepancies appear in all hyetographs that the MuDRain model generated low intensity at hours where the actual intensity was zero during rainy days or generated medium intensities at all day hours where the actual intensity was heavy rainy for some hours which corresponding to the time of occurrence of events for convective periods. The latter appears mostly in July and November for all day hours and this may be unavoidable because of the lowest cross correlation of hourly rainfall for all months.

V. CONCLUSION

Rainfall data are usually gathered at daily timescales due to the availability of daily rain-gauges throughout Malaysia.Rainfall data are often required at a finer scale such as hourly rather than daily for several applications.In this approach, disaggregation daily rainfall to hourly rainfallusing stochastic models formulated based on multivariate approach (MuDRain) is discussed in this paper.

Seven rain gauge stations are considered in this study for different distances from the referred station starting from 4 km to 160 km. For periods of 1973 to 2008, hourly rainfall data waspresented and July and November months wereconsidered as an example of dry and wet periods. This paper discussed the applicability the MuDRain model for disaggregation rainfall in cases where the hourly rainfall is available for the neighbouring stations or not.In addition, the cross-correlation among the raingauges wasfound for the available hourly rainfall information at the neighboring stationsand extracted correlations based on daily to generate spatially and temporally consistent hourly rainfall series at the raingauge of interest. The goodness of fit of the model was based on the reproduction of fitting statistics like the means, variances, coefficients of skewness, lag zero cross-correlation of coefficients and the lag one autocorrelation of coefficients. It is found the correlation coefficients based on extracted correlations are slightly higher than correlations based on available hourly rainfall especially for neighboring stations not more than 28 km like Ampang station. The results showed also the MuDRain model did not reproduce statistics well for stations further than 28 km from the referred stations, in addition to the time of coincidence between the historical and synthetic hourly rainfall where the actual intensity was rainy for some heavy hours which corresponding to the time of occurrence of events for convective periods.These discrepancies are unavoidable because of the lowest cross correlation of hourly rainfall.The overall performance indicated that the MuDRain model would not be appropriate choice for disaggregation daily rainfall.

Table 3Lag-zero and extracted Lag-zero cross correlation coefficients for the seven gauges at hourly level for the month of July

station	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Genting (historical)	historical	1.000	0.408	0.226	0.038	0.075	0.021	0.006
	extracted	1.000	0.439	0.196	0.032	0.042	0.016	0.029
	synthetic	1.000	0.408	0.360	0.079	0.100	0.038	0.018
	synthetic extracted	1.000	**0.408**	**0.336**	**0.045**	**0.083**	**0.022**	**0.054**

	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Gombak (historical)	historical	0.408	1.000	0.179	0.051	0.099	0.031	0.014
	extracted	0.439	1.000	0.199	0.045	0.056	0.038	0.031
	synthetic	0.408	1.000	0.324	0.089	0.144	0.078	0.039
	synthetic extracted	**0.408**	1.000	**0.328**	**0.075**	**0.125**	**0.093**	**0.048**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Ampang (historical)	historical	0.226	0.179	1.000	0.048	0.066	0.030	0.041
	extracted	0.196	0.199	1.000	0.082	0.066	0.038	0.063
	synthetic	0.226	0.179	0.270	0.058	0.058	0.014	0.022
	synthetic extracted	**0.226**	**0.179**	**0.256**	**0.048**	**0.066**	**0.046**	**0.023**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Petaling (historical)	historical	0.038	0.051	0.048	1.000	0.044	0.007	0.014
	extracted	0.032	0.045	0.082	1.000	0.018	0.003	0.018
	synthetic	0.038	0.051	0.019	0.196	0.017	0.001	0.011
	synthetic extracted	**0.038**	**0.051**	**0.021**	**0.216**	**0.023**	**0.002**	**0.015**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Seremban (historical)	historical	0.075	0.099	0.066	0.044	1.000	0.038	0.022
	extracted	0.042	0.056	0.066	0.018	1.000	0.037	0.026
	synthetic	0.075	0.099	0.060	0.022	0.188	0.020	0.014
	synthetic extracted	**0.075**	**0.099**	**0.094**	**0.029**	**0.154**	**0.019**	**0.039**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Kangsar (historical)	historical	0.021	0.031	0.030	0.007	0.038	1.000	0.121
	extracted	0.016	0.038	0.038	0.003	0.037	1.000	0.166
	synthetic	0.021	0.031	0.017	0.006	0.014	0.101	0.028
	synthetic extracted	**0.021**	**0.031**	**0.013**	**0.015**	**0.030**	**0.095**	**0.038**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Sitiawan (historical)	historical	0.006	0.014	0.041	0.014	0.022	0.121	1.000
	extracted	0.029	0.031	0.063	0.018	0.026	0.166	1.000
	synthetic	0.006	0.015	0.018	0.018	0.017	0.028	0.081
	synthetic extracted	**0.006**	**0.015**	**0.032**	**0.020**	**0.010**	**0.029**	**0.107**

The first row is the correlation coefficient obtained by historical hourly data ; the second row isthe extracted correlation coefficient obtained by historical daily data; the third row is thecorrelation coefficient obtained by the historical hourly data and synthetic data; and the fourth row is the correlation coefficient obtained by the historical hourly data and synthetic extracted.

Table 4Lag-zero and extracted Lag-zero cross correlation coefficients for the seven gauges at hourly level for the month of November

station	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Genting (historical)	historical	1.000	0.409	0.225	0.087	0.054	0.035	0.040
	extracted	1.000	0.447	0.274	0.057	0.024	0.004	0.008
	synthetic	1.000	0.409	0.308	0.096	0.062	0.034	0.052
	synthetic extracted	1.000	**0.409**	**0.434**	**0.073**	**0.021**	**0.019**	**0.005**

	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Gombak (historical)	historical	0.409	1.000	0.251	0.119	0.067	0.018	0.034
	extracted	0.447	1.000	0.220	0.085	0.025	0.019	0.004
	synthetic	0.408	1.000	0.337	0.145	0.069	0.034	0.039
	synthetic extracted	**0.408**	1.000	**0.286**	**0.111**	**0.033**	**0.044**	**0.004**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Ampang (historical)	historical	0.225	0.251	1.000	0.122	0.090	0.021	0.052
	extracted	0.274	0.220	1.000	0.121	0.049	0.013	0.014
	synthetic	0.224	0.251	0.229	0.092	0.043	0.020	0.017
	synthetic extracted	**0.224**	**0.251**	**0.238**	**0.062**	**0.020**	**0.025**	**0.026**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Petaling (historical)	historical	0.087	0.119	0.122	1.000	0.159	0.013	0.029
	extracted	0.057	0.085	0.121	1.000	0.105	0.023	0.008
	synthetic	0.086	0.119	0.058	0.186	0.046	0.035	0.007
	synthetic extracted	**0.086**	**0.119**	**0.052**	**0.183**	**0.026**	**0.030**	**0.001**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Seremban (historical)	historical	0.054	0.067	0.090	0.159	1.000	0.018	0.014
	extracted	0.024	0.025	0.049	0.105	1.000	0.001	0.000
	synthetic	0.054	0.067	0.034	0.064	0.129	0.009	0.004
	synthetic extracted	**0.054**	**0.067**	**0.050**	**0.057**	**0.079**	**0.000**	**0.004**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Kangsar (historical)	historical	0.035	0.018	0.021	0.013	0.018	1.000	0.201
	extracted	0.004	0.019	0.013	0.023	0.001	1.000	0.004
	synthetic	0.034	0.018	0.017	0.017	0.001	0.067	0.021
	synthetic extracted	**0.034**	**0.018**	**0.032**	**0.020**	**-0.006**	**0.060**	**0.040**
	correlation	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Sitiawan (historical)	historical	0.040	0.034	0.052	0.029	0.014	0.201	1.000
	extracted	0.008	0.004	0.014	0.008	0.000	0.004	1.000
	synthetic	0.041	0.034	0.022	0.020	0.000	0.014	0.118
	synthetic extracted	**0.041**	**0.034**	**0.016**	**0.009**	**-0.005**	**0.014**	**0.078**

The first row is the correlation coefficient obtained by historical hourly data ; the second row isthe extracted correlation coefficient obtained by historical daily data; the third row is thecorrelation coefficient obtained by the historical hourly data and synthetic data; and the fourth row is the correlation coefficient obtained by the historical hourly data and synthetic extracted.

Table 5 Statistics of hourly rainfall depths at each gauge for the month of July

gauge	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Proportion dry							
value used on disaggregation	0.944	0.944	0.944	0.944	0.944	0.944	0.944
historical	0.948	0.938	0.933	0.954	0.931	0.959	0.959
synthetic	0.946	0.938	**0.946**	**0.836**	**0.920**	**0.938**	**0.914**

synthetic extracted	0.946	0.938	**0.940**	**0.851**	**0.919**	**0.940**	**0.909**
Mean							
value used on disaggregation	2.244	2.244	2.244	2.244	2.244	2.244	2.244
historical	2.081	2.408	0.205	0.963	0.217	0.143	0.142
synthetic	2.077	2.408	0.205	0.963	0.217	0.143	0.142
synthetic extracted	2.077	2.408	0.205	0.963	0.217	0.143	0.142
Standard deviation							
value used on disaggregation	18.230	18.230	18.230	18.230	18.230	18.230	18.230
historical	17.953	18.507	1.854	11.592	2.205	1.409	1.526
synthetic	17.938	18.505	**1.706**	**4.755**	**1.736**	**1.446**	**1.146**
synthetic extracted	17.938	18.505	**1.759**	**5.149**	**1.596**	**1.245**	**1.095**
Skewness							
value used on disaggregation	14.979	14.979	14.979	14.979	14.979	14.979	14.979
historical	16.262	13.693	16.098	27.091	27.069	17.720	21.237
synthetic	16.295	13.697	**16.097**	**15.343**	**34.188**	**30.568**	**20.791**
synthetic extracted	16.295	13.697	**18.261**	**20.767**	**21.426**	**23.634**	**22.036**
Lag-1 autocorrelation							
value used on disaggregation	0.393	0.393	0.393	0.393	0.393	0.393	0.393
historical	0.372	0.414	0.331	0.280	0.329	0.405	0.330
synthetic	0.372	0.414	**0.207**	**0.581**	**0.231**	**0.154**	**0.229**
synthetic extracted	0.372	0.414	**0.205**	**0.556**	**0.217**	**0.166**	**0.19**

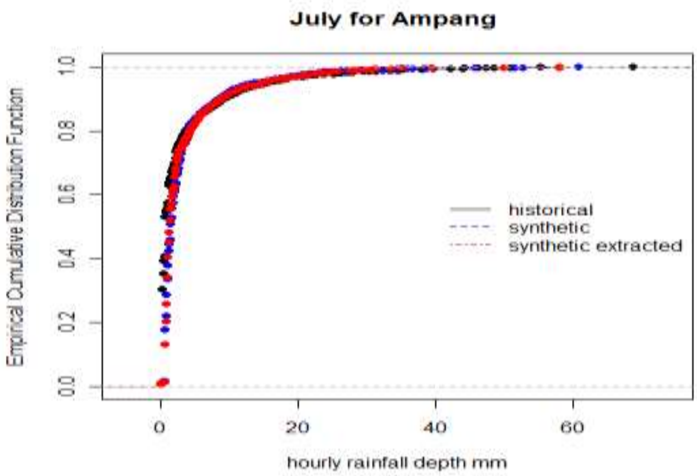
The first row is the value used in the disaggregation model, which is the historical value of stations 1 and 2; the second row isthe historical value, not used in the disaggregation model (apart from values of stations 1 and 2); the third row is the synthetic value; and the fourth row is the synthetic extracted.

Table 6 Statistics of hourly rainfall depths at each gage for the month of November

gauge	Genting	Gombak	Ampang	Petaling	Seremban	Kangsar	Sitiawan
Proportion dry							
value used on disaggregation	0.926	0.926	0.926	0.926	0.926	0.926	0.926
historical	0.925	0.926	0.901	0.920	0.922	0.887	0.921
synthetic	0.922	0.922	**0.917**	**0.708**	**0.927**	**0.874**	**0.858**
synthetic extracted	0.922	0.922	**0.916**	**0.713**	**0.920**	**0.865**	**0.834**
Mean							
value used on disaggregation	2.140	2.140	2.140	2.140	2.140	2.140	2.140
historical	2.147	2.133	0.267	1.978	0.215	0.331	0.407
synthetic	2.147	2.133	0.267	1.978	0.215	0.331	0.407
synthetic extracted	2.147	2.133	0.267	1.978	0.215	0.331	0.407
Standard deviation							
value used on disaggregation	15.940	15.940	15.940	15.940	15.940	15.940	15.940
historical	16.618	15.262	2.054	14.587	1.603	3.344	4.699
synthetic	16.617	15.258	**2.069**	**6.406**	**1.680**	**1.914**	**2.850**
synthetic extracted	16.617	15.258	**1.916**	**6.729**	**1.551**	**1.770**	**2.880**
Skewness							
value used on disaggregation	14.892	14.892	14.892	14.892	14.892	14.892	14.892

historical	16.568	13.213	14.656	17.298	15.033	54.987	33.042
synthetic	16.573	13.219	**17.821**	**11.391**	**16.158**	**16.743**	**23.281**
synthetic extracted	16.573	13.219	**13.966**	**12.501**	**15.294**	**13.561**	**26.729**
Lag-1 autocorrelation							
value used on disaggregation	0.377	0.377	0.377	0.377	0.377	0.377	0.377
historical	0.343	0.411	0.29	0.357	0.385	0.393	0.378
synthetic	0.343	0.411	**0.273**	**0.409**	**0.145**	**0.456**	**0.108**
synthetic extracted	0.343	0.411	**0.219**	**0.42**	**0.127**	**0.575**	**0.09**

The first row is the value used in the disaggregation model, which is the historical value of stations 1 and 2; the second row isthe historical value, not used in the disaggregation model (apart from values of stations 1 and 2); the third row is the synthetic value; and the fourth row is the synthetic extracted.



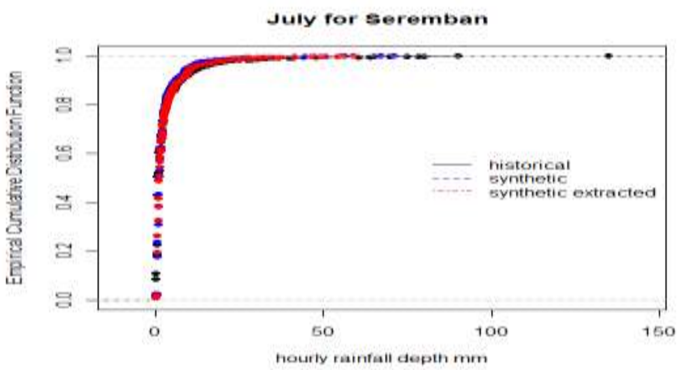


Fig.3: Comparison of the distribution function for historical, synthetic and synthetic extracted hourly rainfall for July for Ampang and Seremban stations

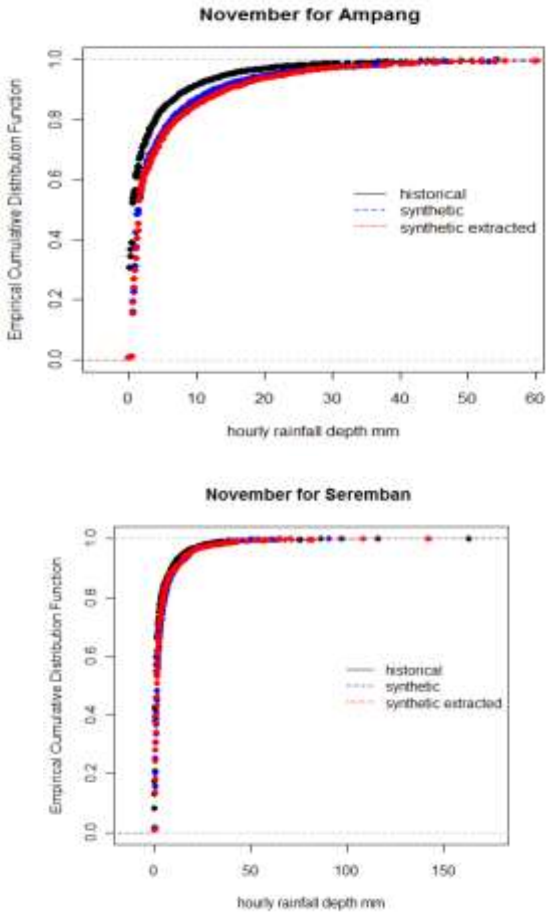


Fig.4: Comparison of the distribution function for historical, synthetic and synthetic extracted hourly rainfall for July for Ampang and Seremban stations

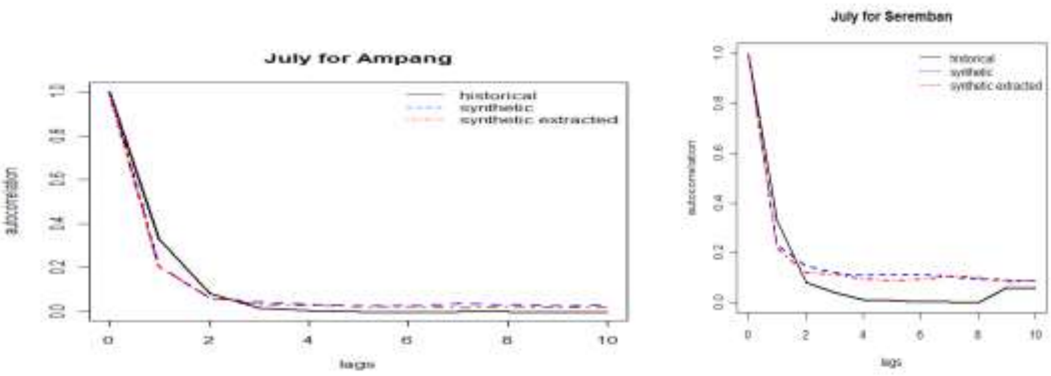


Fig.5: Comparison of autocorrelation function for historical, synthetic and synthetic extracted hourly rainfall for July for Ampang and Seremban stations

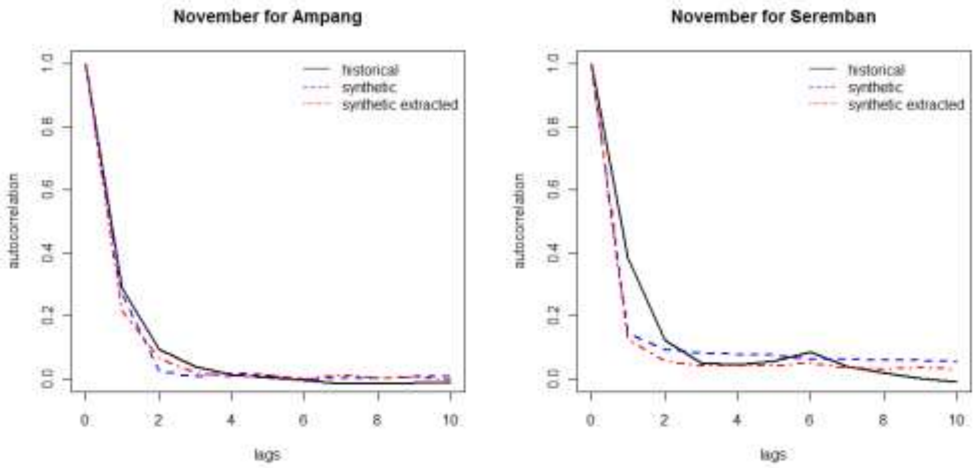


Fig.6: Comparison of autocorrelation function for historical, synthetic and synthetic extracted hourly rainfall for July for Ampang and Seremban stations

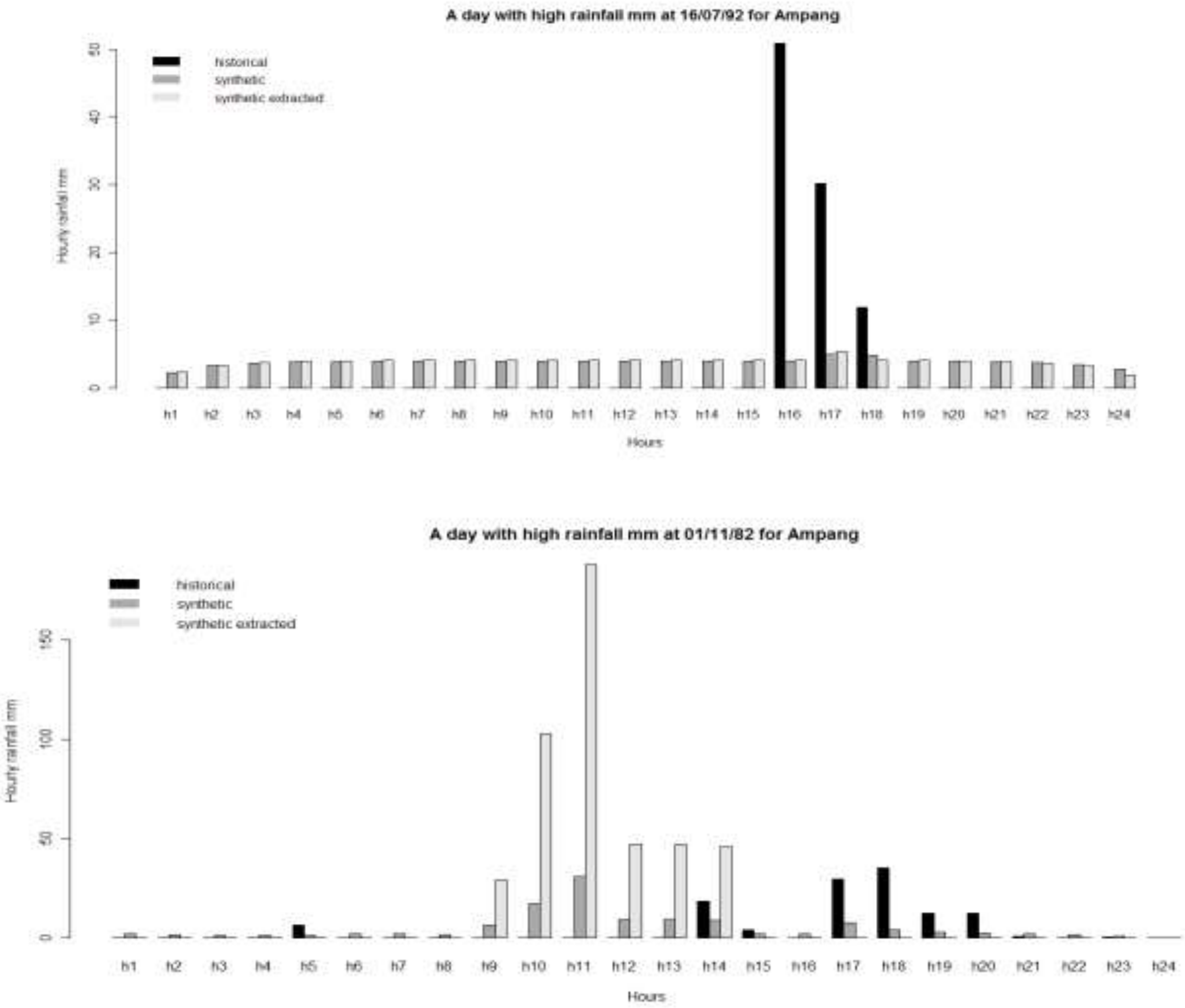
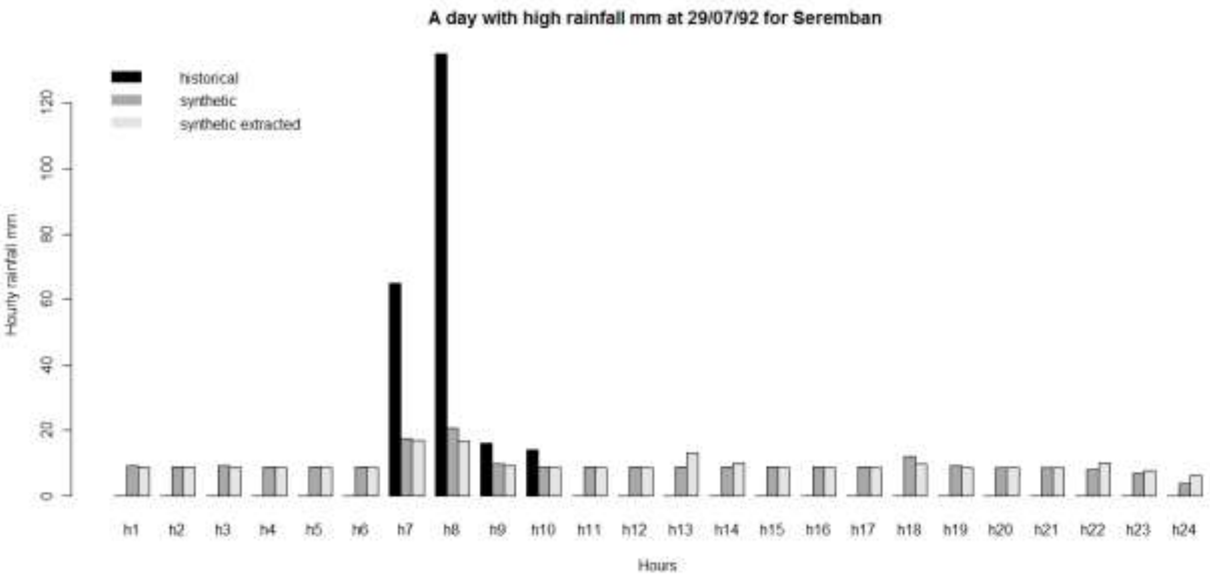


Fig.7: Comparison of hyetographs for historical, synthetic and synthetic extracted hourly rainfall for July and November for Ampang station



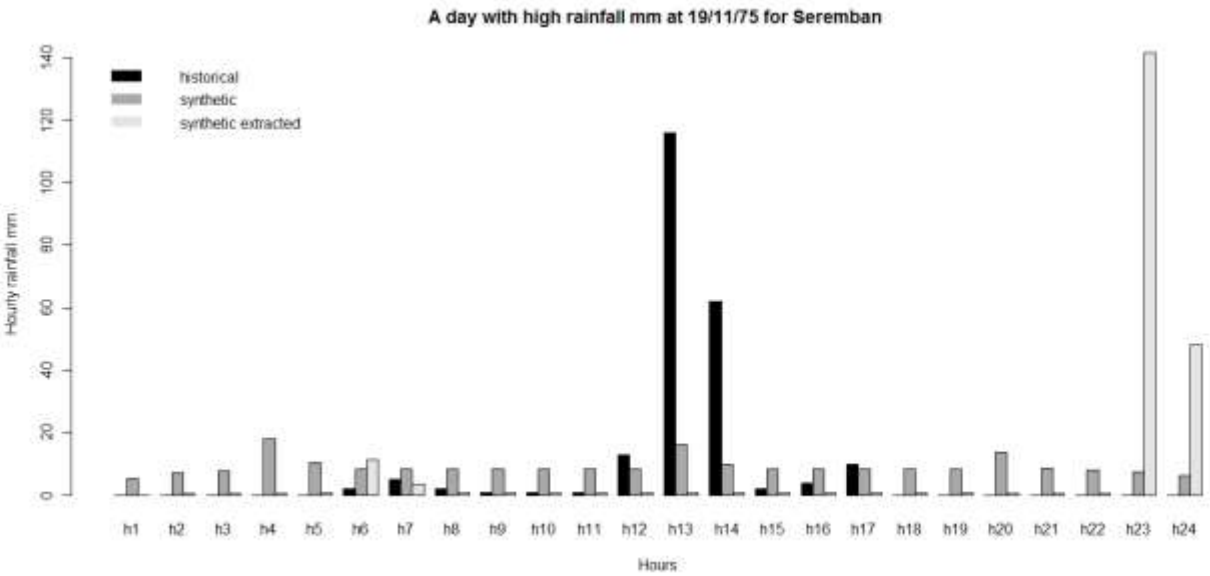


Fig.8: Comparison of hyetographs for historical, synthetic and synthetic extracted hourly rainfall for July and November for Seremban station

REFERENCES

[1] Bekele D, R. S, J. YP (2007) Accuracy evaluation of weather data generation and disaggregation methods at finer timescales. Advances in Water Resources 30:1286-1300
[2] Fytilas P (2002) Multivariate rainfall disaggregation at a fine time scale. Universita degli Studi di Roma "La Sapienza", Roma
[3] Hoshi K, Burges SJ (1979) Disaggregation of streamflow volumes. Journal of the Hydraulics Division, Proceedings ASCE 105 (HY1): 27-41
[4] Koutsoyiannis D (1994) A stochastic disaggregation method for design storm and flood synthesis. Journal of Hydrol 156:193-225
[5] Koutsoyiannis D (1999) Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. Water Resour Res 35 (4):1219-1229
[6] Koutsoyiannis D (2001) Coupling stochastic models ofdifferent time scales. Water Resour Res 37 (2):379-392
[7] Koutsoyiannis D (2003) Multivariate disaggregation of rainfall-MuDrain www.itia.ntua.gr.

[8] Koutsoyiannis D, A. M (1996) Simple disaggregation by accurate adjusting procedures. Water Resour Res 32 (7):2105-2117
[9] Koutsoyiannis D, Onof C (2000) HYETOS a computer program for stochastic disaggregation of fine scale rainfall. Available from http://www.itia.ntua.gr/e/softinfo/3/>.
[10] Koutsoyiannis D, Onof C (2001) Rainfall Disaggregation Using Adjusting Procedures on a Poisson Cluster Model. Journal of Hydrology 256:109-122
[11] Koutsoyiannis D, Onof C, Wheater HS (2003) Multivariate rainfall disaggregation at a fine time scale. Water Resour Res 39 (7):1-18
[12] Koutsoyiannis D, Xanthopoulos T (1990) A dynamic model for short scale rainfall disaggregation. Hydrol Sci J 35 (3):303-322
[13] Lane WL (1979) Applied Stochastic Techniques, User's Manual. In: Reclamation Bo (ed) Engineering and Research Center, Denver, Co.,
[14] Lane WL (1982) Corrected parameter estimatesfor disaggregation schemes, in Statistical Analysis of Rainfall and Runoff. Water Resources Publications, Littleton, Colo
[15] Mejia JM, Rousselle J (1976) Disaggregation models in Hydrology revisited. Water Resour Res 12 (2):185-186
[16] Pereira MVF, Oliveira GC, Costa CCG, Kelman J (1984) Stochastic streamflow models for hydroelectric systems. Water Resour Res 20 (3):379-390
[17] Salas J (1993) Analysis and Modeling of Hydrologic Time Series, vol chapter 19. McGraw-Hill, vol handbook of hydrology. New York
[18] Salas JD, Delleur JW, Yevjevich V, Lane WL (1980) Applied Modeling of Hydrologic Time Series. Water Resources Publications, Littleton, Colo
[19] Schaake JC, Ganslaw MJ, Fothergill JW, Harbaugh TE Multivariate rainfall generator for annual, seasonal, monthly and daily events In: Proceedings of the International Symposium on Mathematical Modeling Techniques in Water Resources Systems, Ottawa, Canada, 1972. pp 437-460
[20] Socolofsky S, Adams E, Entekhabi D (2001) Disaggregation of daily rainfall for continuous watershed modeling. J Hydrol Eng, ASCE 6 (4):300-309
[21] Stedinger JR, Pei D, Cohn TA (1985) A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. Water Resour Res 21 (5)
[22] Stedinger JR, Vogel RM (1984) Disaggregation procedures for generating serially correlated flow vectors. Water Resour Res 20 (1):47-56
[23] Tao PC, Delleur JW (1976) Multistation, multiyear synthesis of hydrologic time series by disaggregation. Water Resour Res 12 (6):1303-1312
[24] Todini E (1980) The preservation of skewness in linear disaggregation schemes. J Hydrol 47:199-214
[25] Valencia D, Schaake JC (1972) A disaggregation model for time series analysis and synthesis. Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics. Massachusetts Institute of Technology, Cambridge, Mass.
[26] Valencia D, schaake JC (1973) Disaggregation processes in Stochastic Hydrology. Wat Resour Re 9 (3):211-219
[27] Woolhiser DA, Osborn HB (1985) A stochastic model of dimensionless thunderstorm rainfall. Water Resour Res 21 (4):511-522