# Evaluating Error of Temporal Disaggregation from Daily into Hourly Rainfall using Heytos Model at Sampean Catchments Area

Entin Hidayah<sup>1</sup>, Nadjadji Anwar<sup>2</sup>, Edijatno<sup>2</sup>, and Nur Iriawan<sup>3</sup>

*Abstract***Developing a rainfall-runoff model sufficient to flood prediction hourly rainfall data. Lack of automatic rain gauge for high resolution rainfall in catchment area can be an obstacle for the modeling. Otherwise, the manual rain gauges may spread on all of catchments areas, providing daily rainfall. Daily rainfall disaggregation to hourly rainfall is an innovation to get higher temporal resolution of the rainfall. This paper attempts to evaluate the implementation of rainfall disaggregation model in Sampean Catchments Area using Heytos. The proposed parameter optimation use Moment Performance model that tested by calibrating it with available hourly data. The results of model indicated that only data within five months had good performance. The estimation result showed that relative error total of January, February, August, November, and December was less than one. In case of March, April, May, June, July, September, and October the model could not result respectively to generate model.**

## *Keywords***disaggregation rainfall, daily, hourly, Heytos**

# I. INTRODUCTION

he high resolution rainfall is a necessary input for The high resolution rainfall is a necessary input for flood simulation in hydrology modeling. There are two important reasons to support the use of it i.e. to anticipate the flood occurrence and to minimize the model uncertainty. First, the difficulty to make the rainfall runoff model is caused by the lack of data hourly rainfall. In fact, it is showed that the high resolution rainfall data is limited. As described in Sampean catchment area  $(1057 \text{km}^2)$ , it only has three automatic rain gauges (hourly rainfall), but they are 17 manual rain gauges (daily rainfall) spreading over in the catchment area (Fig.1). Therefore, availability of hourly rainfall data do not represent the hourly rainfall distributed in the catchments area. Second, supporting the first reason for the need data on high resolution rainfall, based on Carpenter and Georgakakos' (2006) studies, the rainfall modeling employing high resolution rainfall data will produce smaller error of model. Considering those two reasons, a generate model for hourly rainfall is needed for facing the problem in Sampean catchments area.

Some research developments to generate hourly rainfall have been done in many countries; they are popularly called temporal rainfall disaggregation. On of and Arrnbjerg-Nielsen (2009) have constructed this kind

of model. This model was based on random parameter of Bartlett-Lewis rectangular pulse rainfall combined with multi-scale disaggregation applied to an urban area in Denmark. The result of performance model increased 2 - 15% to extreme rainfall condition. Fytilas (2002) has applied the disaggregation rainfall model in Basin of Tiber River in Italy. Disaggregation was performed using hourly data of three rain gauges with Heytos model. This model had a good reproduction of the actual hyetographs and the performance which were much better for the months characterized by a wet regime. Koutsoyiannis and Onof (2001) have developed rainfall disaggregation using adjusting procedures on a Poisson cluster model and examined for hourly data in UK and US. The results indicated good performance methodology.



Fig. 1. Location of selection station in Sampean catchments area used as case studies

Disaggregation rainfall model is applied successfully in several countries; nevertheless, in Indonesia this model is ever implemented. Considering the less of the hourly rainfall data to flood design in Indonesia, this model is necessary to be implemented in Indonesia.

#### II. THEORY

# *A. Disagregation Rainfall Model*

Disaggregation is an important step in the process to obtain lower-level time scale data from higher time scales, i.e. from annually to monthly, monthly to daily or daily to hourly.

The disaggregation is performed on the most promising sets of parameters and is done as a check on the disaggregation procedure. Disaggregation is resulted from a simulation. The statistical generated model may fluctuate

 $\frac{1}{1}$ <sup>1</sup>Entin Hidayah is Student of Civil Engineering Doctorate Program, FTSP, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: entin\_hidayah@yahoo.com.

<sup>&</sup>lt;sup>2</sup>Nadjadji Anwar and Edijatno are with Department of Civil Engineering, FTSP, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: nadjadji@ce.its.ac.id, edidjatno@ce.its.ac.id.

<sup>3</sup> Nur Iriawan is with Department of Statistics, FMIPA, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia. E-mail: nur\_i@statistika.its.ac.id.

depending on the parameters used, as well as the random seed utilized in the simulation.

Therefore, some variations from the historical and modeled statistics are to be expected. Using the simulation also allows us to obtain statistics from the time series that are not explicitly modeled. Higher order statistics, such as the skewness, can be calculated from the simulated disaggregated time series, and this can be compared with the historical data to observe the goodness of it.

There are some stochastic disaggregation rainfall models i.e. BLRPM and Neyman–Scott rectangular pulse (NSRPM), that have shown a result for rainfall characteristic model of 1-24 hour range (Smithers. et al., 2002). Bartlett-Lewis Rectangular Pulse Models (BLRPM) is able to reproduce important feature of rainfall field from the hourly to the daily scale or above, so it was chosen in this research.

## *B. Bartlett-Lewis Rectangular Pulse Models (BLRPM)*

The Bartlett-Lewis Rectangular Pulse Model (BLRPM) is a continuous time rainfall model for a fixed point in space. The BLRPM is a model for point rainfall time series modeling rainfall at a point, as opposed to spatial General Circulation models which model rainfall over an entire area. The model incorporates Poisson cluster processes. This model has been used with considerable success for a wide variety of climates, including the U.K. [1, 2]. The model had proven useful for reproducing the statistics of both daily and sub-daily time scales (Rodriguez-Iturbe et al., 1987, 1988. Demetris et al. (2001) succeeded using this model that was formed in Heytos computer program.

The description of model phenomena for *BLRPM* is storm origins *ti* following a Poisson process (rate *λ*), cell origins *tij* arriving in Poisson process (rate *β*) started each *ti*, cell arrivals terminate after a time *xi* with distribution exponential (parameter  $\gamma$ ) [11, 12]. Each cell has duration *wij* exponentially distributed (parameter *η*). To search a uniform intensity  $P_{ij}$  with a specified distribution as typically assumed exponential (parameter  $1/\mu_x$ ) or other alternative 2-parameter gamma with mean  $\mu_x$  and expected mean square error (EMS) of cell intensity  $\mu_x^2$ . Mean of number cells per storm is:

$$
\mu = 1 + - \tag{1}
$$

Where the equation of k is:

$$
= 2 \ \mu\mu + \frac{\mu\mu}{-1} \ [-1] \\
= \frac{\mu\mu}{-1} \ [-1] \\
= \frac{\mu}{-1}
$$

Parameters used in the temporal model are as follows (Wheater, et al, 2005): (1) Bartlett Lewis model uses 2 parameter cell intensity distribution. (2). Neyman–Scott model employs 2-parameter cell intensity distribution and Poisson number of rainfall cell. (3). Random parameter Bartlett Lewis model utilizes 1-parameter cell intensity distribution. (4). Random parameter *Bartlett*  Lewis model exploits 2-parameter cell intensity distribution. (5) Bartlett Lewis model uses 2-cell types, each with 2-parameter intensity distribution. From the comparison of those model alternatives, according to Wheater (2005), the third model is the most applicable. Wong (2000) had examined the use of the four and six parameters. Using six parameters gives better results than four parameters.

Equation used for statistical model within six parameters by Demetris (2001) can be calculated as follows (equation 2-5):

mean :

$$
= -\mu\mu \qquad \mu = + \qquad (2)
$$

variance :

$$
\frac{v}{1 + \frac{1}{(1 + \
$$

autocovariance :

+ (1 −ݏ)ܶ}(ଷିఈ)(ଶିఈ) = (ݏ − ݈݃) ݅ܽ݊ܿ݁ݎܽݒܿݐݑܽ ߥ(ଷିఈ+ ܶ(ݏ+ 1 (+ ߥ (ଷିఈ− 2(ܶݏ+ ߥ(ଷିఈ} + (ఈିଶ)(ఈିଷ){2(߶ܶݏ+ ߥ(ଷିఈ− [߶ܶ(ݏ− 1 (+ ߥ[ଷିఈ− ఈିଷ] ߥ + (1 +ݏ)ܶ߶] (4)

$$
Pr(zero\text{-}rain) = exp(-\lambda T - f_1 + f_2 + f_3)
$$
\n<sup>(5)</sup>

$$
= \frac{(-1)^{n} (1 + 1)^{n} (-1)^{n} (1 + 1)^{n} (-1)^{n} (-1)^{n
$$

*C. Heytos* 

Heytos is a computer programme that is used to perform the rainfall disaggregation. This piece of software was produced by Koutsoyiannis and Onof (2001). Heytos used the proportionate adjusting procedure. The input is required in Heytos i.e. the six-parameters from the BLRPM and the actual historical rainfall time series.

The current software version does not support estimation of the Bartlett-Lewismodel*.* The result of running model can be compared with the disaggregated simulation and historical statistics. The output of model gives the fully calculated statistics of the hourly time series as well as the simulated time series obtained. The result of disaggregation from Heytos could be evaluated statistically by graph. Graph properties in Heytos consist of autocorrelation, marginal statistic, probabilities of dry spells, hydrograph, and storm. Heytos is essentially a simulation, and therefore, it is expected that the final results will resemble the modeled statistics since the two processes are based on the same set of BLRPM parameters.

#### *D. Parameter Estimation*

There are several methods of fitting the BLRPM model to the historical statistics i.e. Maximum Likelihood, Spectral, and Moments. The Maximum Likelihood method, although very commonly used in other models for parameter fitting, is unwieldy in this case, as the likelihood function is difficult to obtain [11]. The Spectral Method incorporating Fourier analysis can also be used. However, the Method of Moments is a better option, as it is significantly simpler and more practical to use. Also, the method of moments has been found to produce parameters that are significantly better than using spectral methods [11]. The method of moments is set out simply as follows. More complex mathematical formulations of this method are set out in Rodriguez-Iturbe et al. (1988) and Cryer (1986).



Fig. 2. Rainfall series in 2005-2008

The set of parameters to be fitted was given by the set Θ, where = {*λ, μx, κ, φ, α, ν*} for the six parameter model. Let *k* be the number of parameters to be fitted, either six. Next, *p* statistics was chosen from the historical data to fit the parameters, and these were denoted by the set *T*, where  $T = \{t1, t2, t3, \ldots t_p\}$ . These could include the mean, variance, etc. of various time scales. The functions from which the various statistics could be calculated from the parameter values in the BLRPM (Equations 2- 5) denoted by the set *S*, where  $S = \{sI(\Theta), s2(\Theta), s3(\Theta)\}\$ …… *sp*(Θ)}.

If  $k = p$ , then the method of moments requires:  $S = T \forall p$  (9)

This set of *p* equations was then solved for Θ, obtaining the parameter set. However, the functions that were within the set *S* were often highly non-linear; therefore, it was difficult to obtain the Θ explicitly. In order to make it easier in optimization of an error-residual form, numerical methods were used. The objective function of optimization was to find a set of parameters Θ where the expression in (6) is equal to zero, such that *si(*Θ*)=ti* for all *i=*{1, 2, 3, …..p}.

 $\text{Min} \sum \left( \begin{array}{c} (-1)^{n-1} \\ (-1)^{n-1} \\ (-1)^{n-1} \end{array} \right)$  (10)

Where *wi* was the weight attributed to that particular statistic. These weights are usually set to unity in the following optimization schemes. To optimize error value is used The SOLVER function within Microsoft Excel 2007.

#### III. METHOD

Method used in the research was identifying the parameter model the rainfall modeling using the synopsis of Bartlett-Lewis. The proposing of this modeling would be coupled with Heytos approaches to model the temporal disaggregation rainfall data.

Steps to the research are:

- 1. searching for the relation between daily rainfall and total hours in a day,
- 2. searching the characteristic hourly rainfall to 1 hour, 24 hours and 48 hours,
- 3. setting the model parameter,
- 4. entering loading and compiling of the rainfall characteristic data,
- 5. loading and generating initialized,
- 6. optimizing parameter model,
- 7. running Heytos model, and

8. evaluating model performance output.

This research was conducted in the Sentral rainfall station of Sampean Watershed in Bondowoso Regency.



Fig. 3. Comparison of rainfall data between daily and total hourly in 2004-2008

This station has ARR station and manual rainfall station. Hourly rainfall data used were those of the periods 2005-2008 (total number of data was 2952).

## IV. RESEARCH RESULT

The characteristics of climate condition in Sentral rain gauge at Sampean Catchments area in 1991-2008 showed that wet months for this region were considered to start on October and summer months May as shown on Fig. 2. The mean annual rainfall depth for Sentral rainfall station was about 1503 mm and the mean monthly comparison of rainfall data between daily and total hourly (accumulated between in 2005-2008) shows a good agreement, so the daily data ready to make rainfall model.

Value comparison of mean, variance, auto correlation, and proportion dry between the model and historic statistics from January until December can be shown in Table 1 and Table 2. Evaluating for Table 1 could be explained as follow. In general, comparison between mean and variance values show that means value is lower than that of variance value. This means that distribution of hourly rainfall data is skewness. The values of mean have a good performance which is less difference of simulation and history value. However, the comparison of simulation and history value from variance values is not as good as mean value. The differences of simulation and history value from variance, there are over and under estimate that causes of bias parameter.

In Table 2, there are negative value and bigger value differences between historical and simulated value of autocovariance of 24 hr and 48 hr at March, April, Mei, June, July, and October. If auto covariance values are negative, sum of weighted squared errors are resulted a bigger. Variances and autocovariance of 1 hr, 12 hr, and 24 hr show the worst results, with significant variation in the variances and autocovariance. This could indicate that the characteristics of the 24 and 48-hr data were not preserved when aggregation occurred. Proportion dry of 1, 24, and 48 hr shows that differences found in values between historical and simulated were less if compared with the value of variance and auto correlation. The values of overestimate and underestimate are expressed in mean, variance, auto covariance, and proportion dry, so it makes difficulties to obtain optimum value in estimation parameter.

TABLE 2.



The parameters in Table 3 from January until December were resulted from optimization from histories and modeled simulation which were obtained from the lowest local optimum. The parameters were inputs of Heytos program that used to disaggregate the best sets of parameters for each month. Constraints of each parameter for lower level were 1E-07 and those of upper level were 99 in Table 4. The parameter value of each month in Table 3 respectively. Result of estimating parameter shows the resulting sum of weighted squared errors each month from January until December in Table 5 respectively. The better sum of weighted squared errors was in January, February, August, November, and December, whose value is less than one. These months have a good performance of the model implementation because the rainfall data distributions tend to gamma structure (in form of gamma).

For other months, the value of sum of weighted squared errors between the calculated form of statistics (*si*(Θ)) and the actual historical data (*ti*) for the rainfall disaggregation model in Sampean catchments area indicate slightly less than the statistic of less interest, because the distribution tend to normal. In general, distribution of rainfall data on tropical area varies very much, so Heytos program cannot fully be applied for tropical area. In sub tropical area, based on previous research, (Koutsoyiannis and Onof (2001)) model Heytos program has been succesfully applied since the condition of rainfall distribution in the area tends to have Gamma or exponential distribution which is in accordance with the distribution used in Heytos program.

Results from running Heytos show that the disaggregation for January and December followed very closely the results from using the modeled statistics (Fig.4). However, there seems to be small improvement in the estima-

TABLE 1.

tion of the statistics. The simulated statistics tended to lie closer to the modeled statistics rather than the historical statistics.



In summary, disaggregation gives a close fit with the modeled and historical data, hence if a good set of parameters can be found from the 24-hr or 48-hr data, the 1 hr statistics can be estimated by disaggregation to high accuracy. An hourly time series can therefore be derived from daily data. Other statistics, such as skewness and lag-n autocorrelations also can be estimated accurately, providing an excellent set of used BLRPM parameters. Result of comparison between original and disaggregation data for December show that error value from Mean Absolute Error (MAE) is 0.516 respectively. This error value is caused by inaccuracy on disaggregation of hourly period. Otherwise, results shown in the Fig. 5 in July that the simulated statistics show less accurate estimates of the historical data for both the dry and the wet period statistics. The autocovariance for lag periods of more than one are not also closely estimated. In particular, the skewness is also not close to estimation, showing rather inaccurate results, even though this statistic was not explicitly modeled in the BLRPM*.*

Error value that is produced from MAE evaluation is 0.022. This value of July much lower small than December, otherwise the parameter performance of July better than December. That is caused by dry season in July and rainy season in December.

## V. CONCLUSION

Model Heytos is appropriate to disaggregate rainfall in sub-tropic region which has gamma-formed rainfall distribution.

The result of temporal disaggregation model Heytos for tropical region has a good performance for model parameter to generate the rainfall characteristic data which have gamma distribution (such as for January, February, August, November, and December).

Otherwise for the other months for the region, the temporal aggregation model has poor performance, so it may a method to improve statistical model such as on estima-

te parameter optimation, data clustering or using other time series model. TABLE 4.





#### **REFERENCES**

- [1] Onof and H. S. Wheater, 1993, "Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model", *Journal of Hydrology, 149*, 67- 95.
- [2] C .Onof and H. S. Wheater, 1994, "Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model", *Journal* of Hydrology, 157, 177-195.
- [3] Onof and K. Arrnbjerg-Nielsen, 2009, "Quantification of anticipated future changes in high resolution design rainfall for urban area", *Atmospheric Research*, 92, 350-363.
- Koutsoyiannis and A. Manetas, 1996, "Simple disaggregation by accurate adjusting procedures", *Water Resour. Res.*, 32(7) 2105- 2117.
- [5] Koutsoyiannis and C. Onof, 2001, "Rainfall disaggregation using adjusting procedures on a Poisson cluster model", *Journal of Hydrology,* 246, 109-122.
- [6] J. D. Cryer, 1986, *Time Series Analysis*, Massachusetts: PWS Publishers.
- [7] J. C. Smithers, G. G. S. Pegram, and R. E. Schulz, 2002, "Design rainfall estimation in South Africa using Bartlett-Lewis rectangular pulse rainfall models", *Journal of Hydrology,* 258, 83-99.
- [8] K. M. Wong, 2000, Disaggregation of rainfall time series using adjustments.
- [9] P. Fytilas, 2002, Multivariate Rainfall Disaggregation at a Fine Time Scale Diploma Thesis Submitted at the University of Rome "La Sapienza."
- [10] R. Chandler and H. Wheater, 1998, "Climate change detection using generalized linear models for rainfall, A case study from the West of Ireland, II, Modelling of rainfall amounts on wet days", *Technical report, no. 195, Department of Statistical Science, University College London,* (http://www.ucl.ac.uk/ Stats/research/abstracts.html).
- [11] Rodriquez-Iturbe, D. R. Cox, and V. S. Isham, 1987, "Some models for rainfall based on stochastic point process", *Proc R Soc Land A* 410; 269-288.
- [12] Rodriquez-Iturbe, D. R. Cox, and V. S. Isham, 1988, "A point process model: Futher developments", *Proc R Soc Lond A* 417;283-298.
- [13] T. M. Carpenter and K. P. Georgakakos, 2006, *Intercomparison of Lumped versus Distributed Hydrologic Model Ensemble Simulations on Operational Forecast Scales.*
- [14] H. S. Wheater, et.al., 2005, "Spatial temporal rainfall modeling for flood risk estimation*"*. *Stochastic Environmental Research and Risk Assessment,* 19, 403–416.

Autoco



Fig. 4. Results from Heytos for disaggregation with BLRPM



Fig. 5. Results from Heytos for disaggregation with BLRPM parameters optimized for July