Implementing of Temporal Rainfall Disaggregation Model Using Bayesian PAR1 Model Combined with Adjusting and Filtering Procedure in Sampean Catchments Area

Entin Hidayah

Civil Engineering, Jember University.

ABSTRACT

Temporal Rainfall disaggregation model is one ways to cope the lack of high resolution rainfall (hourly rainfall) in single site. Design of rainfall disaggregation model using Bayesian PAR1 model combined with adjusting and filtering procedure had been established in Sentral Station at Sampean Catchments Area. This paper attempts to implement the model with different time variable. Performance of model will be evaluated based on the lowest value of mean absolute error(MAE), root mean square error (RMSE) and mean square error (MSE).

Key word: Temporal, Rainfall, Disaggregation, Bayesian, PAR1.

Introduction

Almost regions in Indonesia is found manual rain gauge (daily) operated with a fairly long period. On the other hand, the presence of automatic rain gauge (hourly) is slightly a little with relatively shorter period of operation. Lack of high resolution rainfall data is an impediment in hydrologic modeling for forecasting flood. An innovation for disaggregation from daily to hourly rainfall is urgently needed to meet the high resolution rainfall.

Disaggregation rainfall model is a method to transform the synthetic hourly rainfall (lower scale) derived from the rainfall data at higher scale (daily, or weekly). The formation of this synthetic data is derived from the rise of a stochastic model (Hidayah et al., 2011). Temporal rainfall disaggregation can be made by generating a stochastic model. Stochastic approach can be done in various ways, among others are: the approach to model time series (such as Periodic autoregressive (PAR), Gamma autoregressive (GAR)) and point processes (such as Battlet-Lewis or Neyman-Scott models). Hidayah et al. (2011) had developed model Rainfall Disaggregation Model Using Bayesian PAR1 model combined with adjusting and filtering procedure. Model structure is adopted from Koutsoyiannis and Manetas (1996), and Yeboah (2005).

The Koutsoyiannis and Manetas (1996) method used is based on the type of periodic autoregressive (PAR1) model with some modification (adjusting) of the modeling process. This approach is intended to maintain the structure modeling PAR1 and claimed that it has been able to produce a simpler arrangement of parameters (parsimony). This model is able to overcome various forms of data distribution from symmetry (bell) to non-symmetric (skewed), and the generation of this model is by adjusting low-level variable in accordance with a high level. However, the weakness of the PAR1 model is not able to cope with over-estimation of variance and extreme values are still under-estimate the depth of rainfall at the maximum. To handle over and under estimation of model parameter, Hidayah et al. (2011) estimated the model using Bayesian Markov Chain Monte Carlo (MCMC) using WinBUGs tools. Furthermore, Yeboah (2005) method, by filtering process, was used to handle the result of model when no rain expected the value was zero.

To assess the ability of the PAR1 model coupled with adjusting and filtering using Bayesian approach, this paper tries to explore it using data at different time.

Model Description:

The rainfall disaggregation model was formed from model time series PAR124 because the pattern of the occurrence of rainfall is periodical. This model assumes that the occurrence of rainfall is periodical, represented by $Y_{v,\tau}$, where $v$ defines year, and $\tau$ defines the season, for example $\tau = 1 \ldots, \omega$, and $\omega$ is the number of seasons during the year. General form of the Period Autoregression (PAR 1) model is (Maidment, 1992):

$$\begin{align*}
Y_{v,\tau} = \mu_\tau + \Phi_{1,\tau}(Y_{v,\tau-1} - \mu_{\tau-1}) + \epsilon_{v,\tau}
\end{align*}$$

(1)
Next Steps that must be done within the parameter estimation of PAR (1)\(^2\) model was to determine the probability density function and the likelihood function (Iriawan, Suhartono, Atok, 2008). Chance density function for hourly rainfall data \((y_t)\) is:

\[
f(Y_t, \theta_0, \theta_1, \theta_2, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(Y_t - 0\theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-24})^2}
\]

(2)

\(\theta_0\) and \(\theta_1\) as autoregressive coefficient
\(\Phi_2\) as seasonal autoregressive coefficient
\(\sigma\) as time series variation

In determining the likelihood function of influential factors are the sample data and probability density function of population. If the population is in form of \(Y_{1i}, Y_{2i}, …, Y_{Ti}\) and the probability density function \(f(Y_t, \theta_0, \theta_1, \theta_2, \sigma)\) then the likelihood function on the order of \(p = 1\) and \(P = 1\) (\(p =\) non-seasonal AR order, \(P =\) non-seasonal AR order) is:

\[
L(Y_1, Y_2, …, Y_T) = f(Y_t, \theta_0, \theta_1, \theta_2, \sigma), …, f(Y_T, \theta_0, \theta_1, \theta_2, \sigma)
\]

(3)

If \(e_t = Y_t - (\theta_0 + \theta_1 Y_{t-1} + \Phi_2 Y_{t-24})\), so the likelihood function can be written as follows:

\[
L(\theta \mid Y_1, Y_2, …, Y_T) = -\frac{1}{2} \log(2\pi \sigma^2) - \frac{1}{2} \sum_{t=1}^{T}(e_t^2)
\]

(4)

From (4) is derived partially on \(\theta_0\), \(\theta_1\), and \(\theta_2\), each is made the same as null and done simultaneously, the estimation \(\theta_0\), \(\theta_1\), and \(\theta_2\) will be obtained.

Bayesian approach was used to estimate parameter of the model. In Bayesian, the information known about the parameter \(\theta\) before the observation is made is called the prior \(p(\theta)\). Furthermore, to determine the posterior distribution \(\theta\), i.e. \(p(\theta \mid x)\), is based on the rules of probability in Bayes theorem as follows:

\[
p(\theta \mid x) = \frac{f(x \mid \theta)p(\theta)}{f(x)}
\]

(5)

with: \(f(x) = E(f(x \mid \theta))\)

\(f(x)\) will have a constant value called normalized constant (Carlin et. al., 2003). Then, the equation (6) can be written as:

\[
p(\theta \mid x) \propto f(x \mid \theta)p(\theta)
\]

(6)

Equation (6) shows that the posterior is proportional to the likelihood multiplied by prior of model parameters. One method used in determining the prior distribution in this study is Conjugate prior.

In Bayesian analysis, Method of Markov Chain Monte Carlo (MCMC) was used. According to Carlin and Chib, (1995) it facilitated fairly complex modeling. MCMC approach was taken by Gibbs sampling method. Gibbs Sampling is a technique for indirectly generating random variables from the marginal distribution without having to calculate its density. By using Gibbs sampling, difficult calculation can be avoided (Casella and George, 1992). Gibbs sampling does not calculate or estimate \(f(x)\) as a marginal density from a combined density of some of the parameters/variables directly, but it is done by generating sample \(x_1, x_2, …, x_m\) from a full conditional distribution (Iriawan, Suhartono, Atok, 2008). By conducting a simulation of a large number of samples \(m\), \(\infty\) with the theorem of strong law of lack numbers, then the average variance or characteristics of the other \(f(x)\) can be calculated from the desired level of accuracy.

The approach to adjusting procedure used to obtain consistency of data between low-level scale and the scale of the level above is by: (1) minimizing the error value of the sum variable low levels of high-level variables. (2) maintaining the statistical distribution of the resulting low level.
This procedure is used to modify the initial value generation \(Y_s\) to get the value of \(X_s\) are in adjusting to the equation:

\[
Y_s = \frac{\sum_{s=1}^{\infty} Y_j}{Z} \quad (s = 1, ..., k)
\]

(5)

with \(Z\) is a high-level variables and \(k\) is the number of low-level variables with one period at a higher level.

Filtering approach is used to divide the processes into two parts, namely when rainfall and when no rainfall. The concept used is assuming that the event is a binary. Simple binary model is used to estimate the unknown population proportion of data series \(y_1, ..., y_n\), where each data can be either 0 or 1.

**Model Structure:**

The structure model as shown in the Doodle by WinBUGs in Figure 1. Doodle illustrates the structure of the model in the nodes and is connected by lines which express the functional relationship between nodes. Variables in the program code can be explained that, the variable \(y[i]\) represents the hourly rainfall data in normal distribution with a reverse word dnorm \((\mu[i], \tau)\) where the logical node \(\mu[i]\) followed the next statement is translated in accordance with the logical doodle image in Figure 1. \(\mu[i]\) is the expectation of PAR1 model prepared by the equation:

\[
\mu[i] <- a + b * y1[i] + c * y24[i]
\]

(7)

with \(\mu\) associated with \(y[i]\) is a variant of PAR1. Parameter \(a\) (as a constant), \(b\) (representing the interval of hourly rainfall), and \(c\) (representing the interval of daily rainfall) was stochastically change. **Prior** for each parameter \(a, b,\) and \(c\) was used in the generation process of model with the same normal distribution.

MAE calculated automatically in the program is to facilitate the process of evaluating the accuracy of the model. MAE is the absolute logical function of the residual \(e[i]\), with \(e[i]\) represents the difference \(\mu[i]\) to \(y[i]\).

Adjusting procedures are marked with 2 boxes loping. The first box is a loping for the generation of hourly rainfall data, while the second box is the adjusting process of data from hourly to daily. \(\mu[i]\) and \(x[i]\) is a link between the two boxes. Where \(\mu[i]\) associated with logical functions of \(x[i]\) and \(zb[h]\) with the equation respectively:

\[
x[i] <- (\mu[i] / ZB[h]) * za[h]
\]

(8)

\[
ZB[h] <- \text{sum (mu [(h - 1) * 24 + 1: h * 24])}
\]

(9)

\(x[i]\) associated with \(za[h]\) is a constant variable for daily rainfall and \(ZB[h]\) is the summation of hourly rainfall in a single day of the simulation results. Between \(za[h]\) and \(zb[h]\) is controlled by the delta \([h]\).

Filtering approach used is to divide the process into two parts, namely rainfall occurs and no rain. This concept uses the assumption of binary events, where if the value that appears similar to the parameter "a" then multiplied by zero, and others are multiplied by one. Filtering components are treated in the \(\mu[i]\). Terms used to generate a binary indicator function can be written as follows.

\[
\text{mu} = \begin{cases} 
0 & \text{if } \mu = a \\
1 & \text{if } \mu \neq a 
\end{cases}
\]

(10)

This binary formulation in the list of programs will be written with the command:

\[
\text{filtering [i] <- step (mu [i] - a) - equals (mu [i], a)}
\]

(11)

Further results from the multiplication of filtering with \(\mu[i]\) produce all \([i]\). The next error will be calculated automatically in the form of MAE, by finding the difference between all \([i]\) and \(y[i]\).

Hierarchical modeling process starts from the generation data of all prior parameter independently. Each time the generation of a prior value of certain parameters always includes a number of \(N\) data residing in the structure of the box (loping) used to calculate the likelihood in the process of its full conditional. Then, the results of the value generation would be used to estimate the average value of each component of the model and finally used to estimate the average of the PAR \((1)_{24}\) model by adjusting and filtering procedures. Generation process model was undertaken by Gibbs sampler iteration. MAE value of model was 0.44.
Results and Discussion

Implementation of this model is performed for two conditions namely to predict the model with the input data in 2005-2008 and to apply the model to input data in 2009.

Model Prediction:

The input of daily rainfall used for predicting of disaggregation rainfall model was the data of December 2009. The parameters used in the model estimation were the same as those of used in the last running result of the modeling of PAR(1)24 by adjusting and filtering procedures. The result of the model validation obtained a good model performance, and the value of MAE error was 0.37. The MAE error value below the value of MAE error of the modeling was 0.44.

The plot of the model validation results as shown in Figure 2 was in form of plot of synthetic hourly rainfall intensity with red color, the intensity of observation rainfall in blue color and the intensity of average daily rainfall in green color. The histogram plot shows that when rain was high, it maximally at the hour 373 had the difference between the disaggregation result and the relatively small observation of 3.76 mm or 12%.

The performance of the model prediction can be shown by the plot quantile. This quantile plot describes the relationship between observation rainfall and the results of rainfall disaggregation. The results of quantile plot of hourly rainfall depth (Figure 3.a) indicated a significant trend approaching the linear position, but the position of rainfall depth value tended to be an under-estimation with the $R^2$ value of 0.921 (see Figure 3.a). This means that hourly rainfall depth between the disaggregation results and observation, there were several improper events.

The plot of daily rainfall depth shows the trend of perfect accuracy where daily rainfall total resulted from simulation was equal to the hourly rainfall total in a single day. It is seen in Figure 3.b the resulted linear
equation \( Y = X +0.0004 \) and the \( R^2 \) value was equal to one. This means that the total result of disaggregation of hourly rainfall in one day was consistent with the observed daily rainfall data.

Results of quantile plot for dry interval period in Figure 3.c shows a trend linearly under-estimation where the resulted \( R^2 \) value of 0.951 and the linear equation was \( Y = 0.986 x +0.105 \). This means that the position of the rainfall arrival between disaggregation and observation was a little different.

The plot of duration quantile in Figure 3.d indicates a trend non-linear with under-estimation. It can be concluded that the duration distribution of rain events in a day between disaggregation and observation were not the same.

![Fig. 2: Histogram plot of hourly rainfall depth between simulation and observation of average of daily rainfall depth.](image)

![Fig. 3: quantile plot compares between simulation and observation on (a) hourly rainfall depth, (b) daily rainfall depth (c) dry interval period and (d) duration distribution of model PAR (1) 24 by adjusting and filtering procedures.](image)

**Model Application:**

Application of model PAR (1) 24 with the adjusting and filtering used the out of sampling rainfall data from 2005 to 2008. Applications rainfall disaggregation model for the data out of the sampling needs to be done on the calibration parameters in advance to produce a model that fits with the data pattern.

**Fitting Model:**

Model calibration results show that the model has a significant parameter for each month in Table 2. Model parameter values in Table 2 have MAE values vary relatively large for January to March. While its value MAE for July have much lower error of 2.87E-15 if it is compared with the results of running Heytos, where the MAE value of running Heytos conducted by Hidayah et.al. (2010a) for 0,022, although the MCMC process on the estimation of several parameters occurs slowly mixing.

The process simulation model using MCMC has not been able to estimate the parameters of all months fastly mixing, but the simulation process is able to generate parameter values and initial values MAE. Fastly mixing condition is achieved when the wet seasons of December, January, February and March, whereas during the dry seasons of the iteration process in fastly mixing only in May and September. The second month has a Lag 1 autocovariance value close to that is equal to 0.0178 and 0.01624. Lag 1 autocovariance of the second month in a kind of smallest value compared to other months. This suggests that there is random work in dry time low temporal correlation structure.
Table 2: The results of the calibration parameters of the model disaggregation of rainfall for January up to December.

<table>
<thead>
<tr>
<th>Month</th>
<th>Parameter</th>
<th>MAE</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>0.2006</td>
<td>0.3862</td>
<td>0.4539</td>
</tr>
<tr>
<td>February</td>
<td>0.367</td>
<td>0.1283</td>
<td>0.08032</td>
</tr>
<tr>
<td>March</td>
<td>0.3948</td>
<td>0.2472</td>
<td>0.03201</td>
</tr>
<tr>
<td>April</td>
<td>0.01906</td>
<td>0.1636</td>
<td>0.1344</td>
</tr>
<tr>
<td>May</td>
<td>0.0196</td>
<td>0.1633</td>
<td>0.1342</td>
</tr>
<tr>
<td>June</td>
<td>0.3954</td>
<td>0.7205</td>
<td>0.3115</td>
</tr>
<tr>
<td>July</td>
<td>0.2006</td>
<td>1.642</td>
<td>0.4539</td>
</tr>
<tr>
<td>August</td>
<td>0.2212</td>
<td>0.1266</td>
<td>0.2327</td>
</tr>
<tr>
<td>September</td>
<td>0.007117</td>
<td>0.07366</td>
<td>0.3566</td>
</tr>
<tr>
<td>October</td>
<td>0.0813</td>
<td>0.03429</td>
<td>0.3152</td>
</tr>
<tr>
<td>November</td>
<td>0.059711</td>
<td>0.6236</td>
<td>0.1026</td>
</tr>
<tr>
<td>December</td>
<td>0.2549</td>
<td>0.19</td>
<td>0.14</td>
</tr>
</tbody>
</table>

The Accuracy of Model:

The evaluation of the accuracy of the model application for each month is reviewed based on root mean square error (RMSE) and mean square error (MSE) from the comparison between simulation results against observations to a depth of maximum hourly rainfall. A review of rainfall maximum depth is intended to demonstrate the ability of the model inputs in response to the flood model. Figure 4 shows that the depth of maximum hourly rainfall from observation is still above their simulation. In January, February, April and September have a relatively high gap that exceeds 10%, while for March and December; the difference in value is small enough that less than 10% (see Table 4.4). The plot results of the depth of maximum hourly rainfall between observation and simulation results as shown in Figure 4 the trend shows that the results of model simulations tend to under-estimate. Underestimation is occurred in January and February, so the model needs to be improved by incorporating the influence **Heteroschedasticity**.

Fig. 4: Maximum hourly rainfall depths between simulation and observation.

Error value from RMSE and MSE of every month was used to statistically compare the result of application. The results of RMSE and MSE values indicate that the best model with the smallest value, which is shown in April for the rainy season and in May for the dry season with the MSE value close to the RMSE.

The results of the implementation of PAR model (1) 24 with the adjusting and filtering for the month of January, February and March seem to have not been able to generate an error model is as good as in December. This is possible due to the high variance of the data as shown in Table 4.1. Subsequent applications for the month of December in addition to more studies need to be careful regarding the distribution of selected data. Further application of the model will be tested for the same month a different time.

Table 4.5: Maximum rainfall depth between observation and simulation.

<table>
<thead>
<tr>
<th>Month</th>
<th>Hourly maximum rainfall depth (mm)</th>
<th>Selisih (mm)</th>
<th>Person Selisih</th>
<th>MSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>30.2</td>
<td>46</td>
<td>15.8</td>
<td>12.868</td>
<td>165.595</td>
</tr>
<tr>
<td>Feb</td>
<td>27.6</td>
<td>49.1</td>
<td>21.5</td>
<td>6.103</td>
<td>6.103</td>
</tr>
<tr>
<td>Mar</td>
<td>47.7</td>
<td>53.3</td>
<td>5.6</td>
<td>11.460</td>
<td>11.460</td>
</tr>
<tr>
<td>Apr</td>
<td>8.07</td>
<td>15.6</td>
<td>6.93</td>
<td>0.903</td>
<td>0.903</td>
</tr>
<tr>
<td>Mei</td>
<td>8.67</td>
<td>15.64</td>
<td>6.93</td>
<td>0.903</td>
<td>0.903</td>
</tr>
<tr>
<td>Jun</td>
<td>-</td>
<td>22.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Jul</td>
<td>-</td>
<td>8.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Agt</td>
<td>-</td>
<td>8.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sep</td>
<td>7.77</td>
<td>9.5</td>
<td>1.73</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>Okt</td>
<td>-</td>
<td>34.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nop</td>
<td>-</td>
<td>26.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusion:

Based on the results of model applications can be summarized as follows.

1. The results of model predictions with hourly rainfall data in December of 2009 obtained a good model performance and the MAE value at 0.37. MAE value was obtained under the MAE values of the initial model that is equal to 0.44.

2. The results of calibration on the out sampling data showed that the model with MCMC simulation process can work for the majority of the wet rainy season i.e. January, February, March, April and December and a small part to the dry season i.e. May and September. The results of model simulations indicate that there is a difference in a relatively small depth of rainfall for the month of March, April, September and December, and conversely there are substantial differences relative to other months. In addition in December, a good modeling result is in March.

References


