ESTIMATION OF DAILY WEATHER DATA BY GENERATING MONTHLY DATA: NORTH SULAWESI CASE STUDY

Eko Supriyadi
Center for Marine Meteorology, BMKG, Jl. Angkasa I No. 2, Jakarta Pusat, 10610
E-mail: eko.supriyadi@bmkg.go.id

Article submitted: September 15, 2020 Article revised: January 12, 2024 Article accepted: January 24, 2024

ABSTRACT

The limited availability of daily data is a major issue when working with weather datasets. It can lead to discontinuities in the data history that impact the accuracy of the model output, which is a function of the observation input. This study presents a simple technique for generating daily weather data from monthly data using synoptic observations from the North Minahasa Climatology Station (1989-2014). Our approach involves the use of logit, Fourier, and gamma functions. The generated rainfall exhibits a similar pattern to the observed rainfall, and the statistical test indicates no significant difference ($p<0.000$). However, the resulting correlation is low (0.33-0.49). Based on the seasonal division, both the generated and observed rainfall values are high in the December-January-February (DJF) season, followed by a decrease in the March-April-May (MAM) and June-July-August (JJA), and an increase in the September-October-November (SON) season. Additionally, there are discrepancies in rainfall generation compared to observations due to the use of uniform distribution random numbers, which tend to overlook temporal specificity. Moreover, other meteorological factors, such as temperature and relative humidity, generate daily values from monthly data that closely resemble observations, with a correlation coefficient greater than 0.8. This is due to the Fourier function's lacks a field variability factor.

Keywords: Generate weather data, Logit, Fourier, Gamma function

1. Introduction

The Indonesia Meteorology Agency (BMKG) currently operates 27 Climatology Stations [1]. However, not all stations can conduct continuous weather observations due to limited equipment and personnel. This results in incomplete data or only partial reporting. Complete data is mandatory when needed and utilized by various sectors, such as early warning, agriculture, safety, and even insurance claims in accident cases. One of the parameters that is often limited in data availability is rainfall. Generally, weather stations are typically equipped with rainfall gauges. However, if these gauges are damaged, errors in data recording or observations may occur during specific periods, leading to a lack of rainfall data. Therefore, developing techniques for generating rain data from existing records is important. In this case, the technique involves generating daily rainfall data from monthly records. Srikanthan and McMahon's study [2] explains the differences in techniques for generating rain data in daily, monthly, and annual categories.

The technique for generating rain data is relatively new in atmospheric science. It was first introduced by Thomas and Fiering [3] and later widely applied by Thyer and Kuczera [4] and Srikathan and McMahon [2]. The technique involves calculating the probability of rainfall rather than directly determining the rainfall rate. This means determining the chance of rain occurring today compared to the conditions of the previous day. The terms p_{01} and p_{11} are used here to represent the probability of rain today given that it did not rain yesterday (p_{01}) and the probability of rain today given that it rained yesterday (p_{11}), which is also known as a first-order Markov chain [5]. It is important to note that this method rarely estimates the probability of non-rain events. Additionally, it should be noted that rain events have a high variability of values. Rain cannot occur continuously every day with the same intensity. Therefore, an equation is required to record the diversity of intensity. In this case, the gamma probability distribution or gamma density function is appropriate, as some studies have shown [6], [7]. The Weibull distribution technique [8] and moving average technique [9] can also be used as rain data generators. Zhu [10] provides detailed techniques for generating rainfall, which are divided into six probability distributions.

Besides rainfall, other weather elements like temperature, radiation, evaporation, and relative humidity (RH) can be generated from monthly to daily data. Calculating these four weather elements is easier than generating rain data, as they tend to have relatively low variability in value. This calculation is done using a Fourier function [11]. Further study of the Fourier function reveals that it can be used to generate weather data influenced by the sun. Both daily and seasonal values of temperature, radiation, evaporation, and RH are affected by the solar cycle. However, it is important to note that the coefficients attached to the Fourier function may differ from one station to another.

Additionally, Richardson [5] explains that modelling weather elements such as temperature and solar radiation is relatively straightforward due to their low proportion of zero-valued temperatures and smoother distribution compared to rain data. The generation of temperature data requires consideration of both maximum and minimum temperatures as part of a multi-variate stochastic process that takes into account daily rain conditions. Bristow and Campbell [12] demonstrated that the variation between maximum and minimum temperatures is correlated with the amount of solar radiation received by the surface. When radiation is low, the temperature difference is low, indicating cloudy weather; conversely, when radiation is high, the temperature difference is high, indicating sunny weather.

Weather generation techniques have been developed in various software tools [13]. These techniques can reach hourly scales with high spatial resolution [14]. Generally, weather generation techniques can be divided into numerical and stochastic models. Each model has complementary advantages and disadvantages [15]. Stochastic models are often preferred over numerical models due to their lower computational requirements. Numerical models, on the other hand, use a mathematical and physical description of the land-atmosphere system, which demands a significant computational load. Recent research has demonstrated that stochastic models can be employed as downscaling tools for impact studies [10].

Given the above explanation, this research has two main objectives. Firstly, to explain the technique for generating rainfall data in terms of rainfall occurrence and intensity. Secondly, to explain the technique for generating data on temperature, radiation, evaporation, and RH. The detailed equations and steps are presented in Chapter 2, while the results are presented in Chapter 3. These two objectives aim to provide insight into weather data generation techniques and their applications and references to the understanding eastern Indonesian weather, which is rarely explored.

2. Methodology

The data used in this study include observations of rainfall, relative humidity, and mean, maximum, and minimum temperature from the North Minahasa Climatology Station (NMCS) located at 124.92^0 E, 1.50^0 N in North Sulawesi, Indonesia, recorded between 1989 and 2014. Before further processing, both sets of data underwent validation to check for time sequence errors, comma-point errors, and extreme values. The value of 8888 in the rain data was changed to 0. Additionally, the tolerance limit of weather elements was checked using the average equation 4 ± st. deviation. If the value falls within this interval, the data is considered valid. The data processing comprises two major parts: modelling rainfall events and estimating them, as well as generating other weather data. The Matlab® student version has been used in this study.

Modelling and estimation rainfall event. The first step is to determine the chance value of rainfall events with the equation:

$$p_{01}(i) = \frac{n_{01}(i)}{n_{01}(i) + n_{00}(i)} \tag{1}$$

$$p_{11}(i) = \frac{n_{11}(i)}{n_{11}(i) + n_{10}(i)} \tag{2}$$

where indices 0 and 1 indicate the occurrence of no-rain and rain on day-i, respectively. Thus n_{01} means the occurrence of today it is raining, but yesterday it did not rain, and p_{01} means the probability of today it is raining, but yesterday it did not rain. Please note that this research only focuses on the "current event is rain".

At this point, the next step is to construct the chance equation for rainfall events using the Fourier regression equation (Eq. 5). However, the problem is that when entering Eq. (1) and (2) directly, it is possible that the odds value resulting from the Fourier regression equation can exceed one and be less than zero. Therefore, in this second step, we will change Eq. (1) and (2) through the logit function:

$$g_{01}(i) = \ln\left(\frac{p_{01}(i)}{1 - p_{01}(i)}\right) \tag{3}$$

$$g_{11}(i) = \ln\left(\frac{p_{11}(i)}{1 - p_{11}(i)}\right) \tag{4}$$

The third step is to construct the Fourier function equation through the form:

$$g_{jk}(i) = a_0 + \sum_{l=1}^{m}\left(a_l \sin(li') + a_{l+1}\cos(li')\right) \tag{5}$$

where $i' = 2\pi i/365$ and i is the number of days in a year (i = 1, 2, 3, ..., 365). To simplify the description, this study uses the form of the Fourier function with harmonic length (l = 2) written as:

$$g_{01}(i) = a_0 + a_1 \sin(i') + a_2 \cos(i') + a_3 \sin(2i') + a_4 \cos(2i') \tag{6}$$

$$g_{11}(i) = a_0 + a_1\sin(i') + a_2\cos(i') + a_3\sin(2i') + a_4\cos(2i') \quad (7)$$

Where a_0, a_1, a_2, a_3, and a_4 are the Fourier coefficients obtained from the regression equation. The results of Eq. (6) and (7) are then plotted together with Eq. (3) and (4) to obtain the Fourier fitting curve for one year simulations.

After obtaining the logit function in Eq. (3) and (4), then in the fourth step, it is converted back into the probability value following the form:

$$p_{01}(i) = \frac{1}{1 + \exp\left(-g_{01}(i)\right)} \quad (8)$$

$$p_{11}(i) = \frac{1}{1 + \exp\left(-g_{11}(i)\right)} \quad (9)$$

The purpose of converting back into chance values is to compare with uniform random numbers in order to obtain rainfall event values. In this study, the unifrnd function is used [16] in Matlab. If the value of the random number is smaller than the value of p_{01} and p_{11} then there is rain, and vice versa.

The fifth step goes into how to determine the alpha (α) and beta (β) parameters of the gamma density function. Please note that, before processing further, the number of n_{01} and n_{11} each year is first sought. The alpha and beta parameters can be determined following the form:

$$y = \ln(\bar{x}) - \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \quad (10)$$

$$\alpha = \begin{cases} \dfrac{0.5000876 + 0.1648852y - 0.0544276y^2}{y} \\[2ex] \dfrac{8.898919 + 9.05995y + 0.9775373y^2}{y(17.79728 + 11.968477y + y^2)} \\[2ex] \dfrac{1}{y} \end{cases} \quad (11)$$

$$\beta = \frac{\bar{x}}{\alpha} \quad (12)$$

Where the value of $\bar{x}$ is the value of the number of each n_{01} and n_{11} each year. So that later two types of gamma parameters will be obtained for each n_{01} and n_{11} . The value of α in Eq. (11) has the conditions $0 \le y \le 0.5772$, $0.5772 \le y \le 17$, and $y > 17$ for the top, middle, and bottom panels, respectively. Please also note that the gamma density function $f(x; \alpha, \beta)$ is another form of probability density function (pdf) formulated as:

$$f(x;\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \quad (13)$$

with $\Gamma(\alpha)$ is the gamma function, as:

$$\Gamma(\alpha) = \int e^{-x}x^{\alpha-1}dx \quad (14)$$

So the value to be sought is actually the x-form, with the value of $f(x; \alpha, \beta)$ used coming from Eq. (8) and (9). In other words, the final result is the daily estimated rainfall value for one year, not the entire observation period (1989-2014). Because it requires complicated calculations, the value of x in Eq. (13) is solved using the gaminv function [17].

The sixth step is compare the observed and estimated rainfall values to obtain the regression coefficient. The regression coefficient value obtained is then used as the estimated rain value to compare with the observed rain. This is done because the estimated rain value generated from Eq. (13) is smaller than the observed rain.

The seventh step, as the final step, is to conduct a statistical analysis of the sixth step and differentiate the rainfall patterns that occur based on the four seasons, namely DJF, MAM, JJA, and SON. The results of the above methodology are presented in the first Chapter 3, results and discussion.

Generating other daily weather data from monthly dataset. This section presents the methodology for generating daily weather data for mean temperature, maximum temperature, minimum temperature, and mean relative humidity from monthly data at NMCS. The same methodology applies to all weather elements except rainfall.

The initial step involves determining the average monthly weather elements for 26 years at NMCS. Subsequently, a single year within the observation period at NMCS is selected to evaluate the conversion of monthly data into daily data. In this case, the year 2014 is used. The objective is to compare the observed weather elements with the generated data to determine if they exhibit similar patterns.

The second step is to apply, Fourier function, which is helpful as a regression function to generate daily data:

$$P(t) = a_0 + \sum_{l=1}^{m}\left(a_l\sin(lt') + a_{l+1}\cos(lt')\right) \quad (15)$$

where $t' = 2\pi t/365$ with t obtained through the relationship:

$$t = T - 0.5 + \frac{m - 0.5}{D} \quad (16)$$

where T and m are the month and day of observation, and D is the number of days in the month of observation. This results in a row equal to the number of days in a year, i.e., 365, and three columns containing the value of the month that repeats from 1 to 12, the value of the day that repeats from 1 to 28, 30, or 31 (depending on the type of month), and the number of days that follow the month of observation. The values of the three columns are entered into Eq. (16) to get t'.

To simplify the understanding of the concept, Eq. (15) used in this study uses a harmonic length of 2 (l = 2), which can be explained as:

$$P(t) = a_0 + a_1 \sin(t') + a_2 \cos(t') + a_3 \sin(2t') + a_4 \cos(2t') \quad (17)$$

Thus, the third step is to find the linear regression coefficients for the first through fifth forms of the right-hand segment of Eq. (17), with P(t) values as monthly data and daily data for each of the four weather elements to be generated. It means that the regression coefficients generated are of two types: the regression coefficients for generating monthly and daily data. Please note that the monthly data here is the weather element value data that has been averaged for 26 years to generate daily data during 2014 and the daily data is the original daily data during 2014. To simplify the calculation, the regress function is used [18].

The fourth step is to multiply the sin and cos forms in Eq. (16) with the regression coefficients (a_0, a_1, a_2, a_3, and a_4) to obtain monthly and daily weather generation data. Furthermore, the two generation data were plotted together with the observed weather data to determine the similarity of the pattern and statistical analysis. All stages of the method in this study refer to Boer [19]. Figure 1 presents the research flow chart for rainfall and non-rainfall datasets.

3. Result and Discussion

Rainfall occurrence and intensity model. Rainfall occurrence refers to the probability of rainfall happening rather than its intensity. The analysis utilizes the logit functions of Eq. (3) and (4) plotted alongside the Fourier functions of Eq. (5). The results are presented in Figures 2 and 3, which show a similar pattern with the minimum value reached in September and the maximum around the end of December. The graph shows a seasonal pattern of rainfall at NMCS, and the following discussion presents the difference between estimated and observed rainfall for each of the four seasons. Two types of logarithmic functions, g_{01} and g_{11}, are used. The function g_{01} represents the influence of a non-rain event the day before on the current rain event, while g_{11} represents the influence of a rain event the day before on the current rain event. Rain events show significant time variability, in contrast to other weather elements which tend to be more consistent.

Note that the logit function here is not the chance value of rainfall events, but the conversion of the chance values from Eq. (3) and (4) in the form of natural logarithms, the aim is to avoid the probability values >1 and <0. However, after plotting Eq. (3) and (4) (not shown here) shows that the probability values of both equations are still within 0-1.

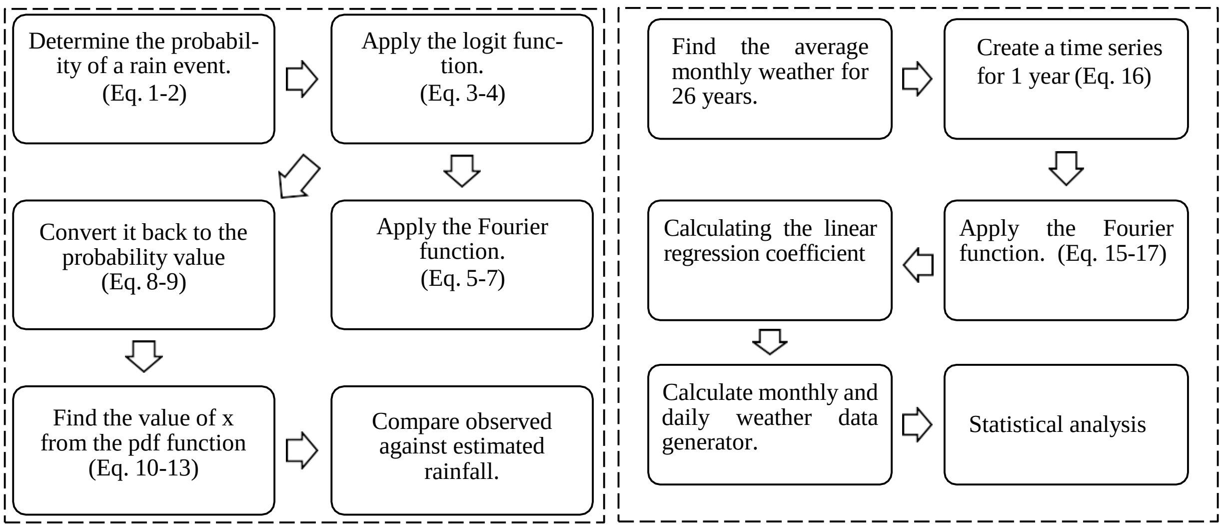

Figure 1. **Flowchart of the study. (a) the process of generating the rainfall event model and its estimation. (b) the process of generating other daily weather data from monthly dataset.**

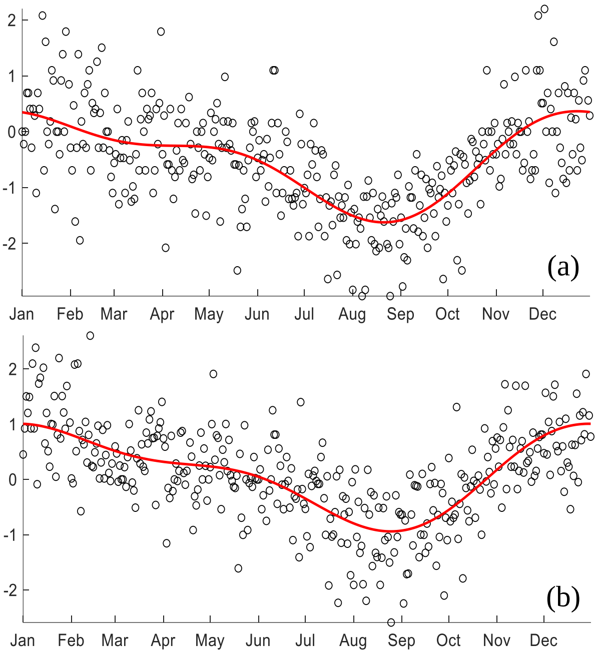

Figure 2. Scatterplot of logit function values of g_{01} (a) and g_{11} (b) (black circle) overlaid on the Fourier regression equation (red line) at NMCS over a 26-year span. Y-axis values are unitless.

This shows that the logit function serves as an anticipation if there is an error in forming the Fourier equation. The study conducted by Boer [20] found that the relationship between g_{01} and g_{11} is exponential to monthly rainfall. This means that we are able to generate daily rain data capitalizing on the available monthly rain data, and we also obtained a correlation of > 75%.

To generate the rain intensity value, the gamma function Eq. (13). Here, the role of the generated rainfall value is the variable x by first finding the value of the gamma parameters (alpha and beta). Both parameters come from the input of events n_{01} and n_{11}, which are summed up every year 26 times. The resulting gamma parameter values (alpha and beta) are listed in Table 1. It can be seen that the alpha is higher than the beta value. Furthermore, Wilks [21] shows that by assuming the alpha value is constant, the gamma function is more variable to changes in beta value. This is due to the beta value will effectively stretch the gamma density function to the right or left, depending on the overall size of the data. Whereas in normal distribution, the larger the data produced, the more independent the data will be and follow the central limit theorem.

Table 1. Gamma parameters of two rain events at NMCS based on Eq. (13).

Gamma parameter	n_{01}	n_{11}
α	73.6803	26.2587
β	0.8415	4.8936

Description: n_{01} : it's raining today, but not yesterday
n_{11} : it rained today and yesterday

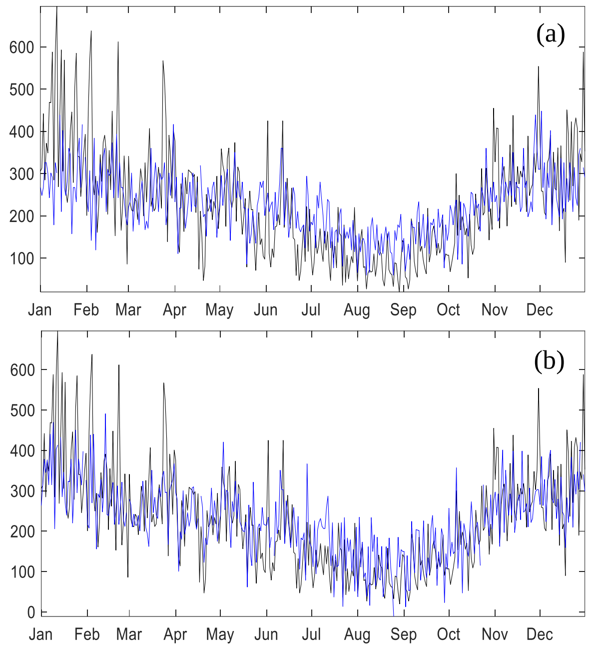

Figure 3. Comparison plots of observed (black line) and estimated rainfall (blue line) generated by logit and gamma functions over a 26-year span for events n_{01} (a) and n_{11} (b). The y-axis values are in mm/day.

To compare the generated rainfall data with the observed rainfall, the observed rainfall data is summed first for 26 years on the same day. This is done because the rainfall data generated from the gamma function is an estimated rainfall that involves the value of rain events (n_{01} and n_{11}) for 26 years. From the rainfall data, the value may be much different from the observed rainfall, so tuning is carried out in order to obtain an estimated rainfall value that is not too far away. The tuning treatment will not change the statistical value of the initial data. The results are presented in Figure 3.

It can be seen in Figure 3 that the observed rainfall has a similar pattern to the estimated rainfall generated for the case of n_{01} and n_{11}. It shows that the rainfall generated using logit and gamma functions works well in the case of n_{01} and n_{11}. In determining the estimated rainfall from Figure 3, the coefficients derived from Eq. (6) and (7) are presented in Table 2. It can be seen that only the coefficients a_1 and a_2 have almost similar values. This shows that in generating rainfall data, only the first harmonic length in the Fourier function plays a role, followed by the second harmonic length and so on. Furthermore, the gamma function that has been fitted together with the Fourier function has an equation:

$$g_{01} = -0.5064 + 0.4327 \cdot \sin(i') + 0.6818 \cdot \cos(i') - 0.322 \cdot \sin(2i') + 0.1752 \cdot \cos(2i')$$

$$g_{11} = 0.1041 + 0.4366 \cdot \sin(i') + 0.6807 \cdot \cos(i') - 0.234 \cdot \sin(2\,i') + 0.2226 \cdot \cos(2i')$$

Table 2. Fourier function coefficients for generating rainfall dataset at NMCS.

Logit function	a_0	a_1	a_2	a_3	a_4
g_{01}	-0.506	0.433	0.682	-0.322	0.175
g_{11}	0.104	0.437	0.681	-0.234	0.223

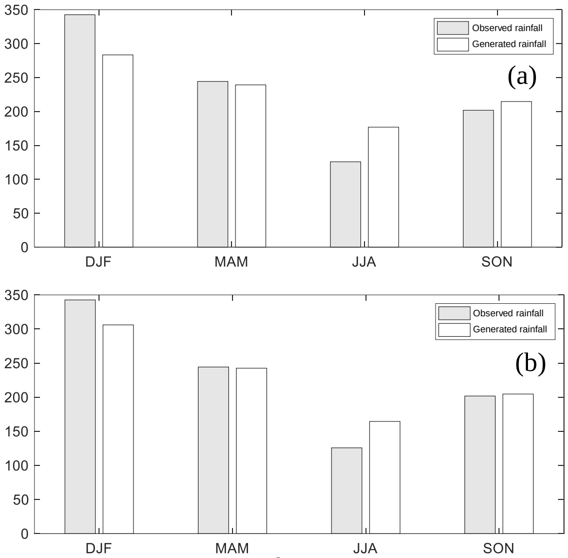

Figure 4. Comparison of observed and generated rainfall for p_{01} (a) and p_{11} (b) events by season. The y-axis value of rainfall is in mm/day.

From the results of statistical tests, both graphs in Figure 3 and Table 3 show that the estimated rainfall and observations are not significantly different. Note that the third column of Table 3 shows the best estimated intercept and variable values to obtain daily generation data from monthly data. For example, to generate daily average rainfall, the equation is $Rainfall_{daily} = Rainfall_{monthly}*18.66 - 884.77$. Note that the equation is not broken down by season. Then the st. error column in Table 3 refers to the notion of standard deviation of intercepts and coefficients from their mean values. It can be seen that a large estimated value results in a large st. error as well. Then, the t-test column shows how far the significant difference in the average intercept and coefficient values of the observed and estimated rainfall relationship. In statistics, the t-test value is associated with the t-table, but in this study, it is enough to see from the p-value column. Then, the p-value column shows the wrong chance of rejecting the statement that there is no good relationship between observed and estimated rainfall. As can be seen, Table 3 also shows that the rainfall generation data and observations are not significantly different (p-value<0.000) from all data (n days) used. It means that the Fourier function (Eq. 5) is suitable to be used as a daily data generator from monthly data at NMCS, but with a correlation (r^2) between the two classifieds as low, which is below 0.50. The low correlation in rainfall generation indicates that the gamma function and Fourier equation applied to this region have a stronger influence from environmental factors that cannot be explained by these two functions. Corroborating evidence of this can be found in the work of Pan et al. [22], who generated monthly rainfall data from annual, which obtained correlations in the interval 0.15-0.50 [22].

After comparing the rainfall for one year, the next step is to divide the rainfall values to obtain the seasonal pattern (Figure 4). It can be seen that both observed and generated rainfall all have the same pattern, which is high in the DJF season, then decreases until JJA and rises again in the SON season. In addition, in the DJF season, the observed rainfall is higher than the generated rainfall, and the difference is small in the MAM season. Different results are shown in the JJA and SON seasons where the generated rainfall is greater than the observations. These results are caused by calculating rainfall generation using random numbers uniform distribution whose values are random and tend to ignore the time specificity factor. Meanwhile, as is well known, the JJA season is a dry season that tends to have minimal rain compared to other seasons. So it is necessary to choose a replacement technique for uniform distribution. In addition, a comparison with other regions is needed to determine whether or not it produces a similar analysis. Furthermore, Sagita [23] examined the periodicity of rain in North Sulawesi obtaining the highest periodicity occurs with a period of 36 and 18 dasarians, meaning that rain events in North Sulawesi are influenced by monsoonal phenomena and the Intertropical Convergence Zone (ITCZ).

Other weather data generation models. In addition to rainfall, other weather elements can be generated similarly, such as temperature, RH, radiation, and evaporation. In this discussion, due to dataset limitations, only the results of generating mean temperature, maximum temperature, minimum temperature, and RH elements from monthly data into daily data from NMCS are presented. The results are presented in Figure 5. It can be seen that the daily data generated from monthly data for all weather elements follow each of the observed weather elements well. It shows that Eq. (15) and (17) can be applied as a daily weather data generator.

The average and maximum air temperatures have similar sinusoidal patterns with maximum values occurring in September-October. However, the minimum temperature is relatively uniform throughout the year. The observed and generated average air temperatures are between 24-29 ^{0}C and 25-27 ^{0}C. The observed and generated maximum air temperatures range between 28-35^0 C and 29-32 ^{0}C. The observed and generated minimum air temperatures are in the range of 18-26 ^{0}C and 22-24 ^{0}C, respectively, relatively more stable than the other two temperature types. The observed and generated average RH values are 50-98 % and 70-90 % respectively. It can be seen that all generated weather data have a tighter range of values (daily to monthly) than the observed weather elements (Figure 5). This is because the equation used (Eq. 17) does not include error factors from the external environment that reflect the seasonal variability of the data during the year. Knowledge of the error factor is influenced by geographical conditions, the seasonal effects that are not properly taken into account, the influence of extreme weather conditions, the instrument used, the data collection method, and the theory applied [24].

A comparison of the statistical analysis of the observed weather elements (mean air temperature, maximum, minimum, and RH) with the results of the generated data is presented in Table 4. The second column in Table 4 states the coefficients of the linear regression equation, which consists of the intercept coefficient and other variables of the relationship of the generated weather to the observations, such as maximum, minimum, mean temperature, and RH. The intercept coefficient is the value that forms the slope, while the variables act as the coefficients of the linear regression equation. It means that the relationship of weather generation to observations can be written as a linear regression equation. Furthermore, the third column of Table 4 shows the best coefficient estimates to derive daily generation data from monthly data. For example, to generate the daily average temperature, the equation is $T_{daily} = T_{monthly}*1.2317 - 5.7198$.

Table 3. Statistical analysis of comparison of rainfall generation with observations at NMCS

	Estimation		st. error		t-test		p-value		n	r^2
	p_{01}	p_{11}	p_{01}	p_{11}	p_{01}	p_{11}	p_{01}	p_{11}		
Intercept	-884.77	-646.59	83.811	47.56	-10.56	-13.60	0.000	0.000	361	0.33
variable	18.66	6.79	1.40	0.37	13.30	18.50	0.000	0.000	357	0.49

Table 4. Statistical analysis of comparison of other generated weather elements with observations at NMCS

Parameter	Coefficient	Estimation	st. error	t-test	p-value	RMSE	r^2
T average	Intercept	-5.7198	0.147	-38.96	0.000	0.051	0.993
	variable	1.2317	0.006	220.40	0.000		
T max	Intercept	-7.4303	0.371	-20.05	0.000	0.177	0.968
	variable	1.2462	0.012	105.27	0.000		
T min	Intercept	-19.924	0.971	-20.52	0.000	0.189	0.844
	variable	1.9779	0.045	44.29	0.000		
RH	Intercept	-25.597	2.476	-10.34	0.000	2.850	0.840
	variable	1.2791	0.029	43.60	0.000		

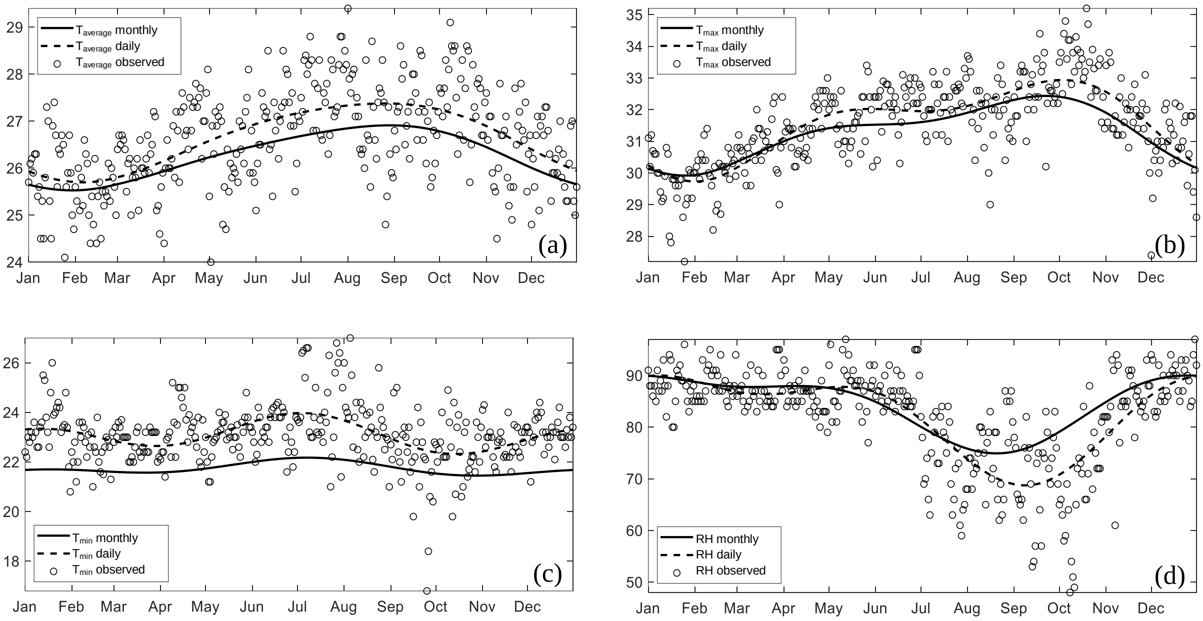

Figure 5. Comparison of observed weather elements (black circle) against monthly to daily weather generation (solid lines) and daily weather dataset (dashed lines) for average temperature (^{0}C) (a), maximum air temperature (^{0}C) (b), minimum air temperature (^{0}C) (c), and average RH (%) (d) in 2014 at NMCS.

Table 4 also shows that all weather elements generated with observations are not significantly different (p-value <0.000). It shows that the Fourier function (Eq. 15) is suitable for generating daily data from monthly data at NMCS. Furthermore, the smallest RMSE is generated in the generation of RH data, which is 2.850, while the correlation is high for all generated data, which is > 0.80. The smaller (larger) of the RMSE (r^2) obtained, the better the model is for generating weather generation data. This kind of weather generation utilization relationship is important in exploring and reconstructing the history of climate change. Recent research conducted by Pan et al. [22],utilizing average temperature and annual rainfall, was able to produce monthly rainfall data from 1930-2016 with a correlation of 0.15-0.50.

4. Conclusion

In the case of generating rainfall data at NMCS, a pattern similar to the observations was obtained. The statistical analysis results also show that the comparison of the two shows results that are not significantly different.

By dividing the observed and generated rainfall into four types of seasons, it is obtained that the observed and generated rainfall are high in the DJF and then decrease until JJA and increase again in the SON season. In addition, in the DJF, the observed rainfall is higher than the generated rainfall with a small difference in the MAM season. However, in the JJA and SON seasons, the generation rainfall is greater than the observations.

Other weather elements, such as temperature and RH, can be generated from monthly data into daily data using Fourier functions. The application to weather elements at NMCS obtained daily data generated from monthly data for the weather elements of average air temperature, maximum, minimum, and RH which follows each of the observation weather elements well. In addition, daily to monthly generated weather data has a tighter range of values than observation data. This is because the generated weather data cannot explain the high variability of field observations due to the influence of error factors. The resulting correlation between the two da-taset is > 0.80.

Future research can include error factors in the Fourier equations used in this study.

References

[1] BMKG, Peraturan Badan Meteorologi, Klimatologi, dan Geofisika Nomor 1 Tahun 2022 Tentang Perubahan Atas Peraturan BMKG Nomor 6 Tahun 2020 Tentang Organisasi dan Tata Kerja Balai Besar Meteorologi, Klimatologi, dan Geofisika, Stasiun Meteorologi, Stasiun Klimatologi, dan Stasiun Geofisika, 2022. https://jdih.bmkg.go.id/dokumen/view?id=4200 diakses tanggal 13 Maret 2023.

[2] R. Srikanthan and T. A. McMahon, “Stochastic generation of annual, monthly and daily climate data: A review,” Hydrol. Earth Syst. Sci., vol. 5, no. 4, pp. 653–670, 2001, doi: 10.5194/hess-5-653-2001.

[3] J. A. Thomas and M. B Fiering, “12. Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation,” Harvard University Press, 2013, pp. 459–493.

[4] M. Thyer and G. Kuczera, “Modeling long-term persistence in hydroclimatic time series using a hidden state Markov Model,” Water Resour. Res. - WATER RESOUR RES, vol. 36, pp. 3301–3310, Nov. 2000, doi: 10.1029/2000WR900157.

[5] C. W. Richardson, “Stochastic simulation of daily precipitation, temperature, and solar radiation,” Water Resour. Res., vol. 17, no. 1, pp. 182–190, Feb. 1981, doi: https://doi.org/10.1029/WR017i001p00182.

[6] N. T. Ison, A. M. Feyerherm, and L. D. Bark, “Wet Period Precipitation and the Gamma Distribution,” J. Appl. Meteorol., vol. 10, no. 4, pp. 658–665, Jun. 1971, [Online]. Available: http://www.jstor.org/stable/26175673.

[7] C. Martinez-Villalobos and J. D. Neelin, “Why Do Precipitation Intensities Tend to Follow Gamma Distributions?,” J. Atmos. Sci., vol. 76, no. 11, pp. 3611–3631, 2019, doi: 10.1175/JAS-D-18-0343.1.

[8] J. S. Selker and D. A. Haith, “Development and Testing of Single-Parameter Precipitation Distributions,” Water Resour. Res., vol. 26, no. 11, pp. 2733–2740, Nov. 1990, doi: https://doi.org/10.1029/WR026i011p02733.

[9] M. Yadnya, K. Astuti, and G. Hendrantoro, "Pembangkitan Curah Hujan dengan model MA (Moving Average) dari Hasil Pengukuran di Surabaya," Conference: Seminar Nasional Aplikasi Sains dan Teknologi IST AKPRIND Yogyakarta, 2008.

[10] B. Zhu, J. Chen, and H. Chen, “Performance of multiple probability distributions in generating daily precipitation for the simulation of hydrological extremes,” Stoch. Environ. Res. Risk Assess., vol. 33, no. 8, pp. 1581–1592, 2019, doi: 10.1007/s00477-019-01720-z.

[11] E. S. Epstein, “On Obtaining Daily Climatological Values from Monthly Means,” J. Clim., vol. 4, no. 3, pp. 365–368, 1991, doi: 10.1175/1520-0442(1991)004<0365:OODCVF>2.0.CO;2.

[12] K. L. Bristow and G. S. Campbell, “On the relationship between incoming solar radiation and daily maximum and minimum temperature,” Agric. For. Meteorol., vol. 31, no. 2, pp. 159–166, 1984, doi: https://doi.org/10.1016/0168-1923(84)90017-0.
[13] S. Mehan, T. Guo, M. W. Gitau, and D. C. Flanagan, “Comparative Study of Different Stochastic Weather Generators for Long-Term Climate Data Simulation,” Climate , vol. 5, no. 2. 2017, doi: 10.3390/cli5020026.
[14] F.-Y. Lin, K.-T. Huang, T.-P. Lin, and R.-L. Hwang, “Generating hourly local weather data with high spatially resolution and the applications in bioclimatic performance,” Sci. Total Environ., vol. 653, pp. 1262–1271, 2019, doi: https://doi.org/10.1016/j.scitotenv.2018.10.433.
[15] K. Breinl, G. Di Baldassarre, M. Girons Lopez, M. Hagenlocher, G. Vico, and A. Rutgersson, “Can weather generation capture precipitation patterns across different climates, spatial scales and under data scarcity?,” Sci. Rep., vol. 7, no. 1, p. 5449, 2017, doi: 10.1038/s41598-017-05822-y.
[16] J. Gentle, Random Number Generation and MonteCarlo Methods. Springer International Publishing, 2003.
[17] D. J. Posselt and C. H. Bishop, “Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter,” Q. J. R. Meteorol. Soc., vol. 144, no. 716, pp. 2331–2349, Oct. 2018, doi: https://doi.org/10.1002/qj.3374.
[18] S. Chatterjee and A. Hadi, “Influential Observations, High Leverage Points, and Outliers in Linear Regression,” Stat. Sci, vol. 1, Aug. 1986, doi: 10.1214/ss/1177013622.
[19] R. Boer, “Model Pembangkit Data Iklim (Climatic Data Generator),” Kapita Selekta Agroklimatologi, hal. 269-290, (tanpa tahun)
[20] R. Boer, I. Las, and K. A. Notodiputro, “Analisis risiko kekeringan untuk pengembangan produksi kedelai di Flores, Nusa Tenggara Timur,” Laporan Riset Unggulan Terpadu IV, Dewan Riset Nasional, 1998.
[21] D. Wilks, Statistical Methods in the Atmospheric Sciences (2nd ed.). Elsevier, 2006.
[22] F. Pan, L. Nagaoka, S. Wolverton, S. F. Atkinson, T. A. Kohler, and M. O’Neill, “A Constrained Stochastic Weather Generator for Daily Mean Air Temperature and Precipitation,” Atmosphere , vol. 12, no. 2. 2021, doi: 10.3390/atmos12020135.
[23] N. Sagita, “Spectral Analysis of Rainfall Data in North Sulawesi,” J. MIPA UNSRAT ONLINE, vol. 2, pp. 133–139, Nov. 2013.
[24] H. Mohajan, “Two Criteria for Good Measurements in Research: Validity and Reliability,” Ann. Spiru Harat Univ., vol. 17, pp. 59–82, Dec. 2017, doi: 10.26458/1746.