Determination of General Circulation Model Grid Resolution to Improve Accuracy of Rainfall Prediction in West Java

Nanda Fadhli*, Aji H. Wigena*, Anik Djuraidah*

* Department of Statistics, IPB University

DOI: 10.29322/IJSRP.9.07.2019.p91106
http://dx.doi.org/10.29322/IJSRP.9.07.2019.p91106

Abstract- The statistical downscaling technique is used to predict rainfall using general circulation model (GCM) data. One important factor to improve prediction accuracy is selection of GCM grid resolution. The purpose of this study is to determine the best GCM grid resolution to predict rainfall. The data used was GCM data from CFSRv2 with grid resolutions of 0.312°×0.312°; 0.5°×0.5°; 2.5°×2.5° and local rainfall data in West Java. The determination of best grid resolution began with determining the GCM domain in each rainfall observation station based on minimum correlation value of 0.3 between GCM data and local rainfall data. The domains were evaluated by LASSO regression based on the smallest RMSEP and the largest correlation. The results showed that the best grid resolution is 2.5°×2.5°.

Index Terms- GCM grid resolution, Statistical downscaling, LASSO regression.

I. Introduction

Rainfall as one of the climate elements has a serious impact especially on the agricultural sector [1]. Increasing rainfall above the normal average potentially cause flood [2], whereas decreasing rainfall below-normal can cause drought [3]. Floods and droughts also harm the area of harvest and production [4].

The negative impact of high rainfall and low rainfall is felt in West Java. Indramayu Regency as the biggest rice producer area in West Java was hit by floods that soak ten thousand hectares of rice fields [12]. Meanwhile, Dewi (2019) revealed that West Java had no rain more than 61 days which had the potential to cause drought. This will cause significant impacts on food production. Based on this description, rainfall information is needed to minimize these negative impacts and maximize production results.

General Circulation Model (GCM) data were often used in climate studies, such as precipitation [5]. GCM is a numerical model that produces a number of data on various climate parameters such as precipitation, temperature, and humidity [6]. The data generated from the model are the climate parameter values on the GCM grids. GCM data are available at various grid resolutions, including 0.312°×0.312°; 0.5°×0.5°; and 2.5°×2.5° in the form of the Climate Forecast System Reanalysis (CFSR) model. The grid resolution includes a rough resolution due to $1°\times1° \approx 100\text{x}100\ \text{km}^2$. The precipitation value in the grid cannot be used directly to predict rainfall in local scale areas that have a smaller grid resolution [7]. Therefore, rainfall estimating requires statistical downscaling (SDS) techniques to bridge global scale data with local scale data.

Rainfall prediction using the SDS technique requires a GCM area called the GCM domain. The selection of the GCM domain is an important factor in the SDS technique and will determine the predicting results [8]. The grid resolution is considered to improve the rainfall prediction accuracy.

This study aims to determine the best GCM grid resolution and to improve the rainfall prediction accuracy. The grid resolution determination began with determination of GCM domain using concept of a minimum correlation of 0.3 [14] between the GCM grid data (predictors) and the local rainfall data (response) for the grid in the four compass directions. This is supported by Busuioc *et al.* (2001) which states that one of the conditions in the SDS technique is the close relationship between response and predictors.

II. METHOD

A. Data

In this study, two types of data were used. The first data, as predictor variables, were monthly precipitation data of the GCM output in 1981 to 2009 (348 months). The GCM data was issued by the National Centers for Environmental Prediction (NCEP) in the form of CFSR model and could be downloaded on the website https://rda.ucar.edu/. CFSR is a model that describes a global interaction between land, sea, and earth air that is measured every 6 hours or 4 times a day. The GCM data were in grids above the territory of Indonesia, located at 12°N to -15°S latitude and 90°E to 150°E longitude. The grid resolutions in this study were 0.312° × 0.312°; 0.5° × 0.5°; and 2.5° × 2.5°. The second data, as a response variable, were monthly rainfall data (mm/month) of 27 local rainfall stations located at -7.78°S to -6.28°S latitude and 108.40°E to 107.87°E longitude in West Java Province. The rainfall data obtained from the Meteorological, Climatological, and Geophysical Agency (BMKG).

B. Procedure of Analysis

The procedure used R 3.5.2 software with the following steps:

1. Preparing and exploring data
   a. Extracting CFSR data from **netCDF** format into R dataframe format using packages **ncdf4**.

   b. Combining predictor variables (CFSR data) with response variables (local rainfall data).
   c. Cleaning the missing data so that the data had a different length of data at each station, but in general, this data had length of 348 months.
   d. The data was divided into two parts, named modeling data and validation data. Modeling data was the entire monthly data except validation data. While the validation data was monthly rainfall data that had been cleaned in item (1.c) in the last 12 months period.
   e. Exploring data using patterns of local rainfall distribution.
2. Determining the best grid resolution.
   These processes were carried out on three local rainfall stations, namely Leles station (lowland 0-200 metres above sea level (masl)), Lengkong station (medium land 200-500 masl), and Sangiang station (plateau> 500 masl) [10].
   The processes were as follows:
   a. Determining the GCM grid that would be used as a reference. The grid was the closest coordinates to the local rainfall station.
   b. Calculating correlation between the local rainfall station and GCM grid precipitation on four compass directions with the reference grid as the center [7].
   c. Selecting grids from step (2b) with correlation value at least 0.3.
   d. Forming a square / rectangular domain that includes the grids in point (2.c). The process illustration of determining the GCM domain is shown in Figure 1.

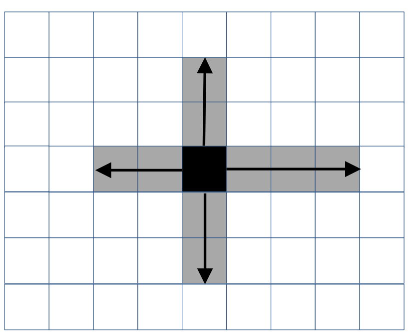


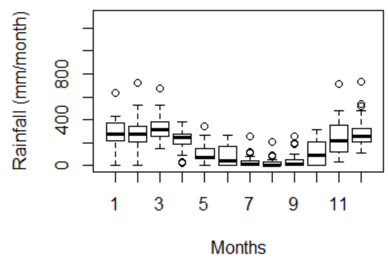
Figure 1: Illustration of determining domain GCM

Information:
■ : The Reference grid
■ : The grid with correlation of at least 0.3

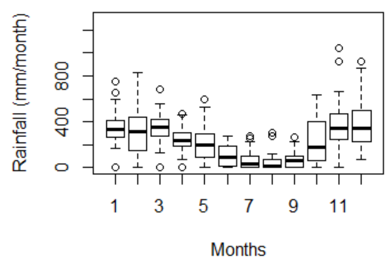
   e. Performing LASSO regression modeling with the domain obtained in (2.d) using the **glmnet** function in the R program, and finding the best lambda (λ) using the **cv.glmnet** function.
   f. Doing predictions and validations using validation data.
   g. Steps 2.a to 2.f were repeated for each GCM grid resolutions, including 0.312°×0.312°; 0.5°×0.5°; and 2.5°×2.5°.
   h. Choosing the best grid resolution based on RMSEP value and the correlation coefficient between the predicted and actual rainfall.

III. RESULT

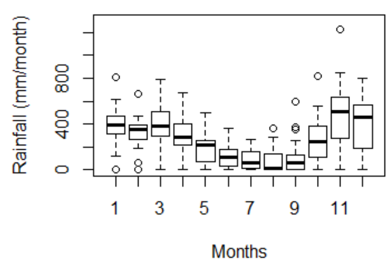
The data exploration was carried out to three different observation stations based on altitude above sea level. The stations were Leles (lowland 0-200 masl), Lengkong (moderate land 200-500 masl), and Sangiang (high altitude> 500 masl) [10]. These stations were used to determine the best grid resolution. The rainfall pattern in Figure 1 showed monthly rainfall increases with altitude above sea level increases. This was in line with the results of the study by Marpaung (2010) which revealed that land with a height of 600-1300 masl has the highest annual rainfall average compared to the land with altitude below 600 meters above sea level. Each station had a pattern similar to the letter U, which means that high rainfall occurs at the beginning and end of the year. This situation was following the monsoon rainfall pattern in West Java.



(a)



(b)



(c)

Figure 1: Rainfall patterns of (a) Leles station (b) Lengkong station (c) Sangiang station

The results of determining grid resolution were shown in Table 1. The grid resolution of 2.5°×2.5° has relatively high correlation and low RMSEP in each type of land. This could be seen from the correlation value respectively for low, medium and plateau were 0.918; 0.868; and 0.790. Also, the RMSEP of 2.5°×2.5° was also relatively low for each type of lands, which were 50.48; 129.60; and 142.37 respectively for low, medium and high land. The best GCM grid resolution was 2.5°×2.5°.

The visualizations of RMSEP and correlation values for each type of land in Table 1 shown in Figure 2 and Figure 3.

Table 1: RMSEP and Correlation for each GCM grid resolution

Station	Type of land	Grid resolution	GCM domain	RMSEP	Correlation y and ŷ
Leles	Low	2.50°×2.50°	13×6	50.48	0.92
		0.50°×0.50°	64×18	62.52	0.84
		0.31°×0.31°	107×28	64.42	0.83
Lengkong	Medium	2.50°×2.50°	13×6	129.60	0.87
		0.50°×0.50°	69×26	129.71	0.87
		0.31°×0.31°	111×26	128.75	0.91
Sangiang	Plateau	2.50°×2.50°	13×6	142.37	0.79
		0.50°×0.50°	63×14	153.49	0.67
		0.31°×0.31°	103×23	152.35	0.67

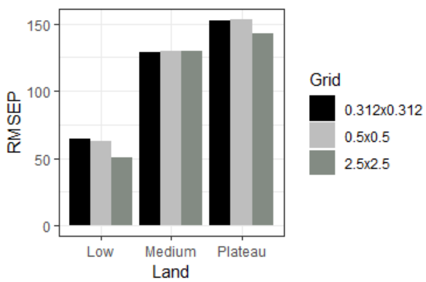


Figure 2: RMSEP for each type of land

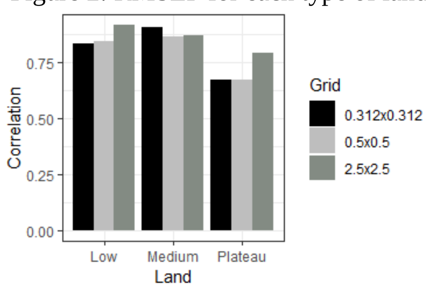


Figure 3: Correlation for each type of land

Previous research, Wigena (2006) which used an 8×8 domain in Indramayu resulted in RMSEP in the range 63-98 and correlation in the range of 0.50-0.76. Santri (2016) which used an 8×8 domain in Indramayu produced RMSEP in the range 67-148 and correlations in the range 0.84-0.94. So that it could be concluded that the grid resolution of 2.5°×2.5° could improve the accuracy of rainfall prediction in the low and medium land

IV. CONCLUSION

The results showed that grid resolution of 2.5°×2.5° was the best grid resolution based on the smallest RMSEP value and the largest correlation in each type of land.

ACKNOWLEDGMENT

This work is fully supported by Kemenristek DIKTI (Kementrian Riset Teknologi dan Pendidikan Tinggi) of Indonesia.

REFERENCES

[1] E. Surmaini, E. Runtunuwu, and I. Las, "Upaya Sektor Pertanian dalam Menghadapi Perubahan Iklim," *J. Litbang Pertan.*, vol. 30, pp. 1–7, 2011.

[2] A. Rosyidie, "Banjir: Fakta dan Dampaknya, Serta Pengaruh dari Perubahan Guna Lahan," *J. Perenc. Wil. dan Kota*, vol. 24, no. 3, pp. 241–249, 2013.

[3] F. Aprian, Y. D. Setianingsi, U. Muntia, A. S. Kartika, S. I. Wicakson, and A. Faru, "Analisis Curah Hujan Sebagai Upaya Meminimalisasi Dampak Kekeringan di Kabupaten Gunung Kidul Tahun 2014," *Khazanah. J. Mhs.*, vol. 6, no. 2, pp. 13–22, 2014.

[4] Suciantini, "Interaksi iklim (curah hujan) terhadap produksi tanaman pangan di Kabupaten Pacitan," vol. 1, no. April, pp. 358–365, 2015.

[5] E. Zorita and H. Von Storch, "The Analog Method as a Simple Statistical Downscaling Technique : Comparison with More Complicated Methods," *J. Clim.*, pp. 2474–2489, 1999.

[6] A. M. Soleh, A. H. Wigena, A. Djuraidah, and A. Saefuddin, "Statistical downscaling to predict monthly rainfall using linear regression with L_1 regularization (LASSO)," *Appl. Math. Sci.*, vol. 9, no. 108, pp. 5361–5369, 2015.

[7] A. H. Wigena, "Pemodelan Statistical Downscaling Dengan Regresi Projection Pursuit Untuk Peramalan Curah Hujan Bulanan (Kasus Curah hujan bulanan di Indramayu)," *Disertation*, 2006.

[8] R. L. Wilby and T. M. L. Wigley, "Precipitation predictors for downscaling: Observed and general circulation model relationships," *Int. J. Climatol.*, vol. 20, no. 6, pp. 641–661, 2000.

[9] A. Busuioc, D. Chen, and C. Hellstro, "Performance of Statistical Downscaling Models in GCM Validation and Regional Climate Change Estimates : Application for Swedish Precipitation," *Int. J. Climatol.*, vol. 21, pp. 557–578, 2001.

[10] B. Nuryanto, A. Priyatmojo, and B. Hadisutrisno, "Pengaruh Tinggi Tempat dan Tipe Tanaman Padi terhadap Keparahan Penyakit Hawar Pelepah," *J. Penelit. Pertan. Tanam. Pangan*, vol. 33, no. 1, p. 1, 2018.

[11] S. Marpaung, "Pengaruh Topografi terhadap Curah Hujan Musiman dan Tahunan di Provinsi Bali Berdasarkan Data Observasi Resolusi Tinggi," *Pros. Semin. Penerbangan dan Antariksa*, pp. 104–110, 2010.

[12] A. S. Ichsan. 2014. Banjir di Indramayu, Ribuan Hektar Sawah Jadi Sungai [Internet]. Jakarta (ID): Republika; [24 June 2019]. https://www.republika.co.id/berita/nasional/jawa-barat-nasional/14/01/19/mzmydz-banjir-di-indramayu-ribuan-hektar-sawah-jadi-sungai

[13] M. K. Dewi. 2019. Potensi Kekeringan Meteorologis di Beberapa Wilayah di Indonesia [Internet]. Jakarta (ID): Meteorological, Climatological, and Geophysical Agency (BMKG); [25 June 2019]. https://www.bmkg.go.id/press-release/?p=potensi-kekeringan-meteorologis-di-beberapa-wilayah-di-indonesia&tag=&lang=ID

[14] E. Garcia. 2010. *A Tutorial on Correlation Coefficients*. https://pdfs.semanticscholar.org/c3e1/095209d3f72ff66e07b8f3b152fab099 edea.pdf .[28 Februari 2019]

[15] D. Santri, "Pemodelan Statistical Downscaling dengan Regresi Kuantil Menggunakan LASSO untuk Pendugaan Curah Hujan Ekstrim, " *Thesis*, 2016

AUTHORS

First Author – Nanda Fadhli S.Pd, college student, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and nanda_senyo@apps.ipb.ac.id

Second Author – Prof. Dr. Ir. Aji Hamim Wigena, M.Sc, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and aji_hw@apps.ipb.ac.id

Third Author – Dr. Ir. Anik Djuraidah, MS, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and anikdjuraidah@apps.ipb.ac.id

Correspondence Author - Prof. Dr. Ir. Aji Hamim Wigena, M.Sc, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, and aji_hw@apps.ipb.ac.id