Statistical Bias Correction of RegCM4 Output Data (Study Area: Indramayu Regency)

Anria1,2, Rizaldi Boer2,3 and Akhmad Faqih3
1Natural Resources and Environmental Management Program, Graduate School of Bogor Agricultural University, Kampus IPB Baranangsiang, Jl. Pajajaran No. 1, Bogor 16143
2Center for Climate Risk and Opportunity Management in Southeast Asia and Pacific, Bogor Agricultural University, Kampus IPB Baranangsiang, Jl. Pajajaran No. 1, Bogor 16143
3Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Kampus IPB Dramaga, FMIPA Building, Wing 19, Level 4, Jalan Raya Dramaga, Bogor, 16680

Corresponding author: andria.anria@gmail.com

Abstract. Climate data that has limitation on spatial and temporal scale becomes an obstacle in perform the climate-related research such us crop simulation, climate risk assessment and many others. Regional climate models such as RegCM4 can produce adequate outcomes of climate data in term of spatial and temporal sides. Output data from RegCM4 can be used for further research once it has been through correction process. On this study, statistical bias correction with three different transfer function (i.e. linear regression, second order polynomials and third order polynomials) were tested. The correction process that has been done on the RegCM4 rainfall data in Indramayu Regency by using third order polynomials-transfer function is able to produce climatology data with value close to observation data. Similarity between observation data and the corrected RegCM4 output data is indicated by similar temporal and spatial distributions. Based on the historical climate and projection of rainfall (1981-2065), it is known that annual rainfall will decrease by 0.84 mm/year.

1. Introduction
Climate data that is difficult to obtain sometimes becomes a barrier in conducting simulations. Regional climate models such as RegCM4 can produce adequate climate data outputs both spatially and temporally. Several previous studies have shown that RegCM output data has a systematic bias [1, 2]. Bias in RCM output data can be caused by several things. Bias in GCM data used as input to RCM is one source of bias in RCM output data [3]. The results of Dosio and Paruelo's research [4] and Macias et al. [5] even shows that bias in GCM data is the main source for bias in RCM output data. Bias in RCM output data can also come from inadequate parameters or from model structures that cannot describe the physical processes of interest [6, 7]. Bias that are often found in RCM output data include inadequate values, a little days without rain caused by high frequency of days with low rainfall intensity and estimates of seasonal variations that are incorrect (under/over-estimate ). The correction process needs to be done to get RegCM data with values close to the observation data.

Many methods have been developed and used in correcting bias in RCM output data. Linear scaling [8-11] and statistical bias correction [12-15] are methods that are often used in the correction process of RCM output data bias. Dasanto et al. [16] applied a statistical bias correction to correct TRMM monthly
rainfall data in Citarum watershed, whereas Jadmiko et al. [17] used a statistical bias correction method to correct RegCM4 monthly rainfall for the West Kalimantan region. In this study, correction was made to the daily rainfall data of RegCM4.

2. Methodology

2.1. Study Area
The area of study in this research is Indramayu Regency, which is part of West Java Province. This regency is located at 107° 52' - 108° 36' E and 6° 15' - 6° 40' S. The regency with an area of 2,099.42 km² is located on the north coast of Java Island (Figure 1). The selection of Indramayu as a study area is based on the important role of this region as a national rice producer. Climate-related studies, such as crop modelling, which require adequate climate data are needed to support role of this area but unfortunately the availability of climate data is limited.

![Figure 1. Study area.](image-url)

2.2. Data
The data used in this study include observational rainfall data, Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS; [18, 19]) and climate data from RegCM4 output with RCP 4.5 scenario produced by Faqih [20].

2.3. Bias Correction of RegCM4 data
To get reliable correction results, the RegCM4 output data correction process is carried out for each grid. The process requires observation data which is also in the form of a grid. Observed rain data available for the Indramayu region is not evenly distributed. Of the 21 existing observation stations, most are in the eastern region (Figure 2). The condition of the data distribution becomes an obstacle in making rain observation in the form of a grid that covers the entire Indramayu region. To overcome this, the CHIRPS data available in grid form is used. The regridding process is first performed to obtain CHIRPS data with the same size and position of the grid with RegCM4 output data.
The correction process for RegCM4 output rain data goes through two stages. The first step is the correction of the CHIRPS data which will replace the observation data and the second step is the correction of the RegCM4 output data itself. In this study, at each stage a correction process is carried out using the linear scaling method and statistical bias correction to get the best correction results. In the statistical bias correction method, the distribution in the CHIRPS and RegCM4 data is corrected with three different transfer functions. The transfer function used is in the form of (1) linear regression, (2) second order polynomials and (3) third order polynomials. Intercepts on all three transfer functions are at points x and y which are 0.

2.3.1. First step - correction of the CHIRPS rainfall data. In this first step, the CHIRPS rainfall data will be corrected, which later will be used to replace the observation data in the RegCM4 rainfall data correction. The data used are observational data and daily CHIRPS data with a period of 1981-2005. The observation rain data used is the average daily rainfall value of 21 stations, while the CHIRPS data is the average daily rainfall value from the grid that has an observation station. The correction process in the first step is carried out for each month. Correction of rainfall data using linear scaling method and correction of statistical bias.

**Linear scaling method**

CHIRPS rain data correction with linear scaling method is multiplication. The initial CHIRPS data is multiplied by the correction factor which is the ratio of the average monthly rainfall observed to the monthly average initial CHIRPS data. The process is explained in equation (1).

\[
CHIRPS_{corr} = CHIRPS_{int} \times \left( \frac{\mu_b P_{obs}}{\mu_b CHIRPS_{int}} \right)
\]

\[(1)\]

**Statistical bias correction method**

The statistical bias correction method begins by making corrections to the frequency of rainy days on the CHIRPS data. It is intended that the frequency of rainy days in the CHIRPS data resembles the frequency of rainy days in the observation data. This process is carried out for each month by cutting the CHIRPS data distribution by a threshold \(P_{ab}\) that can change the frequency of rain in the CHIRPS data to be the same as the observation data. CHIRPS data smaller than the threshold value are converted to 0, while other data that meet the threshold are considered constant (equation (2)).

\[
CHIRPS_{int'} = \int_0^{CHIRPS_{int}} \frac{(CHIRPS_{int} \times P_{ab})}{(CHIRPS_{int} < P_{ab})}
\]

\[(2)\]
Correction to the CHIRPS rain intensity is done after going through the correction process to the frequency of rain. The initial step in the process of correcting the CHIRPS rain intensity is to determine the probability density function (pdf) from the observation data and the model in the form of a gamma distribution using equation (3). Based on the pdf of each data, it can also be known the cumulative distribution function (cdf) by using equation (4).

\[
\text{pdf}(x) = \frac{e^{-\gamma x} x^{\theta - 1}}{\Gamma(\theta) \beta^\gamma} 
\]

(3)

\[
\text{cdf}(x) = \int_0^x \frac{e^{-\gamma x} x^{\theta - 1}}{\Gamma(\theta) \beta^\gamma} \, dx' + cdf(0)
\]

(4)

where: \( \Gamma = \text{gamma function}, k = \text{form parameter}, \theta = \text{scale parameter}, e = \text{natural number} \)

Based on the observation rain data \((y)\) and CHIRPS data \((x)\) with the same cdf value a curve with a certain transfer function can be formed. The transfer function is a correction factor to change the distribution in the CHIRPS data to resemble the distribution in the observas data. In this study an experiment was carried out using three different transfer functions, namely linear regression (equation (5)), second order polynomial (equation (6)) and third order polynomial (equation (7)). Intercepts on all three transfer functions are at points \(x\) and \(y\) which are 0.

\[
\text{CHIRPS}_{\text{corr}} = a \cdot \text{CHIRPS}_{\text{int}} 
\]

(5)

\[
\text{CHIRPS}_{\text{corr}} = a \cdot (\text{CHIRPS}_{\text{int}})^2 + b \cdot \text{CHIRPS}_{\text{int}} 
\]

(6)

\[
\text{CHIRPS}_{\text{corr}} = a \cdot (\text{CHIRPS}_{\text{int}})^3 + b \cdot (\text{CHIRPS}_{\text{int}})^2 + c \cdot \text{CHIRPS}_{\text{int}} 
\]

(7)

Based on the correction method that was tried in the first step, one correction method that is able to correct the CHIRPS data properly will be selected. The assessment is done by looking at the ability of the correction method to improve the average monthly rainfall and the number of rainy days in the CHIRPS data. The best correction method is a method that is able to produce corrected CHIRPS data with an average monthly rainfall value and the number of rainy days close to the observation data. The corrected CHIRPS data is then used for the second stage of the correction process.

2.3.2. Second step- correction of the RegCM4 rainfall data. In this step, corrections were made to RegCM4 output rain data. The data used are corrected CHIRPS data and daily RegCM4 output rain data for the period 1981-2005. The correction process is carried out for each month on each grid. The correction methods used in the second stage are the same as the previous step.

**Linear scaling method**

Correction of RegCM4 output rainfall data with linear scaling method is done through multiplication between the initial RegCM4 data with the correction factor which is the ratio of the monthly rainfall CHIRPS-corrected to the average monthly rainfall from the initial RegCM4 data (equation (8)).

\[
\text{RegCM4}_{\text{corr}} = \text{RegCM4}_{\text{int}} \times \left[ \frac{\mu_b \cdot \text{CHIRPS}_{\text{corr}}}{\mu_b \cdot \text{RegCM4}_{\text{int}}} \right] 
\]

(8)

**Statistical Bias Correction**

Corrections to RegCM4 rainfall frequency are carried out using equation (9) below.

\[
\text{RegCM4}_{\text{int}} = \left\{ \begin{array}{l}
\text{RegCM4}_{\text{int}} \cap P_{ab} \\
\text{RegCM4}_{\text{int}} < P_{ab}
\end{array} \right\} 
\]

(9)
The transfer function used in the rainfall intensity correction process of the RegCM data with the statistical bias correction method are as follows:

\[ R_{\text{RegCM}} \text{corr} = a \cdot R_{\text{RegCM}, \text{int}}' \]  \hspace{1cm} \hspace{1cm} (10)\\
\[ R_{\text{RegCM}} \text{corr} = a \cdot (R_{\text{RegCM}, \text{int}}')^2 + b \cdot R_{\text{RegCM}, \text{int}}' \]  \hspace{1cm} \hspace{1cm} (11)\\
\[ R_{\text{RegCM}} \text{corr} = a \cdot (R_{\text{RegCM}, \text{int}}')^3 + b \cdot (R_{\text{RegCM}, \text{int}}')^2 + c \cdot R_{\text{RegCM}, \text{int}}' \]  \hspace{1cm} \hspace{1cm} (12)

The best correction method that is able to correct RegCM4 rain data will be chosen. The assessment criteria used at this stage are the same as the previous stage, which is based on the ability of the correction method to improve the average monthly rainfall value and the number of rainy days in RegCM4 rainfall data. The best correction method is a method that is able to generate corrected RegCM4 rain data with an average monthly rainfall value and the number of rainy days close to the corrected CHIRPS data. In addition, the results of the correction with the spatial distribution that is closest to the observation data are also taken into consideration in determining the best correction method.

3. Results

To get reliable correction results, the RegCM4 output data correction process is carried out for each grid. The process requires observation data which is also in the form of a grid. Observed rain data available for the Indramayu region is not evenly distributed. Of the 21 existing observation stations, most are in the eastern region (Figure 2). The condition of the data distribution becomes an obstacle in making observation rain in the form of a grid that covers the entire Indramayu region. To overcome these obstacles, CHIRPS data are available in the form of grids. The regridding process is first performed to obtain CHIRPS data with the same size and position of the grid with RegCM4 output data (Figure 3).
3.1. First step - correction of the CHIRPS rainfall data

Based on climatological monthly rainfall data (monthly averages from 1981-2005 data), the CHIRPS data has the same pattern as observational data, but the values tend to be higher than observational rainfall data (Figure 4a) which indicate with green dots in Figure 4b. The correction process for CHIRPS data is carried out to get the CHIRPS data corrected with values close to observation.

**CHIRPS data that has been corrected by the linear scaling method (Figure 4c) as well as the statistical bias correction method with transfer functions in the form of polynomials of order 3 (Figure 4d) and order 2 (Figure 4e) can improve the average monthly rainfall in Indramayu Regency. Statistical bias correction with linear transfer function has not been able to improve the value of rain during high rainfall such as in November, December, January and February (Figure 4f).**

**Figure 4.** Monthly rainfall on observation data, CHIRPS and CHIRPS-corrected by four different methods.
Figure 5. The number of rainy days on observational data, CHIRPS and CHIRPS-corrected by four different methods.

The number of rainy days for each month in the CHIRPS data, especially in the rainy season (November, December, January and February), tends to be higher than the observation data (Figure 5a) which indicate with blue dots in Figure 5b. Correction using the linear scaling method is not able to improve the amount of rain in the CHIRPS data (Figure 5c). The correction process using the statistical bias correction method that begins with the determination of the rain threshold in the CHIRPS data is proven to be able to improve the condition so that the number of rainy days approaches the observation condition. Statistical bias correction with Order 3 polynomial transfer function (Figure 5d) shows the number of rainy days which is the closest to the number of rainy days in the observation data.

By considering the average monthly rainfall conditions and the number of rainy days from the CHIRPS data corrected for each method, it was concluded that the statistical bias correction method with the transfer function in the form of polynomial order 3 is the best correction method that is able to improve the CHIRPS data so that it can resemble observational data. Comparison between observational data, CHIRPS and CHIRPS corrected with the best method for daily (Figure 6) and monthly (Figure 7) rain conditions can be seen below.
For daily rainfall conditions, especially for very high rainfall events, the corrected CHIRPS data cannot resemble observational data. Judging from the monthly rain conditions, the corrected CHIRPS data has a pattern that resembles observational data. The correlation between the two data is also high ($R^2=0.78$).

3.2. Second step- correction of the RegCM4 rainfall data

The second stage of the correction process for RegCM4 output rain data is carried out for each grid by utilizing the corrected CHIRPS data. Correction of bias to RegCM4 output rain data needs to be done because the data has a value that tends to be higher than the corrected CHIRPS data (Figure 8a) which indicate with green dots in Figure 8b. The second stage of bias correction was carried out to obtain the corrected rain output RegCM4 data with values approaching the observation data represented by the corrected CHIRPS data.
Figure 8. Monthly rainfall on CHIRPS-corrected, RegCM4 and RegCM4-corrected by four different methods.

RegCM4 rainfall data that has been corrected by the linear scaling method (Figure 8c) is able to improve the average monthly rainfall value in Indramayu Regency. Similar to the linear scaling method, correction method by using statistical bias correction (Figure 8d-8f) has also been shown to be able to improve the average monthly rainfall value.
The number of rainy days for each month in the RegCM4 output rain data is always higher than the corrected CHIRPS data (Figure 9a), which indicate with green dots in Figure 9b, even the value is very high in the dry months. Correction with the linear scaling method can improve the amount of rain in RegCM4 output rain data, but is still higher than the corrected CHIRPS data (Figure 9c). The correction process with a statistical bias correction method that begins with the determination of the rain threshold in RegCM4 output rain data is able to improve these conditions so that the number of rainy days approaches the corrected CHIRPS condition. Statistical bias correction with Order 3 polynomial transfer function (Figure 9d) shows the number of rainy days which is the closest to the number of rainy days in the corrected CHIRPS data.

Taking into account the average monthly rainfall conditions and the number of rainy days from the corrected RegCM data in each method, it was concluded that the statistical bias correction method with the transfer function in the form of polynomial order 3 is the best correction method that is able to improve RegCM4 output rain data so that it can resemble the data CHIRPS corrected. Comparison between corrected CHIRPS data, RegCM4 and RegCM4 corrected with the best method for daily (Figure 10) and monthly rain conditions (Figure 11) can be seen below.
Judging from the monthly rainfall conditions, the chosen correction method has been able to improve RegCM4 output rain data whose intensity exceeds the corrected CHIRPS data. The rainfall intensity in the corrected RegCM4 data is close to the corrected CHIRPS data. Corrected RegCM4 has a monthly rainfall pattern that resembles corrected CHIRPS data. The correlation between the two data is also high with a value of 0.7.

Comparison between corrected CHIRPS data, RegCM4 and RegCM4 spatially corrected for Indramayu Regency can be seen in Figure 12. The corrected RegCM4 data has a distribution that resembles corrected CHIRPS data, where rainfall in the south tends to be higher than in the northern region. Both data also show that the peak of the rainy season occurs in January and the dry season occurs in August-September.
CHIRPS-corrected  RegCM4  RegCM4-corrected

Jan

Feb

Mar

Apr

Mei

Jun
Figure 12. Spatial distribution of monthly rainfall for CHIRPS-corrected, RegCM4 and RegCM4-corrected.
3.3. Indramayu Rainfall Conditions based on corrected RegCM4 output data

By setting the historical data time period is 1981-2005, while the projection data is divided into two, namely 2011-2035 and 2041-2065, historical monthly rainfall and projections in Indramayu Regency can be seen in Figure 13a. Based on this picture, it is known that the rainfall patterns in Indramayu Regency, both for historical periods and projections, are monsoonal. The peak of the rainy season is in December-January-February (DJF), while the peak of the dry season occurs in June-July-August (JJA). In the monsoonal rainfall pattern, the transition from rain to dry season occurs in March-April-May (MAM) and the transition season from dry to rain in September-October-November (SON).

**Figure 13.** (a) Historical and projected monthly rainfall in Indramayu Regency and (b) changes in projected monthly rainfall (%) compared to historical monthly rainfall in Indramayu Regency.

Changes in the projected rainfall over historical rainfall can be seen in Figure 13b. The highest decrease in monthly rainfall in the future occurs in the dry season (JJA). In the future rainfall in the dry season will decrease further, where the highest decline occurs in August. Based on rainfall data for 1981-2065 it is known that annual rainfall in Indramayu Regency has a tendency to decrease (Figure 14). The annual reduction in rainfall is 0.84 mm/year.

**Figure 14.** Annual rainfall for the period 1981-2065 in Indramayu Regency.

The trend of changes in seasonal rainfall for the period 1981-2065 can be seen in Figure 15. Rainfall in the rainy season (DJF), dry season (JJA) and the transition season from rain to dry (MAM) has decreased. Different conditions occur during the transition from the dry season to the rainy season (SON) where there is a tendency to increase in the future.
Figure 15. Seasonal rainfall for the period 1981-2065 in Indramayu Regency.

References

[1] Christensen JH, Boberg F, Christensen OB, Lucas-Picher P 2008 On the Need for Bias Correction of Regional Climate Change Projections of Temperature and Precipitation Geophysical Research Letters. 35 20
[2] Varis O, Kajander T, Lemmelä R 2004 Climate and Water: from Climate Models to Water Resources Management and Vice Versa Climatic Change 66 321-44
[3] Fowler HJ, Blenkinsop S, Tebaldi C 2007 Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling International journal of climatology 27 1547-78
[4] Dosio A, Paruolo P 2011 Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: evaluation on the present climate Journal of Geophysical Research: Atmospheres 116 (D16)
[5] Macias D, Garcia-Gorriz E, Dosio A, Stips A, Keuler K 2016 Obtaining the correct sea surface temperature: bias correction of regional climate model data for the Mediterranean Sea. Climate Dynamics 1-23
[6] Allen M, Frame D, Kettleborough J and Stainforth D 2006 Model error in weather and climate forecasting in: Predictability of Weather and Climate. Edited by: Palmer T and Hagedorn R Cambridge University Press 391–427
[7] Schoetter R, Hoffmann P, Rechid D, Schlünzen KH 2012 Evaluation and bias correction of regional climate model results using model evaluation measures. Journal of Applied Meteorology and Climatology 51 1670-84
[8] Lenderink G, Buishand A, Deursen WV 2007 Estimates of Future Discharges of The River Rhine Using Two Scenario Methodologies: Direct Versus Delta Approach Hydrology and Earth System Sciences 11 1145-59
[9] Graham LP, Andréasson J, Carlsson B 2007 Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods–a case study on the Lule River basin *Climatic Change* 81 293-307

[10] Teutschbein C dan Seibert J 2012 Bias Correction of Regional Climate Model Simulations for Hydrological Climate-Change Impact Studies: Review and Evaluation of Different Methods *Journal of Hydrology* 456 12-29

[11] Fang G, Yang J, Chen YN, Zammit C 2015 Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China *Hydrology and Earth System Sciences* 19 2547-59

[12] Piani C, Haerter JO, Coppola E 2010 Statistical bias correction for daily precipitation in regional climate models over Europe *Theoretical and Applied Climatology* 99 187-92

[13] Watanabe S, Kanae S, Seto S, Yeh PJ, Hirabayashi Y, Oki T 2012 Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models *Journal of Geophysical Research: Atmospheres* 117(D23)

[14] Ines AV, Hansen JW 2006 Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and forest meteorology* 138 44-53

[15] Faqih A 2017 A Statistical Bias Correction Tool for Generating Climate Change Scenarios in Indonesia based on CMIP5 Datasets. *In IOP Conference Series: Earth and Environmental Science* 58(1):012051

[16] Dasanto BD, Boer R, Pramudya B, Suharnoto Y 2014 Evaluasi Curah Hujan TRMM Menggunakan Pendekatan Koreksi Bias Statistik *Jurnal Tanah dan Iklim* 38 15-24

[17] Jamdiko SD, Murdiyarso D, Faqih A 2017 Koreksi Bias Luaran Model Iklim Regional untuk Analisis Kekeringan *Jurnal Tanah dan Iklim* 41 25-35

[18] Funk CC, Peterson PJ, Landsfeld MF, Pedreros DH, Verdin JP, Rowland JD, Romero BE, Husak GJ, Michaelsen JC, Verdin AP 2013 A quasi-global precipitation time series for drought monitoring. US Geological Survey No.832

[19] Funk CC, Verdin A, Michaelsen J, Peterson P, Pedreros D, Husak G 2015 A global satellite-assisted precipitation climatology *Earth Syst. Sci. Data* 8 401-425

[20] Faqih A 2016 Historical climate and Future climate scenarios In Indonesia. [Final Report of Climate Modelling and Analysis for Indonesia 3rd National Communication]. Ministry of Environment and Forestry, Republic of Indonesia (MoEF). Indonesia