Determining best East Java monthly rainfall projection using spatial-based validation

R R Mahmudiah\(^1\,\!^2\), A Kurniawan\(^1\,\!^2\) and E E S Makmur\(^3\)

\(^1\)Malang Climatological Station, Malang, 65152, Indonesia
\(^2\)Indonesian State College of Meteorology Climatology and Geophysics, Perhubungan Street 5, Pondok Aren, South Tangerang, 15221, Indonesia
\(^3\)Research and Development Center, Meteorology Climatology and Geophysics Agency, Angkasa I Street 2, Kemayoran, Jakarta, 10720, Indonesia

rikharizki.mahmudiah@gmail.com

Abstract. In production of climate change information, determining the best projection is a very important step. Missed projection can lead to loss of public trust in climate change. In some previous studies, most analyses were conducted on point approach due to the lack of observational data covering large spatial areas. In this study, rainfall data from 197 observation points are used to get correction coefficients by comparing it to monthly rainfall historical model from 1972-2005. Area of study is 110.89°E–116.27° E and 8.78°S–5.04°S. Two RCP scenarios (4.5 and 8.5) under CSIROMK3.6 RegCM4 BMKG SEA-Cordex are compared to get its ensemble weighting factor. Validation parameters used are mean absolute error and quartiles of error. Shifting correction is introduced in this study with some evidence based on visual analysis and high correction coefficients from the unshifted one. It is also shown that bilinear resampling gives better result than nearest neighbour. Ensemble weighting factors are then set 0.4 for RCP 8.5 and 0.6 for RCP 4.5. In next 10 years (2017-2026), dry condition average rainfall condition is expected to happen between May and October.

Keywords: Projection, rainfall, correction coefficient, bilinear, shifting correction

1. Introduction

Projection is information about how far climatic parameter will change. It is calculated using greenhouse gases (GHG) scenarios, global climate models (GCMs), and regional climate models (RCMs). The variation of its possibilities is so high due to uncertainty carried by its modelling process. Determining or choosing the best one is a very important step. Climate change itself refers to statistically significant variation in either the mean state of the climate or in its variability, including for periods extended (typically decades or longer). Climate change may be due to persistent anthropogenic changes in the atmosphere of land-use. The change is clear in global level but the complexity is increasing drastically for more local area [1].

In Indonesia, rainfall is more important than temperature. Indonesian season also determined not by temperature but rainfall. Seasonal prediction is then focused only on rainy-dry season and its properties such as its start, length, and rainfall amount compared to normal condition [2]. Rainfall is also the primary driver of streamflow in tropical regions [3]. Because of it, this climate change study is also will be focused on rainfall.
BMKG (Indonesian meteorology climatology and geophysics agency) also contributes to the South-East Asia Coordinated Regional Climate Downscaling Experiment (SEA-Cordex). In this framework, there are two Representative Concentration Pathways (RCP) used. RCP 8.5 is assumption that condition similar to A2R SRES is happen or business as usual is carried out [4] while RCP 4.5 is assumption that realistic adaptive and mitigative policies are agreed so a constant population after 2050 [5]. GCM used is CSIROMK3.6 [6] and RCM used is RegCM4 [7]. There is possibility to identify the best projection due to that it is been at least 10 years after the last historical year (2005) has passed.

East Java Province is the provincial administration with the biggest in number of local government under it. 38 local governments are under its administration. For making comprehensive climate information for those stakeholders, Malang Climatological Station collects and maintains at least 197 rainfall data each month. This number is also the highest among other province. It is interesting to investigate climate change in term of monthly rainfall projection in this area (spatial-based). So, the aim of this study is determining best East Java monthly rainfall projection using spatial-based validation.

2. Methodology

2.1. Area of Study and Data

This study is focused in East Java Province which is astronomically bordered by 110.89°E–116.27° E and 8.78°S–5.04°S. 197 points used in this study are the same points used for operational monthly climate monitoring. In figure 1, the dots from right image are the 197 observation points and the black filled area shows the area position inside maritime continent.

![Figure 1. Area of the study.](image)

The rainfall data are collected by Malang Climatological Station. The data is not available at all time. Each observation point began to operate at different times. Data that is not available or suspicious is removed and denoted as NA values. The figure 2 shows the data availability. A rising pattern happened near 1981, 1991, and 2001 due to regular update of climate data normal. Availability reaches above 80% after 1990.

Projection data used in this study are monthly rainfall projections output of CSIROMK3.6 RegCM4 models ran by BMKG with RCP 4.5 and 8.5 in the period for 2006-2099 (94 years). The model has historical data from 1970-2005. This data is so popular among students of Indonesian State College of Meteorology Climatology and Geophysics. It is being used as under graduate research material and also be one of data used in climate change analysis course.

In addition, polygon shapes file from Malang Climatological Station also used in shifting detection. The shapes file is under standard of National Geospatial Agency (BIG). Shuttle Radar Topography Mission (SRTM) data [8] is also used to create elevation contour.
2.2. Method
Because this study orientation is spatial through some area not points, the first thing to do is making sure that the model data is spatially correct. The historical data is then mapped using Grid Analysis and Display System (GrADS) with shaded output. Default colours level and Indonesian National Standard no. 8196:2015 about colours level of monthly rainfall mapping issued by Badan Standardisasi Nasional (BSN) are used. Contour of elevation using 1000 m as interval is also used in the mapping process.

In investigating spatial correctness of the model, correction coefficient (CC) is used. CC is similar to multiplicative factor used in climate change bias correction study [9]. CC is calculated for every available time if there is observation data at the time. The historical model data is extracted \( P_{\text{mod}} \) then used as divisor to observation data \( P_{\text{obs}} \) at same point and month \( (p,m) \). Model output having zero value can make some error in CC calculation. 0.01 is then used as substitution for zero value due to possibility of error from division by zero. The result will be 197 x 12 = 2364 CC for every extraction method.

\[
CC(p,m) = \frac{\sum_{y=1970}^{2005} P_{\text{obs}}(p,m,y)}{\sum_{y=1970}^{2005} P_{\text{mod}}(p,m,y)}
\]  

(1)

Extraction method used in this study is nearest neighbor (NN) and bilinear interpolation (BL). NN will take one nearest model point from observation coordinate so the possibility of same value between close points is relatively higher. On the other hand, BL uses four nearest points to calculate some value in observation coordinate [10].

The CC is then multiplied with 2006-2016 average monthly rainfall from projection model for both RCP 4.5 and RCP 8.5. To determine the better one, mean absolute error (MAE) is used due to its uncomplicated interpretation [11]. MAE is calculated from 2364 error values. Quartile1 (Q1), median (Med), and quartile 3 (Q3) from those error values are also calculated.

\[
MAE = \frac{1}{197} \sum_{n=1}^{197} |P_{\text{obs}}(p,m) - P_{\text{mod}}(p,m) \times CC(p,m)|
\]  

(2)

The plan was actually only choosing between NN and BL so that the one with lower error will be used as extraction method. Due to its spatial orientation, this study interpolates the best CC with inverse distance weighting (IDW) [12] with parameter power =1.5, point used in interpolation = 9, and resolution = 0.01°. Resampling the average projection model data from period 2017-2026 with the best extraction method to resolution 0.01 is also done so the resolution is same with CC raster. Ensemble projection using RCP 4.5 and RCP 8.5 is done by weighting. The weighting factor is got from percentage ratio of when and where RCP 4.5 is better than RCP 8.5.
3. Result and Discussion

This first step of the method is factually unconsidered before, but plotting experience near Madura islands made it is clear that there is something wrong about the historical data. The first case that is found is Jan 1972 historical rainfall as shown in figure 3. For making sure that it is not the only case, the first step of method is done.

The result of mapping shows that using GrADS default colours, at least there is 32 cases that Madura islands rainfall is look-like being shifted to North-East. Using SNI colours template, 33 cases is found. Justification that the model is wrong cannot be done in this study but the fact shows that there is enough evidence that there is shifting in coordinate of model. So, for this study, using some basic intuitive visual analysis, it is decided that there is possibility to correct it by shifting the historical data model to South-West one resolution each. The example of change due to shifting correction (SC, this term will also will be used as shifting corrected) can be seen in Figure 1.

![Figure 3. Mapping of rainfall historical data: (a) GrADS default colours, (b) SNI colours, and (c) shifting-corrected data.](image)

The impact of SC can also be analysed using CCs that are produced. Lower CC is produced when SC is done. Figure 4 shows an example of how the high CC values near south coast of East Java is decreased when SC is applied. It encourage some principle that differently from continental area [13] or global scale [9], maritime continent climate change study needs more attention to its spatial data precision. This study can also give more detail of some bias that may exist from previous study with larger scale [14]. CC in figure 4 is calculated using NN extraction method. It can be stated that maybe there is some other mechanism that makes it better in corrected but at least by this one jump to South-West, the CC is more logically accepted.

![Figure 4. CC for March of (a) before SC and (b) after SC.](image)
For further study and practical use in course of climate change analysis, the extraction script for R Statistics user that can be used is as follows. The raster package used in the script is written by Hijmans et al [15].

```r
>library(raster)
>SC=raster("...\csromk3.6-hist-pr.nc")
>res(SC)
>cor_lon=-1*res(SC)[1]
>cor_lat=-1*res(SC)[2]
>SCm=extent(SC)
>SCm[1]= SCm [1]+cor_lon
>SCm[2]= SCm [2]+cor_lon
>SCm[3]= SCm [3]+cor_lat
>SCm[4]= SCm [4]+cor_lat
```

Then apply “SCm” as the new extent before extracting historical model.

Other impact of using SC to CC can also be shown in figure 5. In this figure, CC is distributed each month. CC of NN without SC shows values as high as 100. It means that in some case, when rainfall model value is 1 mm, the observation is above 100 mm. Very high CC is not acceptable because if CC itself shows the ratio between model and reality captured in observation. Furthermore, using BL method as extraction method can also decrease high CC values. It can be decided that to determine best projection using no SC will give bad result.

![Figure 5. Correction coefficients boxplot for rainfall data.](image)

Although CC from BL method does having lower CC, it does not mean the best projection will be come from BL. Using MAE, the error from model projection in 2006-2016 is calculated. From table 1, BL method’s superiority can be shown. Its MAE is lower than NN method. Errors’ Q1, Med, and Q2 of BL are also lower than NN method. It can be confirmed that BL is better than NN for extracting the model data.

| Table 1. Errors from shifting corrected NN and BL methods. |
|-----------------------------------------------------------|
| **Nearest Neighbour** (NN) | **Bilinear** (BL) |
| MAE | RCP 4.5 | RCP 8.5 | RCP 4.5 | RCP 8.5 |
| Errors Q1 | 18.36 | 24.62 | 15.51 | 17.62 |
| Errors Med | 55.05 | 72.18 | 39.61 | 45.79 |
| Errors Q3 | 138.06 | 179.01 | 90.13 | 98.04 |
It is simple that BL is better than NN but choosing the best between RCP 4.5 and RCP 8.5 cannot be differentiated clearly. MAE of RCP 8.5 is lower than RCP 4.5 but 3 quartile parameters show the opposite. Because of its inconsistency, the ensemble method is then used. For making better ensemble, weighting is needed but the weighting factor is calculated need some justification.

In this study, how better RCP 4.5 than RCP 8.5 is then examined using both spatial and temporal approach. Spatially, conditions that RCP 4.5 better than RCP 8.5 cannot be stated clearly. 58.01% from 197 points states that RCP 4.5 is better. It means that average percentage of condition from 12 months at the point is 58.01%. For example, at a specific point, if condition RCP 4.5 is better than RCP 8.5 is 7 month so the percentage at the point will be 7 divided by 12 or around 58%. Percentages from all 197 points are then averaged. From temporal perspective, it is also unclear which RCP is better. RCP 4.5 is better in all month except August, October, November, and December. Using same averaging approach with the spatial separation, 57.92% conditions show RCP 4.5 is better. Figure 6 shows both temporal and spatial (RCP 4.5 oriented).

Figure 6. Temporal (left) and spatial (right) distribution of condition that RCP 4.5 better than RCP 8.5.

In projection, separated model based on month can be used. For example, RCP 4.5 will be used for eight months and the rest will use RCP 8.5. Climate change is a global system so that the impact to regional or more local one can be separated like that. From the ratio of spatial and temporal that RCP 4.5 better than RCP 8.5 is around 60%, the ensemble weighting in this study is decided to be 0.6 for RCP 4.5 and 0.4 for RCP 8.5. Other weighting can be taken but these possibilities will not be done in this study.

The CC from BL with SC is then interpolated using System for Automated Geoscientific Analyses (SAGA) [16]. The interpolation results then multiplied as raster with resampled RCP 4.5 projection. Same procedure is carried out for RCP 8.5. The RCP 4.5 raster is then multiplied with 0.6 and the RCP 8.5 raster with 0.4. These two rasters then combined by raster calculator for every month. So, the result can be seen in figure 7.

Period of next 10 years is used in this so that if there is some urgency, stakeholder can use the information. Longer period will only decrease practical benefit of projection. It is projected that in next 10 years, average condition from November till April in mountainous area will be wet. Topography still plays a big role in determining rainfall distribution. Based on the result, peak dry condition will happen between July and September. Relatively wetter condition will back in November. May till October is expected as dry period for almost East Java. Further analyses can be done by making raster from this study as input.
4. Conclusion
The first aim of this study is factually determining best projection, but the side result give wider perspective. This study brings new standard of pre-testing in spatial correctness of model before using it. Correction coefficient (CC) is proposed not only as corrector but also parameter to determine correctness of model’s coordinate. Although bring no good answer, this study at least shows that extraction using bilinear is better and demonstrate the possibility of making ensemble weighting factor justification from spatio-temporal error analysis. In the nutshell, dry condition for next 10 years (2017-2026) is expected to happen between May and October.

Acknowledgments
We would like to thank Malang Climatological Station and Indonesian State College of Meteorology Climatology and Geophysics (STMKG) for supporting this study by providing the data and opportunities.

References
[1] IPCC 2014 Climate Change 2014: Synthesis Report In Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change ed Pachauri RK and Meyer LA (Geneva: IPCC) p 151
[2] BMKG 2012 Verifikasi Prakiraan Iklim Indonesia (Jakarta: BMKG)
[3] Yanto, Livneh B, Rajagopalan B and Kasprzyk 2017 Hydrological model application under data scarcity for multiple watersheds, Java island, Indonesia J. of Hydrology: Regional Studies 9 127-139
[4] Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N and Ralaj P 2011 RCP 8.5—a scenario of comparatively high greenhouse gas emissions Climatic Change 109 33–57
[5] Thomson A M, Calvin K V, Smith S J, Kyle G P, Volke A, Patel P, Delgado-Arias S, Bond-Lamberty B, Wise M A, Clarke L E and Edmonds J A 2011 RCP4.5: A pathway for stabilization of radiative forcing by 2100 Climatic Change 109 77–94

[6] Jeffrey S, Rotstayn L, Collier M, Dravitzki S, Hamalainen C, Moeseneder C, Wong K and Syktus J 2013 Australia’s CMIP5 submission using the CSIRO Mk3.6 model Aust. Meteor. Oceanogr. J. 63 1-13

[7] Giorgi F, Coppola E, Solmon F, Mariotti L, Sylla M B, Bi X, Elguindi N, Divo G T, Nair V, Giuliani G, Turuncoglu U U, Cozzini S, Güttler I, O’Brien T A, Tawfik A B, Shalaby A, Zakey A S, Steiner A L, Storda F, Sloan L C and Brankovic C 2012 RegCM4: model description and preliminary tests over multiple CORDEX domains Clim. Res. 52 7-29

[8] Yang L, Meng X and Zhang X 2011 SRTM DEM and its application advances Int. J. Remote Sensing 32 3875-3896

[9] Ngai S T, Tangang F and Juneng L 2016 Bias correction of global and regional simulated daily precipitation and surface mean temperature over Southeast Asia using quantile mapping method Global and Planetary Change 149 79-90

[10] Teegavarapu R S V, Meskele T and Pathak C S 2012 Geo-spatial Grid Based Transformations of Precipitation Estimates Using Spatial Interpolation Methods Computers & Geosciences 40 28–39

[11] Willmot C and Matsuura 2005 Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance Clim. Res. 30 79-82

[12] Yanto, Livneh B and Rajagopalan B 2017 Development of gridded meteorological dataset over Java island, Indonesia 1985-2014 Sci. Data 4 170072

[13] Ahmed K F, Wang G, Silander J, Wilson A M, Allen J M, Horton R and Anyah R 2019 Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast Global and Planetary Change 100 320-332

[14] Siew J H, Tangang F T and Juneng L 2014 Evaluation of CMIP5 coupled atmosphere–ocean general circulation models and projection of the Southeast Asian winter monsoon in the 21st century Int. J. Climatol. 34 2872-2884

[15] Hijmans R J, Etten J V, Sumner M, Cheng J, Bevan A, Bivand R, Busetto L, Canty M, Forrest D, Ghosh A, Golicher D, Gray J, Greenberg J A, Hiemstra P, Karney C, Mattiuzzi M, Mosher S, Nowosad J, Pebesma E, Lamigueiro O P, Racine E B, Rowlingson B, Shortridge A, Venables B and Wueest R 2017 Geographic Data Analysis and Modeling https://cran.r-project.org/web/packages/raster/raster.pdf accessed September 2018

[16] Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V and Böhner J 2015 System for Automated Geoscientific Analyses (SAGA) v. 2.1.4 Geosci. Model Dev. 8 1991-2007