Evaluation of the North American multi-model ensemble for monthly precipitation forecast

D Y Faidah\textsuperscript{1,2*}, H Kuswanto\textsuperscript{2} and Suhartono\textsuperscript{2}

\textsuperscript{1}Department of Statistics, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia
\textsuperscript{2}Department of Statistics, Kampus ITS Sukolilo, Surabaya 60111, Indonesia

*defi.yusti@unpad.ac.id

Abstract. The North American multi-model ensemble (NMME) is a multi-model seasonal forecasting system consisting of a collection of models generated from several climate modelling centers. This research examined the monthly precipitation in North Maluku generated by five NMME models. The purpose of this research is to assess the performance of monthly precipitation prediction by using RMSE and Rank Histogram analysis. The NMME models are verified against observed precipitation. The analysis shows that they are biased and underdispersive. Among the five NMME models, the Center for Ocean-Land-Atmosphere Studies (COLA) exhibits the best predictive skill. The performances of the Canadian Meteorological Centre (CMC) are relatively worse than that of the other models. The COLA model shows relatively high skill when used to forecast May-November monthly precipitation. Meanwhile, the National Oceanic and Atmospheric Administration (NOAA)’s Geophysical Fluid Dynamics Laboratory (GFDL) model shows high skill in December-April periods. The ensemble forecast is calibrated with the BMA approach in order to obtain reliable forecasts.

1. Introduction
Precipitation information is continuously processed to develop more accurate forecasting methods. There are various methods of predicting precipitation, i.e. from individual and hybrid models to a combination of several individual forecasting models. Combining the results of predictions made using several individual models into a forecast value is one form of ensemble forecast. The North American Multi-Model Ensemble (NMME) is a collection of seasonal forecast models generated by the modeling center in the United States. Kirtman, et al. \cite{1} explained that the ensemble forecast is more accurate in forecasting weather and climate. The NMME dataset has been extensively used to verify monthly precipitation forecast \cite{2,3,4,5}. However, so far there have been only a few studies examining the performance of NMME forecasts in Indonesia \cite{6,7,8}. Kuswanto, et al. \cite{9} explained that the results of these studies cannot be generalized to all regions since Indonesia has three rainfall patterns with different characteristics. The purpose of this research is to evaluate the performance of NMME in North Maluku as one of the areas with high precipitation in Indonesia. The ensemble prediction performance will be evaluated using RMSE and rank histogram. Furthermore, the NMME dataset is calibrated to improve the reliability of the forecast. Precise and reliable precipitation forecasts are used as the basis for early warning of disasters in the hydrological field \cite{10,11}. One of the methods that can be used to calibrate the NMME dataset is Bayesian Model Averaging (BMA). BMA has optimal properties in the form of intervals and their use has performed well in several case simulations and real situations \cite{12}.
2. Method

2.1. Goodness of Fit

The analysis begins with a time series plot to determine the precipitation pattern. Furthermore, the ensemble forecast is verified using RMSE and rank histogram. RMSE is used to evaluate the performance of the NMME dataset in predicting precipitation. Rank histogram is obtained by sorting the ensemble members from the lowest to the highest. The criterion for a good rank histogram is the formation of a straight line in which every member of the ensemble has the same frequency [13]. The criteria for rank histogram are shown in Figure 1.

![Figure 1. Criteria of rank histogram](image)

Source: [13]

2.2. Bayesian Model Averaging (BMA)

The forecast results of the ensemble are usually under dispersive or over dispersive, which means that the forecast is unreliable. This problem can be solved by calibrating the ensemble forecast. One method that can be used to calibrate ensemble forecasts is BMA, which was first introduced by Raftery. The idea of this method is to calibrate the ensemble model in one equation that represents all ensemble models [12]. The mechanism of BMA is to give precise weight to the forecast ensemble. Weights are given based on the predictive ability of each ensemble model. Suppose that, \( y \) is the observed value where \( f_1, f_2, ..., f_M \) is the ensemble predicted value obtained from the different \( M \) model, then the BMA’s prediction model for ensemble forecast follows equation [14]:

\[
p(y | f_1, f_2, \ldots) = \sum_{m=1}^{M} w_m g_m(y | f_m)
\]  

(1)

where \( w_m \) is the weight that describes the predictive ability of \( m \)-model. \( g_m(y | f_m) \) is a conditional PDF of \( y \) at \( f_m \) if \( y \) is the calibrated forecast and \( f_m \) is the best forecast in the ensemble member.
### 3. Result and Discussion

#### 3.1. Research Data

The data used in this study consisted of monthly precipitation predictions from the ensemble model and precipitation observations in North Maluku from January 1982 to December 2018. The data were obtained from the official websites of NMME and the official website of the Meteorological, Climatological, and Geophysical Agency. In this study, the NMME models used for precipitation prediction are Canadian Coupled Climate Model versions 3 (CanCM3), Canadian Coupled Climate Model versions 4 (CanCM4), Community Climate System Model version 3 (CCSM3), Community Climate System Model version 4 (CCSM4) and Geophysical Fluid Dynamics Laboratory Climate Model version 2p1 (GFDL-CM2p1).

| Model     | Modeling Center                                      | Ensemble Members | References |
|-----------|------------------------------------------------------|------------------|------------|
| CanCM3    | Environment Canada's Canadian Meteorological Centre (CMC) | 10               | [15]       |
| CanCM4    | Environment Canada's Canadian Meteorological Centre (CMC) | 10               | [15]       |
| CCSM3     | The Center for Ocean-Land Atmosphere Studies (COLA)   | 6                | [16]       |
| CCSM4     | The Center for Ocean-Land Atmosphere Studies (COLA)   | 6                | [17]       |
| GFDL-CM2p1| National Oceanic and Atmospheric Administration (NOAA)’s Geophysical Fluid Dynamics Laboratory (GFDL) | 10               | [18]       |

#### 3.2. Evaluation of the NMME dataset

The analysis begins with an exploration of precipitation data using a time series plot. Time series plot can describe the characteristics and patterns of precipitation. Figure 2 shows that each ensemble member can follow a general precipitation pattern. If the trend of precipitation increases, the forecast results will also increase. Based on Figure 2, it can be seen that the forecast is close to the observed value, although there are events that overfitting and underfitting.

![Time series plot between ensemble prediction and precipitation observation](image)

*Figure 2. Time series plot between ensemble prediction and precipitation observation*
The next step is to evaluate the ensemble forecast. Table 2 shows that based on the RMSE, the COLA consisting of the CCSM3 and CCSM4 has a smaller RMSE compared to the CMC and GFDL. A smaller value of RMSE indicates that the forecast model is more accurate.

### Table 2. RMSE of monthly precipitation individual model

| Model      | RMSE  |
|------------|-------|
| CanCM3     | 17.17 |
| CanCM4     | 16.54 |
| CCSM3      | 14.86 |
| CCSM4      | 14.59 |
| GFDL-CM2p1 | 15.11 |

Figure 3 shows the performance of the forecast ensemble based on the percent bias value for each observation. The forecast bias results from the difference between the observations and the estimates. CCSM3 and CCSM4 have high frequency in January. This means that in certain observations, CCSM4 is the best model with the smallest RMSE. However, on other observations, CCSM4 has a large bias. CCSM3 and CCSM4 indicate good performance when predicting rainfall from November to April. Meanwhile, GFDL-CM2p1 outperformed other models from May to October.

![Figure 3](image)

In addition to using RMSE, the evaluation of the precipitation forecast model is carried out using the rank histogram shown in Figure 4. The rank histogram is used to determine whether the ensemble prediction is under dispersive or over dispersive. Based on Figure 4, the rank histogram has a shape similar to the letter U. Therefore, a calibration model is needed to produce more accurate and precise forecasts.
The initial stage in the BMA calibration process is to determine the training window (this study used 12 and 24 months). The training window is the amount of data used to estimate BMA parameters. To improve the forecast accuracy the ensemble needs a calibration using BMA. BMA gives weights for each model in January 2019 as shown in Table 3. The average weights for all periods show the same results. Table 1 shows that the CMC, especially CCSM4, has better performance than the COLA and NOAA’s GFDL. It is because the CCSM4 has the highest weight value.

Table 3. BMA weights for each model

| Training Window | CanCM3        | CanCM4        | CCSM3 | CCSM4 | GFDL-CM2p21 |
|-----------------|---------------|---------------|-------|-------|-------------|
| 12 months       | $1.251 \times 10^{-14}$ | $2.215 \times 10^{-13}$ | 0.035 | 0.965 | $6.254 \times 10^{11}$ |
| 24 months       | $3.84 \times 10^{-07}$ | $3.84 \times 10^{-06}$ | 0.058 | 0.518 | 0.424       |

The final step in the BMA calibration process is to get a calibrated PDF of the ensemble forecast. The probabilistic estimates are shown in Figure 5. The black vertical lines that represent the observations. It appears that the observations are within the predicted interval indicated by the dashed black line. The observation is indicated by an orange vertical line. The point forecast bias in the 12 months training window is smaller than 24. This can be seen from the difference between the yellow and black lines. The wider the distance between the yellow and black lines, the greater the forecast bias.
4. Conclusion
Based on the results of the ensemble forecast performance evaluation, it is known that COLA and NOAA perform better compared to CMC due to have the smaller RMSE. COLA model performed well in the period May to November. From December to April, the NOAA model performed well. Based on this research, BMA can improve the accuracy of the ensemble prediction.

Acknowledgments
The author would like to thank the Directorate of Research and Community Service (DRPM) of the Indonesian Ministry of Research, Technology, and Higher Education who has provided financial assistance for this research through a PDD research grant.

References
[1] Kirtman B P, Min D, Infanti J M, Kinter J L, Paolino D A, and Wood E F 2014 The North American Multi Model Ensemble (NMME); phase-1, seasonal-to- interannual prediction; phase-2, toward developing intraseasonal prediction Bulletin of American Meteorological Society 95 585–601
[2] Becker E, Van den Dool D, and Zhang Q 2014 Predictability and forecast skill in NMME Journal of Climate 27 5891–5906.
[3] Wang S, Zhang N, Wu L, and Wang Y 2016 Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method Renew Energy 94 629–636
[4] Nie H, and Guo Y 2019 An evaluation of East Asian summer monsoon forecast with the North American Multimodel Ensemble hindcast data International Journal of Climatology 39 4838–4852
[5] Ma F, Ye A, Deng X, Zhou Z, Liu X, Duan Q, Xu J, Miao C, Di Z, and Gong W 2015 Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China International Journal of Climatology 36 132–144.
[6] Faidah D Y, Kuswanto H, Suhartono, and Ferawati K 2019 Rainfall forecast with best and full members of the North American Multi-Model Ensemble Malaysian Journal of Science 38 113-119
[7] Setiawan A M, Koesmaryono Y, Faqih A, and Gunawan D 2017 North American Multi Model Ensemble (NMME) performance of monthly precipitation forecast over South Sulawesi, Indonesia IOP Conference Series: Earth and Environmental Science 58 012035
[8] Faidah D Y, Kuswanto H, and Suhartono 2019 The comparison of bayesian model averaging with gaussian and gamma components for probabilistic precipitation forecasting AIP Conference Proceedings 2192 090003
[9] Kuswanto H, Rahadiyuza D, and Gunawan D 2019 Probabilistic precipitation forecast in (Indonesia) using NMME models: case study on dry climate region Advances in Science, Technology, and Innovation 13–16
[10] Demargne J, Brown J D, Liu Y, Seo D J, Wu L, Toth Z, and Zhu Y 2010 Diagnostic verification of hydrometeorological and hydrologic ensembles Atmospheric Science Letters 11 114–122
[11] Seo D J, Herr H D, and Schaake J C 2006 A statistical post processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction Hydrology and Earth System Science Discuss 3 1987–2035
[12] Raftery AE, Gneiting T, Balabdoul F, and Polakowski M 2005 Using bayesian model averaging to calibrate forecast ensembles Monthly Weather Review 133 1155–1174
[13] Hamill T M 2001 Interpretation of rank histograms for verifying ensemble forecasts *Monthly Weather Review* **129** 550–560

[14] Sloughter J M, Gneiting T, and Raftery A E 2010 Probabilistic wind speed forecasting using ensembles and bayesian model averaging *Journal of the American Statistical Association* **105** 25–35

[15] Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou Y-T, Chuang H, Iredell M, Ek M, Meng J, Yang R, Mendez M P, van den Dool H, Zhang Q, Wang W, Chen M, and Becker E 2014 The NCEP climate forecast system version 2 *Journal of Climate* **27** 2185–2208

[16] Merryfield W J, Lee W-S, Boer G J, Kharin V V, Scinocca J F, Flato G M, Ajayamohan R S, Fyfe J C, Tang Y, and Polavarapu S 2013 The canadian seasonal to interannual prediction system. part I: Models and initialization *Monthly Weather Review* **141** 2910–2945

[17] Collins W D, Bitz C M, Blackmon M L, Bonan G B, Bretherton C S, Carton J A, Chang P, Doney S C, Hack, James J, Henderson T B, Kiehl J T, Large W G, McKenna, Daniel S, Santer B D, and Smith R D 2006 The community climate system model version 3 (CCSM3) *Journal of Climate* **19** 2122–2143

[18] G Delworth T L, Broccoli A J, Rosati A, Stouffer R J, Balaji V, Beesley J A, Cooke W F, Dixon K W, Dunne J, and Dunne K A 2006 GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics *Journal of Climate* **19** 643–674