North American Multi Model Ensemble (NMME) Performance of Monthly Precipitation Forecast over South Sulawesi, Indonesia

A M Setiawan 1, 2*, Y Koesmaryono 1, A Faqih 1 and D Gunawan 2

1 Department of Geophysics and Meteorology, Bogor Agricultural University (IPB), Bogor, Indonesia
2 Centres for Climate Change Information, Indonesia Agency for Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia

Email: amsari.setiawan@gmail.com

Abstract. The North American Multi-Model Ensemble (NMME) as one of the multi-model seasonal forecasting system, regularly generate monthly precipitation forecast for all globe with 0.5 – 11.5 months lead time. This useful information can be used as general input to regional and local precipitation forecast. This study quantifies monthly precipitation hindcast data performance in South Sulawesi provided by seven coupled models from the NMME during 29 years (1982 – 2010) period. Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dataset and the Standardized Verification System (SVS) for long-range forecasts (LRF) analysis applied to asses monthly precipitation prediction. Almost all individual model shows relatively high skill when used to make June – November monthly precipitation prediction and low skill when used to forecast monthly precipitation in December – May periods. Multi-Model Ensemble (MME) performed using Simple Composite Method (SCM) increased performance of monthly precipitation prediction. Nevertheless, this improvement only applies at short lead time (< 3.3 months). This result also shows that NMME is more promising when it used to make precipitation forecast application during dry period (i.e. drought prediction) rather than wet period over South Sulawesi region.

Keywords: NMME performance; Monthly Forecast Verification; South Sulawesi; SCM-MME; SVS-LRF

1. Introduction
The North American Multi-Model Ensemble (NMME) is one of the multi-model seasonal forecasting system, consisting of coupled models from US modeling centers, including the NOAA National Centers for Environmental Prediction (NOAA/NCEP), Center for Ocean-Land-Atmosphere Studies (COLA), NOAA’s Geophysical Fluid Dynamics Laboratory (NOAA/GFDL), National Aeronautics and Space Administration/ Global Modeling and Assimilation Office (NASA/GMAO) and Canadian modeling centers [1]. The routine generation of global monthly precipitation forecasts of each individual model are freely available from and easily downloaded from International Research Institute for Climate and Society/ Lamont-Doherty Earth Observatory (IRI/LDEO) website.
This dataset has been widely used in previous studies to verify monthly mean precipitation [2] and 2-meter temperature forecasts[3], monthly prediction of Sea Surface temperature as ENSO indices [4,5], seasonal precipitation [6,7], seasonal global drought onset [8] and interannual prediction [1]. Information about forecasting skill for each model needed before it can be used to improve operational routine monthly deterministic forecast provided by National Meteorological and Hydrological Services (NMHS).

Precipitation over South Sulawesi as one of the primary national rice production centers strongly affected by El Niño event [4,9,10], indicated by wider drought coverage area [11]. Precipitation information become very important to this region. The objectives of this study are to quantifies monthly precipitation hindcast data performance in South Sulawesi provided by seven individual coupled models from the North American Multi-Model Ensemble (NMME) and evaluate the forecasting skill of Simple Composite Method Multi Model Ensemble (SCM-MME) formed from NMME. This study investigates model performance for all time period (1982 – 2010) and stratified according to each calendar month and lead time.

2. Data and Methodology

2.1. NMME
The gridded monthly precipitation predictions of each individual model were downloaded from the International Research Institute for Climate and Society/ Lamont-Doherty Earth Observatory (IRI/LDEO) collection of climate data (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). All time coverage with all lead and ensemble mean [12] of each NMME member used and averaged over South Sulawesi. Table 1 provides a brief description of the models used to in this study.

| Model       | Lead Time (Months) | Time Coverage | Ensemble Members | Modeling Centre                                      | References |
|-------------|--------------------|---------------|------------------|------------------------------------------------------|------------|
| CFSv2       | 0.5 – 9.5          | 1982 – 2010   | 24               | National Centers for Environmental Prediction (NOAA/NCEP) | [13]       |
| CMC1-CanCM3 | 0.5 – 11.5         | 1981 – 2010   | 10               | Environment Canada's Canadian Meteorological Centre (CMC) | [14]       |
| CMC2-CanCM4 | 0.5 – 11.5         | 1981 – 2010   | 10               | Environment Canada's Canadian Meteorological Centre (CMC) | [14]       |
| CCSM3       | 0.5 – 11.5         | 1982 – Present| 6                | The Center for Ocean-Land-Atmosphere Studies (COLA)   | [15]       |
| CCSM4       | 0.5 – 11.5         | 1982 – Present| 6                | The Center for Ocean-Land-Atmosphere Studies (COLA)   | [16]       |
| GFDL-CM2p5-FLOR-B01 | 0.5 – 11.5 | 1982 – 2012   | 10               | NOAA’s Geophysical Fluid Dynamics Laboratory (GFDL) National Aeronautics and Space Administration/ The Global Modeling and Assimilation Office (NASA/ GMAO) | [17] [18] |
| GEOS5 GMAO-062012 | 0.5 – 8.5 | 1981 – Present| 12               |                                                       |            |

An MME of seven individual coupled models was formed using the simple composite method [19], with equal weights assigned to ensemble mean predictions of each model, assuming that each model is independent. The mean bias from each model was removed by forming anomalies with respect to each model’s own seasonal climatology. Because the lead times and hindcast periods were different in each model, a common range of 0.5–8.5 months lead time and the 1982–2010 hindcast period were selected for the MME.
2.2. Observation
This study quantifies NMME monthly precipitation hindcast performance using 0.05° x 0.05° blended
gauge-satellite of daily and monthly precipitation estimates of the Climate Hazards Group InfraRed
Precipitation with Stations (CHIRPS) dataset [20]. This grided dataset were obtained from Climate
Hazards Group/ The Department of Geography, University of California Santa Barbara
(ftp://ftp.chg.ucsb.edu/pub-org/chg/products/CHIRPS-2.0/) for the 30-year period from 1981 – 2010.
South Sulawesi CHIRPS dataset averaged over the study area (119.25E – 121.75E and 7S – 2S)
(figure 1). General topographical characteristics of study area described with gridded elevation data
were obtained from National Geophysical Data Center, NOAA, with 2 – minute Gridded Global Relief
Data (ETOPO2)v2 [21].

![Monthly Precipitation over South Sulawesi](image)

**Figure 1.** Scatter plot between all rain-gauged monthly precipitation observation and CHIRPS
averaged over study area (left); and elevation characteristic (right) over study area (blue box) (meter
above sea level/ m. a. s. l) with station location (black circle).

Observational-based stations are distribute over 4 region and district in South Sulawesi (Table 2)
with various altitude between 14 – 81 meter above sea level (m. a. s. l). Various altitude and station
type such as rain gauge (Obs Gauge), Agricultural Meteorological station (AgriMet), and BMKG
weather station (BMKG). This observational dataset provided by The Indonesia Agency for
Meteorology Climatology and Geophysics (BMKG). Station form four different districts located at
West Coast (WC), South Coast (SC), East Coast (EC) and North Coast (NC) South Sulawesi Province
are used to determine their representation in CHIRPS dataset. Good correlation between CHIRPS
values and recorded insitu precipitation over study area also shown (Figure 1).
Table 2. Observational station used for comparison with CHIRPS monthly precipitation dataset. Missing data information calculated from 30-years period from January 1981 to December 2010.

| No | StationName | District       | Latitude (dms) | Longitude (dms) | Altitude (m.a.s.l) | Station Type | Missing Data (%) |
|----|-------------|----------------|----------------|-----------------|-------------------|--------------|------------------|
| 1  | STAMET HASANUDDIN | MAROS (WC)      | 05 04 15.6 S  | 119 33 07.7 E   | 14                | BMKG         | 3.61             |
| 2  | BATUKAROPA   | BULUKUMBA (SC)  | 05 28 09.1 S  | 120 12 29.2 E   | 81                | AgriMet      | 0.28             |
| 3  | SIWA         | Wajo (EC)       | 03 44 59.1 S  | 120 21 50.2 E   | 25                | Obs Gauge    | 8.61             |
| 4  | STAMET MASAMBA | Luwu Utara (NC) | 02 33 16.0 S  | 120 19 27.0 E   | 50                | BMKG         | 11.39            |

2.3. Verification

The Standardized Verification System (SVS) for long-range forecasts (LRF) analysis [22] was used on the hindcast data and monthly precipitation observation from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dataset [20] to evaluate model skills for predicting South Sulawesi precipitation. The SVS consists of four parts including diagnostic measures, key parameters, verification data sets, and details of forecast systems [22]. The monthly precipitation anomaly verification was stratified according to each calendar month and lead time. The Mean square skill score (MSSS) is defined as one minus the ratio of the mean square error (MSE) of the forecasts to the MSE of climatology forecast. The values can be used to compare forecast performance to climatology forecast. MSSS decomposition provides information on phase errors through forecast/observation correlations, amplitude errors through the ratio of the forecast to observed variances, and overall bias error. Ratio of the square roots of the variance explains how well the variance of model can explain the variance of observation. The information can be used to adjust or weight forecasts when they are used as an input into regional and local forecasts. If model performance is good with a strong correlation and small amplitude error, the model can be considered suitable for use in forecasts. The MSSS will not exceed zero if forecasts have the same amplitude with the observations and no overall bias unless the correlation exceeds approximately 0.5 [22]. A good forecast is achieved when the ratio of the forecasted to observed variances is close to one, overall bias (the difference between the forecast and observation) is near zero, and the MSSS is close to one. Negative/zero MSSS values indicate that deterministic forecasts are worse than/the same as climatology forecasts.

3. Result and Discussion

3.1. Monthly Performance of Individual Model

The monthly climatological mean precipitation calculated for each station with 30-years period (1981-2010). CHIRPS monthly precipitation data averaged over study area and monthly climatological mean calculated with same period with station data. In general, lower precipitation amount in study area occurred in August – October. Higher precipitation with high variation value between each region (especially between west and east coast) observed during December – June period (figure2).
Figure 2. Monthly South Sulawesi climatological mean (1981-2010) of precipitation based on regional averaged CHIRPS data (blue bar) and all gauged-based data (green bar). Line chart represent precipitation in west coast (WC_prec/ circle dashed line), south coast (SC_prec/ cross line), east coast (EC_prec/ triangle dashed line) and north coast (NC_prec/ square line).

Monthly climatological mean pattern obtained from station are similar and slightly underestimates compared to gauged based observation. Furthermore, relatively highly significant relationship with monthly precipitation from observation ($R^2 = 0.63$, $\alpha = 0.05$) makes CHIRPS data is considered to represent the rainfall observation in the study area. This precipitation dataset has been used to evaluate NMME skill [23] and verified in various past studies with good correlation and small error between CHIRPS values and recorded insitu precipitation [24–26].

The monthly precipitation verification was stratified according to each calendar month and lead time. A lead time of 0.5 month implies a prediction made at the very beginning of the target month. Comparisons of model performance based on anomaly correlation values are presented in figure 3.

The correlation coefficient values are plotted for each month and for each individual model lead time. Noticeable skill differences based on anomaly correlations are founded in monthly precipitation forecasts for June - November at medium lead times (4.5 – 9.5). The NASA/GMAO has relatively high skill (anomaly correlations > 0.6) while other individual models show lower skill (0.3 – 0.8). These correlations are statistically significant at 95 % confidence level. Almost all individual model shows low skill (anomaly correlations < 0.4) when used to forecast monthly precipitation in January – May periods. MME performed using simple composite method increased performance of monthly precipitation prediction. Nevertheless, this improvement only applies at short lead time (< 3 months) and not seen in January – May prediction.

Model performance comparison based on overall bias values between each model output and observed precipitation are presented in figure 4. Similar with figure 3, overall bias values are plotted for each month and for each individual model lead time. Almost all model show overestimates prediction during dry period (June – August) and underestimates during wet period (January – March), except for CFSv2 which show overestimates prediction for all month. Significant MME improvement related to smaller overall bias value applies at three month forecast and shorter lead time.
**Figure 3.** Performance of monthly precipitation individual model and Multi Model Ensembles (MME) forecast (1982 – 2010) based on anomaly correlation coefficient value of monthly precipitation over South Sulawesi with CFSv2 (a), COLA-CCSM3 (b), COLA-CCSM4 (c), GFDL-CM2p5-FLOR-B01 (d), CMC1-CanCM3 (e), CMC2-CanCM4 (f), NASA-GMAO-062012 (g) and Simple Composite Method (SCM)-MME (h). Target month indicated by horizontal axis and lead time on the vertical axis.
Figure 4. Similar with figure 2, except for overall bias between each model and observation. Positif (negatif) overall bias values indicated by color ramp, associated with overestimates (underestimates) of model prediction.
Figure 5. Same as figure 3, except for Mean Square Skill Score (MSSS) of each model.
Figure 5 describes the Mean Square Skill Score (MSSS) of each individual model as a function of target month and lead time. The most noticeable skill difference based on MSSS is found in forecasts for June - November at 2.5 to 5.5 month lead times. The NASA-GMAO has relatively high skill (MSSS of 0.4 or higher) while other individual models show lower skill. This result is slightly different from previous results that reviewed the model performance only based on overall bias. The MSSS not only is based on correlations, but also takes into account the phase errors, amplitude errors and overall bias error of the forecasts. Worse forecast skills than climatology were found in forecast for January to March (CFSv2 at all lead time), January to May (COLA-CCSM3 at all lead time), January to April (COLA-CCSM4 at 1.5 months lead time or longer), December to April (GFDL-CM2p5-FLOR-B01 at all lead time), January to May (CMC1-CanCM3 at 1.5 months lead time or longer ) and January to February (CMC2-CanCM4 at all lead time). Simple Composite Method (SCM)-MME make significant performance improvement especially during June – November target month with less than 4.5 month lead time.

Climatological monthly precipitation variability between each region during December – May are suspected to be reason of lower skill prediction in study area. There are opposite monthly rainfall pattern in this period especially between east coast and west coast region (figure 2), although general rainfall condition well captured by NMME prediction during June – November. Therefore, rainfall prediction for December – May should not generalized and further downscaling procedure may apply to capture regional precipitation characteristic.

3.2. Multi Years Performance of NMME
Further analysis was conducted to determine the performance comparisons between each individual model and multi-models ensemble forecast for monthly precipitation forecast during multi years (1982 – 2010) period. This comparison presented in figure 6.

The MME, COLA-CCSM4 and NASA/GMAO have relatively high skill (correlations of 0.6 or higher) while GFDL-CM2p5-FLOR-B01 models show lowest skill (correlation < 0.2) if model performance only stratified by coefficient value. A CFSv2 and CMC2-CanCM4 show underestimates prediction at almost all lead time, while GFDL-CM2p5-FLOR-B01, CMC1-CanCM3 and NASA/GMAO show overestimates prediction. These overall bias issues for each individual model are well addressed in MME and improvements showed by near zero overall bias for all lead time.

COLA-CCSM4 shows good performance according to model sensitivity explained by near one ratio of the square roots of the variance. Based on the MSSS, the most noticeable skill difference was found in 4.5 to 8.5 month lead times forecast. The GFDL has relatively low skill (MSSS of 0.1 or lower) while other models show higher skill (0.2 – 0.5) due to overall bias and the ratio of square root of the variance value.
Figure 6. Performance comparison of multi years monthly precipitation from Multi Model Ensembles (MME) forecast (1982 – 2010) over South Sulawesi based on Standardized Verification System (SVS) for Long-Range Forecasts (LRF) analysis. Each figure indicates anomaly correlation value (a), overall bias (b) and ratio of square roots of the variances (c) as decomposition of Mean Squared Skill Score (MSSS) (d). The individual model name is indicated on the horizontal axis, and lead time prediction on the vertical axis.

4. Conclusion

Noticeable skill differences in monthly precipitation prediction were found between each individual model, especially at medium to long lead times. The performance of monthly precipitation MME forecasts for all time period (1982–2010) using the Simple Composite Method (SCM)-MME method show model improvement. Significant performance improvement by SCM-MME especially during June – November target month with less than 4.5 month lead time. Climatological monthly precipitation variability between each region during December – May are suspected to be reason of lower skill prediction in study area. Therefore, rainfall prediction for December – May should not generalized and further downscaling procedure may apply to capture regional precipitation characteristic. This result also shows that NMME is more promising when it used to make precipitation forecast during dry period rather than wet period over South Sulawesi region.
Acknowledgements
The first author thanks the Indonesia Agency for Meteorology Climatology and Geophysics (BMKG) for providing a scholarship for his PhD program. This research is funded by Department of Geophysics and Meteorology, Bogor Agricultural University (IPB) as part of BMKG – IPB scholarship program. The NMME project and data dissemination was supported by NOAA, NSF, NASA and DOE.

References
[1] Kirtman B P, Min D, Infanti J M, Kinter J L, Paolino D A, Zhang Q, van den Dool H, Saha S, Mendez M P, Becker E, Peng P, Tripp P, Huang J, DeWitt D G, Tippett M K, Barnston A G, Li S, Rosati A, Schubert S D, Rienecker M, Suarez M, Li Z E, Marshak J, Lim Y-K, Tribbia J, Pegion K, Merryfield W J, Denis B and Wood E F 2014 Bull. Am. Meteorol. Soc. 95 585–601
[2] Wang H 2014 Hydrol. Process. 28 4472–86
[3] Becker E, van den Dool H, Zhang Q and Peña M 2013 Sci. Technol. Infus. Clim. Bull.
[4] Setiawan A M 2014 Drought Characteristics in Indonesia Related to Warm ENSO Episodes (Busan)
[5] Barnston A G, Tippett M K, Van Den Dool H M and Unger D A 2015 J. Appl. Meteorol. Climatol. 54 1579–95
[6] Infanti J M and Kirtman B P 2014 J. Hydrometeorol. 15 529–50
[7] Ma F, Ye A, Deng X, Zhou Z, Liu X and Duan Q 2016 Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China 144 132–44
[8] Yuan X and Wood E F 2013 Multimodel seasonal forecasting of global drought onset 40 4900–5
[9] Setiawan A M 2011 Determination of Reference ENSO Index for Indonesian Region Based on Correlation Analysis of Spatial and Temporal Pattern with Standardized Precipitation Index (SPI) (Bandung: Institut Teknologi Bandung (ITB))
[10] Hidayat R, Ando K, Masumoto Y and Luo J J 2016 IOP Conf. Ser. Earth Environ. Sci. 31 12043
[11] Surmaini E, Hadi T W, Subagyro N and Puspito N T 2015 Theor. Appl. Climatol. 121 669–84
[12] Ishizaky Y, Nakagawa T and Takayabu I 2012 Clim. Dyn. 39 185–206
[13] 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 J. Clim. 27 2185–208
[14] Merryfield W J, Lee W-S, Boer G J, Kharin V V., Scinocca J F, Fyfe J C, Tang Y and Polavarapu S 2013 Mon. Weather Rev. 141 2910–45
[15] Collins W D, Bitz C M, Blackmon M L, Donohue E A, Marshak J, Lim Y-K, Tribbia J, Pegion K, Merryfield W J, Denis B and Wood E F 2014 Bull. Am. Meteorol. Soc. 95 585–601
[16] Gent P R, Danabasoglu G, Donner L J, Holland M M, Hunke E C, Jayne S R, Lawrence D M, Neale R B, Rasch P J, Vertenstein M, Worley P H, Yang Z L and Zhang M 2011 J. Clim. 24 4973–91
[17] 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 J. Clim. 19
[18] Vernieres G, Rienecker M M, Kovach R and Keppenne C L 2012 The GEOS-iODAS: Description and Evaluation vol 30(Greenbelt, Maryland: National Aeronautics and Space Administration)
[19] Jeong H-I, Lee D Y, Ashok K, Ahn J-B, Lee J-Y, Luo J-J, Schemm J-K E, Hendon H H, Braganza K and Ham Y-G 2012 Clim. Dyn. 39 475–93
[20] Funk C C, Peterson P J, Landsfeld M F, Pedreros D H, Verdin J P, Rowland J D, Romero B E,
Husak G J, Michaelsen J C and Verdin a. P 2014 *U.S. Geol. Surv. Data Ser.* **832** 4

[21] National Geophysical Data Center 2006 ETOPO2v2 2-minute Global Relief Model

[22] World Meteorological Organization 2010 Standardized Verification System (SVS) for Long-Range Forecasts (LRF) *Manual on the Global Data Processing and Forecasting System* WMO-No. 485 vol I-Global (Geneva: World Meteorological Organization) p II.8.1-II.8.17

[23] Wanders N, Wood E F, Lott F C, Gordon M, Graham R J, Macleod D A, Jones A, Di F, Shukla S, Funk C and Hoell A 2014 *Environ. Res. Lett.* **9**

[24] Duan Z, Liu J, Tuo Y, Chiogna G and Disse M 2016 *Sci. Total Environ.*

[25] Katsanos D, Retalis A and Michaelides S 2015 *Atmos. Res.* **Volume 169** 459–464

[26] Hessels T M 2015 *Comparison and Validation of Several Open Access Remotely Sensed Rainfall Products for the Nile Basin* (Delft University of Technology)