Performance of probabilistic forecast of the onset of the rainy season over Java Island based on the application of Constructed Analogue (CA) method on Climate Forecast System Version 2 (CFSV2) model output

F Lubis¹ and I J A Saragih¹,²*

¹ BMKG - Kualanamu Meteorological Station, Deli Serdang
² Department of Physics, Faculty of Mathematics and Natural Sciences – Universitas Sumatera Utara, Medan

* immanuel.saragih@bmkg.go.id

Abstract. The onset of the rainy season is one of the forecast products that is issued regularly by the Indonesian Agency of Meteorology, Climatology, and Geophysics (BMKG), with deterministic information about the month of which the initial 10-days (dasarian) of the rainy season will occur in each a designated area. On the other hand, state-of-the-art of seasonal forecasting methods suggests that probabilistic forecast products are potentially better for decision making. The probabilistic forecast is also more suitable for Indonesia because of the large rainfall variability that adds up to uncertainty in climate model simulations, besides complex geographical factors. The research aims to determine the onset of the rainy season and monsoon over Java Island based on rainfall prediction by Constructed Analogue statistical downscaling of CFSv2 (Climate Forecast System version 2) model output. This research attempted to develop a method to produce a probabilistic forecast of the onset of the rainy season, as well as monsoon onset, by utilizing the freely available seasonal model output of CFSv2 operated by the US National Oceanic and Atmospheric Administration (NOAA). In this case, the output of the global model is dynamically downscaled using the modified Constructed Analogue (CA) method with an observational rainfall database from 26 BMKG stations and TRMM 3B43 gridded dataset. This method was then applied to perform hindcast using CFS-R (re-forecast) for the 2011-2014 period. The results show that downscaled CFS predictions with initial data in September (lead-1) give sufficient accuracy, while that initialized in August (lead-2) have large errors for both onsets of the rainy season and monsoon. Further analysis of forecast skill using the Brier score indicates that the CA scheme used in this study showed good performance in predicting the onset of the rainy season with a skill score in the range of 0.2. The probabilistic skill scores indicate that the prediction for East Java is better than the West and Central-Java regions. It is also found that the results of CA downscaling can capture year-to-year variations, including delays in the onset of the rainy season.

Keywords: probabilistic forecasts, the onset of the rainy season, statistical downscaling, constructed analogue, monsoon onset
1. Introduction
Probabilistic forecasts have an opportunity value for an event or not while deterministic forecasts have no opportunity value. Probabilistic forecasts are more useful for decision-makers than forecasts that state explicitly that circumstances will occur because the user can make a decision based on probability, and special knowledge about the cost of decision-making, gains, or losses that depend on the weather to occur [1].

Probabilistic forecasts have better potential value for decision-makers rather than deterministic forecasts because of their ability to measure uncertainty [2]. Forecasts of the onset of the rainy season in the Indonesian territory have great uncertainty due to the high rainfall variability in Indonesia. Seasonal until interannual variability of Indonesian rainfall is affected primarily by monsoon [3] and El Nino-Southern Oscillation (ENSO) [4]. The territory of Indonesia is influenced by two monsoon systems, the Asian monsoon and Australian monsoon, the transitional region between the Asian summer monsoon and the Australian summer monsoon [5,6]. The monsoon is a global circulation that undergoes the initial phase (onset), active phase, break phase, and final phase (withdrawal). One of the signs of an active monsoon is the onset of the rainy season [7], so the monsoon circulation is considered active when indicated by the occurrence of monsoon onset which is followed by several rainy days (onset of rainy season). The onset of the rainy season in Indonesia is related to the monsoon onset where the onset of the rainy season is a monsoon onset in Indonesia [8]. The onset of the rainy season is a product that is used operationally by the Indonesian Agency of Meteorology, Climatology, and Geophysics (BMKG).

Forecasts for the next three to six months with Global Climate Models (GCM) are widely developed today. However, GCM output has a relatively coarse resolution (above 50 km) and is not able to represent processes smaller than 50 km. To obtain regional and local scale information from GCM output, a technique called downscaling [9] is required. The approach used is to use statistical methods to derive an empirical relationship between GCM output and local historical observation data known as statistical downscaling. Statistical downscaling has advantages in terms of resource requirements and a much smaller calculation time than dynamical downscaling necessaries [10]. Until now, there are various statistical downscaling techniques developed and applied in various places in the world and the CA (Constructed Analogue) method has several advantages compared to other statistical downscaling techniques. In terms of computing, CA is simpler than other statistical downscaling techniques, with performance not much different [9].

In this study, we will use Climate Forecast System version 2 of the National Center for Environmental Prediction (NCEP CFSv2). NCEP CFSv2 is a marine-atmospheric coupling forecast system that provides season prediction operational with re-forecasts data to evaluate and calibrate the model. The study of Zhang et al. [11] shows many features of the Maritime Continent rainfall is simulated CFSv2 well. Syahputra [12], who implements the Constructed Analogue (CA) method, has successfully improved the monthly rainfall prediction performance of the CFS (Climate Forecast System) model in Java and southern Sumatra. This study focuses on the implementation of the Constructed Analogue (CA) method for probabilistic forecasts of the onset of the rainy season and monsoon onset in Java Island. For verification of Constructed Analogue (CA) of onset of rainy season forecasts, data from 26 observation stations of BMKG in Java are recorded, while CA result of monsoon onset forecast verified with TRMM 3B43 data. The verification method used is the Brier score. The verification results show the prediction skill of the onset of the rainy season and monsoon onset have good performances with the Brier score value in the range of 0.2.

2. Data and Methods
The data used in this research include model output data and observation data. The model output data is daily CFS-R (Forecast Climate-Reforecast) wind data including meridional wind (u-wind) and zonal wind (v-wind) on layer 850hPa of the period 1998-2010 with a hindcast of 9 months from every 5 days and running four cycles from that day (00, 06, 12 and 18 UTC), starting from 1 January every year. Besides, this research also used the CFSv2 operational prediction data (October-November-December) in the 2011-2014 periods. The output models are used in 2 lead times, September and August, and produce 30 ensemble members after downscaling with CA. In this study, ensemble members are
generated from 2 wind field variables namely $\psi$ (current function, PSI) and $\chi$ (potential velocity, PHI) on layer 850 hPa. Observation data used are daily rainfall data of 26 observation stations of BMKG in Java which of 1998-2010 and 2011-2014 (October-November-December) periods and daily rainfall data TRMM 3B43 which of 1998-2010 and 2011-2014 (October-November-December) periods.

Constructed Analogue (CA) applied in this study largely follows the methods proposed by Hidalgo et al. [10] with modifications. The CA method is generally divided into two processes: the diagnosis process and the prognosis process. This research is divided into several stages:

1) Diagnosis process
   The diagnosis process is the predictor pattern adjustment in the target time towards the predictor pattern in the database. The analogue search process is performed using the Cosine Similarity method [13]. Cosine Similarity is used to calculate the degree of similarity between vectors $\vec{a}(u)$ and $\vec{a}(t)$ by the equation:
   $$S(u) = \frac{\vec{a}(u) \cdot \vec{a}(t)}{|\vec{a}(u)||\vec{a}(t)|}$$  \hspace{1cm} (1)

   In this study, some modifications of the CA method, in general, are made. This modification is done in the diagnosis process and application of the multi-window scheme in the CA method [12]. The use of the same dataset for model targets and model databases (CFS-R) is done to avoid biases between the database and the targets which will be compared. Meanwhile to overcome the sensitivity of the CA method to the chosen domain, in this study, 5 (five) windows are considered to affect the distribution of rainfall in Indonesia [14].

2) Prognosis process
   The prognosis process aims to establish an analogue of rainfall at each target $t$ time, based on the best analogue (subset) group obtained from the diagnosis process. In this study, the weighted average method is used in the prognosis process [15]. First, the weight ($W$) of each member of the analogue subset of predictors is determined by the correlation coefficient and RMSE toward target predictors. Furthermore, this weight is used in the formation of constructed analogue predictand in target time $t$:
   $$Z_{CA}(t) = \sum_{i=1}^{N} W_i P_i$$  \hspace{1cm} (2)

   where $P$ is a predictand analogue subset which is coupled with a predictor analogue subset and $Z$ is the constructed analogue of predictand.

3) Determining the occurrence of onset of the rainy season and monsoon onset
   The onset of the rainy season is defined based on the operational framework used by BMKG is the onset of the rainy season is the amount of rainfall in one dasarian (10 days) equal to or more than 50 millimetres and followed by the next several dasarian. While the monsoon onset is defined based on the Australian Monsoon Rainfall Index (AMRI) modification criterion [16] which is running mean five-day rainfall that exceeds 150% of the mean annual cycle of daily rainfall and difference 150% mean annual cycle with the mean annual cycle of daily rainfall is greater or equal with 1 mm.
   $$AMRI = \begin{cases} 
1, & 5_{rm \text{t} \text{prec}} > 150 \% \text{tprec}_{long} \\
0, & 5_{rm \text{t} \text{prec}} \leq 150 \% \text{tprec}_{long} 
\end{cases}$$  \hspace{1cm} (3)

   with $\frac{5_{rm \text{t} \text{prec}}}{\text{tprec}_{long}}$ = running-mean 5-daily rainfall (mm) and $\text{tprec}_{long}$ = long-term daily average rainfall (mm)
where P is a predictand analogue subset which is coupled with a predictor analogue subset and Z is the constructed analogue of predictand.

4) Verify the forecasts of onset of the rainy season and monsoon onset

Prediction skills of the onset of the rainy season and monsoon onset are calculated by Brier score [17] which is a mean-square error measure of forecast probability for dichotomy events (two categories), such as rain/no rain events.

\[
BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - a_i)^2, \quad (0 \leq BS \leq 1) \tag{4}
\]

To obtain the relationship of the Brier score and the probability of a deeper forecast, Murphy [18] decomposes the Brier Score into three parts: reliability, resolution, and uncertainty.

\[
BS = \text{Reliability} - \text{Resolution} + \text{Uncertainty} \tag{5}
\]

\[
\text{Reliability} = \sum_{i=1}^{L} \left( p_i - \frac{M_i}{N_i} \right)^2 \frac{N_i}{N} \tag{6}
\]

\[
\text{Resolution} = \sum_{i=1}^{L} \left( \frac{M}{N} - \frac{M_i}{N_i} \right)^2 \frac{N_i}{N} \tag{7}
\]

\[
\text{Uncertainty} = \frac{M}{N} \left( 1 - \frac{M}{N} \right) \tag{8}
\]

with

- \( N \) = number of samples
- \( p_i \) = prediction probability
- \( a_i \) = observation (value 1 for observed events and 0 for un-observed)
- \( \frac{M_i}{N_i} \) = relative frequency of observation
- \( M \) = frequency of prediction

3. Results

3.1. Lead time influence on CA prediction results

Box plot of ensembles predictions (Figure 1) shows the prediction’s spread of CFS-R output in September (lead-1) is better than CFS-R output in August (lead-2). For CFS-R lead-1, some observed precipitations are within the predictive rainfall range of the ensembles, but not always around the median. In some instances, the observed rainfall lies outside the predictive range of the ensembles.

The results of CA simulations of 30 members are then averaged into mean-ensemble members for daily and dasarian rainfall. The mean-ensemble members of the CA are correlated with the observed rainfall from 26 BMKG stations (Figure 2). The correlation distribution for dasarian lead-1 rainfall shows a good simulation performance in Java Island, especially in East Java with a correlation value between 0.4 - 0.72. While the distribution of dasarian rainfall in lead-2 shows generally lower performance than lead-1. Likewise, the daily lead-1 rainfall shows better performance than the daily lead-2 rainfall (Figure 2 and Figure 3).
Figure 1. Distribution of 30 ensemble members of dasarian rainfall predicted for August / lead-2 (a and c) output and September / lead-1 (b and d) outputs. The round plot (blue) is the observational rainfall per dasarian of 2011-2014.

Figure 2. The correlation coefficient between observed daily rainfall and predicted lead-1 (a) and lead-2 (b).
3.2. Prediction skill of the onset of the rainy season

The area of Java is divided into 3 clusters based on the spatial correlation value of dasarian lead-1 rainfall (Figure 4). The region is divided into East Java (Karangkates, Kalianget, Karangploso, Perak I, Perak II, Juanda, Tretes), West Java (Tangerang, Cengkareng, Kemayoran, Serang, T. Priok, Citeko, P.Betung) and Central Java (Jatiwangi and Tegal).

In each cluster, the probability of AMH (onset of rainy season) is calculated for the October-November-December period 2011-2014 (Figure 5) based on the BMKG definition. The CA results show that AMH’s probability in East Java generally occurs on Nov-III (42%) and Dec-I (reaches 66%). The probability of AMH in West Java generally occurs on Oct I (reaches 50%) and Oct-III (reaches 26%). While in Central Java, the probability of AMH generally occurs on Nov-III (reaches 67%) and Dec-I (reaches 77%).
The probability of AMH (onset of rainy season) result of predicted CA (Figure 5) is grouped into bins based on its probability value (%). Then the Brier score of AMH is calculated based on equations (4) and (5).

**Figure 5.** The probability of AMH (onset of rainy season) result of predicted CA in (a) East Java (b) West Java and (c) Central Java

**Figure 6.** Brier score of AMH in Java Island
In general, the Brier scores in all three areas are quite good. Figure 6 shows that East Java has the best brier score followed by Central Java and then West Java. This shows that AMH’s prediction skill in East Java is better than Central Java and West Java. This is consistent with the previous deterministic forecasts showing East Java has the best skill in AMH prediction.

3.3. Prediction skill of monsoon onset
Rainfall in Java is analyzed by Principal Component Analysis (PCA). The calculated data in PCA is the rainfall of running mean of 5-day (penta-day) in October-December in 1999-2015. Then the cluster analysis (clustering) grid is done based on the equation between two objects with the shortest distance called Euclidean Distance. The cluster analysis yields 4 clusters (Figure 7).

The monsoon onset is calculated based on TRMM 3B43 observation data and is calculated in all ensemble members in each grid on Java by using a modified AMRI threshold [16]. Monsoon onset dates are calculated annually on each grid in 30 ensemble members of CA. To find out how the performance of rainfall of running means of 5-day of predictions of CA, then a sample of rainfall of running mean of 5 days in each cluster is correlated with TRMM observation rainfall.

Table 1. The correlation of rainfall of running-mean of 5 days in each cluster

| Cluster  | Correlation |
|----------|-------------|
|          | Lead-1      | Lead-2      |
| Cluster-1| 0.408       | 0.405       |
| Cluster-2| 0.327       | 0.22        |
| Cluster-3| 0.317       | 0.33        |
| Cluster-4| 0.419       | 0.414       |

To find out how the date performance of monsoon onset is calculated the percentage of monsoon onset hit and the probability of the date of occurrence of monsoon onset in each grid on a map of contour as in Figure 8. The percentage of contours that hit the onset of monsoon is generally quite large in West Java and East Java. The percentage of contours that hit the onset of monsoon reaches over 70% in West Java in 2011 and 2013 and reaches over 60% in East Java in 2013. For probability, the date of occurrence of onset can be seen in Figure 9, Figure 10, and Figure 11.
Figure 8. Percentage of contours hit of monsoon onset in Java Island in 2011-2014

Figure 9. Probability of monsoon onset of CA prediction on October 6-31 (Dasarian-I Oct, Dasarian-II Oct, and Dasarian-III Oct) in 2011-2014
Figure 10. Probability of monsoon onset of CA prediction on 1-30 November (Dasarian-I, Dasarian-II Nov, Dasarian-III Nov) in 2011-2014

Figure 9, Figure 10, and Figure 11 show that the probability of monsoon onset on October 6-31 (Dasarian-I Oct, Dasarian-II Oct, Dasarian-III Oct) is greater than any other date. The probability of monsoon onset CA prediction is then grouped into a bin based on its probability values (%). Then calculated the Brier score of monsoon onset as the onset of the rainy season. In general, the Brier score in each cluster is quite good. Figure 12 shows that cluster 2 (Central Java region) and cluster 4 (East Java region) have the best brier score. This indicates that the prediction skill of dates of monsoon onset in Central and East Java are the best. This is following the previous prediction results of probabilistic rainfall that show the areas of East Java and Central Java have the best prediction skills.
Discussions

CA results in the prediction of onset of rainy season show different performance in all three areas of Java. East Java region shows the best performance with a correlation value of dasarian lead-1 rainfall reaches 0.7, followed by Central Java and West Java. This is in contrast to previous studies showing similar performance across Java with the correlation value of monthly lead-1 rainfall in the range 0.61 - 0.8 [12]. The monsoon onset in this study used a modified AMRI threshold based on Azizah [16]. CA results in the prediction of monsoon onset indicate the date of onset in Java on average occurred in October. The spread of the date of October's onset (Dasarian-I Oct, Dasarian-II Oct, and Dasarian-III Oct) propagates from west to east from 2011 to 2014. This is consistent with the determination of
monsoon onset by Tanaka (1994) that defines the onset of monsoons based on cloud cover and wind patterns. Probabilistic prediction skill of onset of the rainy season and monsoon onset is quite good with Brier score 0.209–0.262. However, the probabilistic predictions of the onset of the rainy season are more consistent with the deterministic results rather than the probabilistic monsoon onset. These are indicated by the correlation values of dasarian lead-1 rainfall and Brier score in East Java which shows the best performance, followed by Central Java and West Java in the predictions of onset of the rainy season.

5. Conclusions
The prediction skill of daily or dasarian rainfall of the CA results of the August (lead-2) output can not be used in this study because the skills are too low qualitatively and quantitatively. While the skill of predicted daily and dasarian rainfall output of September (lead-1) can be used because it has better skill than the output of August (lead-2) with a correlation of 0.7. In general, the CA schemes used in this study show considerable skill in the probabilistic forecasts of onset of the rainy season and monsoon onset with Brier score values in the range of 0.2. In the prediction of the onset of the rainy season, East Java has the best skill followed by Central Java and West Java. In the prediction of monsoon onset, the areas of East Java, Northern Central Java, and some areas of West Java have the best skills.

References
[1] Gordon N and Shaykewich J 2000 Guidelines on Performance Assessment of Public Weather Services (Geneva, Switzerland: World Meteorological Organization)
[2] Murphy A H 1977 The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation Mon. Weather Rev. 105 803–16
[3] Ramage C S 1971 Monsoon Meteorology Int. Geophys. Ser. 15
[4] Fujita M, Yoneyama K, Mori S, Nasuno T and Satoh M 2011 Diurnal convection peaks over the eastern Indian Ocean off sumatra during different MJO phases J. Meteorol. Soc. Japan. Ser. II 89 317–30
[5] Chang C-P, Wang Z and Hendon H 2006 The Asian winter monsoon (Heidelberg Berlin: Praxis - Springer)
[6] Chang C P, Harr P A, McBride J and Hsu H H 2004 Maritime continent monsoon: Annual cycle and boreal winter variability East Asian Monsoon pp 107–50
[7] Webster P J and Fasullo J 2003 Tropical Meteorology and Climate: Monsoon: Dynamical Theory Encyclopedia of Atmospheric Sciences: Second Edition (Elsevier Science Ltd.) pp 1370–86
[8] Giannini G, Dupe Z L and Mustofa M A 2012 Kajian Awal Musim Hujan dan Awal Musim Kemarau di Indonesia J. Meteorol. dan Geofis. 13 1–8
[9] Fowler H J, Blenkinsop S and Tebaldi C 2007 Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling Int. J. Climatol. 27 1547–78
[10] Hidalgo H G, D. D M and Cayan D R 2008 Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States Calif. Clim. Chang. 48
[11] Zhang T, Yang S, Jiang X and Zhao P 2016 Seasonal-Interannual Variation and Prediction of Wet and Dry Season Rainfall over the Maritime Continent: Roles of ENSO and Monsoon Circulation J. Clim. 29
[12] Syahputra M R 2013 Kajian Prakiraan Curah Hujan Musiman dengan Metode CA (constructed Analogue) dari Keluaran Model CFS (Climate Forecast System) Studi Kasus: Pulau Jawa dan Sumatera (Institut Teknologi Bandung)
[13] Garcia E 2006 Cosine Similarity and Term Weight Tutorial
[14] Surmaini E, Hadi T W, Subagyno K and Puspito N T 2015 Prediction of Drought Impact on Rice Paddies in West Java Using Analogue Downscaling Method Indonesia. J. Agric. Sci. 16 21–30
[15] Fernández J and Sáenz J 2003 Improved field reconstruction with the analog method: Searching the CCA space Clim. Res. 24 199–213
[16] Azizah S 2017 Onset monsun dan hubungannya dengan awal musim hujan di Indonesia (Institut Teknologi Bandung)
[17] Brier G W 1950 Verification of Forecasts Expressed in Terms of Probability *Mon. Weather Rev.* **78** 1–3

[18] Murphy A H 1973 A New Vector Partition of the Probability Score *J. Appl. Meteorol.* **12**