An alternative index to NINO3.4 for predicting monthly rainfall anomaly in East Java

B E A Haq1*, M Ryan1, A Kurniawan2 and A M Rafi2

1 Meteorology Climatology and Geophysics Agency Head Quarter, Angkasa I Street 2, Kemayoran, Jakarta, 10720, Indonesia
2 Malang Climatological Station, Malang, 65152, Indonesia

*E-mail: bagasega2a@gmail.com

Abstract. ENSO and NINO3.4 index are known to have some relation with Indonesian monthly rainfall anomaly. There is a gap between scientific studies on one hand and forecasting operational problems on the other hand since previous studies are not giving enough attention to the N+1,2,3 concept. The concept is about giving three next month rainfall anomaly prediction rather than connecting ENSO index with three-monthly rainfall anomaly. Here we propose an alternative index for ENSO. The median of categorical gridded rainfall anomaly of East Java is used as a general representation. Plots of correlation between the median and anomaly sea surface temperature from ReynSmithOIv2 are used to determine locus candidate to be compared with NINO3.4. The Near Maritime Continent (NMC) index is selected and proven to have a significant average difference in correlation between based on bootstrap technique. Verification of prediction used in this study is simulation-based and only uses binary hit-miss final result. Prediction is generated by simple linear regression with three lag times (2, 3, and 4). Verification based on three categories shows that NMC’s hits are higher than NINO.34 in lag-2 and lag-3. In lag-2, NMC’s verification is 57.5% compared to only 38.7% for NINO3.4. However, NINO3.4 is a still better predictor in lag-4. Radar charts of monthly verifications are also developed.

Keywords: NINO3.4, NMC, simulation-based

1. Introduction
The relation between rainfall and sea surface temperature (SST) is known. Rainfall in Indonesia has some relation with the El Nino Southern Oscillation (ENSO). In East Java, Supari’s work using composite analyses shows that El Nino brings drier condition than neutral and La Nina brings wetter on seasonal rainfall especially in JJA (June-July-August) and SON (September-October-November) [1]. Another study also used composite analysis for examining El Nino and El Nino Modoki modulation on seasonal rainfall in Indonesia [2]. Composite analyses depend on the average of many events. ENSO itself is not something with a single index to determine whether the condition is La Nina or El Nino. There are at least seven indices for monitoring this phenomenon [3]. ENSO pictured...
by SST anomalies tend to slowly vary in time. It can drive cloud development so that it relates to a mesoscale convective system in Indonesia so that can increase rainfall [4].

All studies that have been cited in the paragraph before show an ENSO relation with rainfall. Unfortunately, it is not in monthly scale. We identify a problem on the time-scale if we see ENSO through climate forecasting in East Java. Information about rainfall anomaly is one of the obligations to be produced with format N+1,2,3. It means that if this month is January, prediction for rainfall anomaly in February, March, and April should be produced. There is a gap between scientific information and operational level here. We want to investigate ENSO relation to monthly rainfall in East Java more straightforwardly.

Hendon’s and Aldrian’s works on investigating the relation between rainfall and local SST motivates us to study better relation between SST around East Java and monthly rainfall [5][6]. We hypothesize that Tobler’s First Law in Geography can also be applied to SST and rainfall. “Everything is related to everything else, but near things are more related than distant things” [7]. The study for searching new SST-based index is not something new. Wang’s work in 2017 tried to find a new index of El Nino Modoki on purpose to China’s rainfall information [8]. Earlier modelling studies show substantial sea surface and atmosphere interaction on narrow bay. The bay condition is related to convection and the monsoon system in India [9]. East Java has many narrow bays, so we think that the interaction of SST near East Java will have a heavier impact. We want to find and prove a new index based on SST. It aims for predicting monthly rainfall anomaly in East Java. We hypothesize that it will be a better predictor than the ENSO-based index.

2. Data and Method

![Rainfall raster and rainfall anomaly raster of November 2018](image)

Figure 1. Rainfall raster and rainfall anomaly raster of November 2018

We use gridded rainfall data used by Malang Climatological Station from January 1991 until January 2020 (349 rasters). We used data used at operational level. These rasters are produced using inverse distance weighting (IDW) because IDW is better than Regression-Kriging and Ordinary-Kriging for East Java’s case [10]. These raster data are then divided by the raster of normal (based on 1981-2010) for each month, so it becomes rainfall anomalies. At the operational level, exact values are not important. Values of rainfall anomaly are then plotted in 7 categories, as shown below.

| Category          | Category | Percentage |
|-------------------|----------|------------|
| Below Normal      | 1        | 0 – 30     |
|                   | 2        | 31 – 50    |
|                   | 3        | 51 – 84    |
| Normal            | 4        | 85 – 115   |
|                   | 5        | 116 -150   |
| Above Normal      | 6        | 151 -200   |
|                   | 7        | >200       |
Hence, we convert 349 rasters to 349 categorical rasters. Instead of using principal components as representation of monthly rainfall data, we use the median of categorical rainfall anomaly as East Java’s representation due to its simplicity [11]. NA (not available/empty) values are omitted in the calculation of median.

We use the principle of apple to apple. For making monthly rainfall anomaly, we should use another anomaly, in this case, SST anomaly. Rainfall anomaly should be coupled with the anomaly of SST. Hence, SST is not used in this study. The anomaly of SST in this study is based on ReynSmithOIv2 Monthly SSTA data downloaded from IRIDL Columbia web [12]. It is a free-open SST data so next research will be easier to be done. Niño-3–4 zone is used as ENSO-based sea surface temperature representation. These anomalies in the zone are averaged over 5°–5°N and 160°E–150°W [13][14].

We performed correlation analyses between the East Java median categorical monthly rainfall anomalies with all grid of available anomaly of SST (ASST). We examine correlation in three lag systems, lag-2, lag-3, and lag-4. Lag-2 means that ASST from January 1991 until November 2019 is coupled with the median data from March 1991 until January 2020. We use only a three-month lag system due to the principle of N+1,2,3 at the operational level. All correlation values are then plotted. The locus candidate will be selected through this process. Bootstrap will be done to make sure the significance of the average difference. The bootstrap method can be used to assign significance of two different groups of data. Supari’s work did it for assessing composite difference [1]. In this study, we use it to assess the average correlation on a zone.

For making justification about which metric be the best, verification-based competition will be used. Using straightforward linear regression, we will do an ASST-based prediction simulation. Simulation-based mean we do not use unavailable data at a specific time [15]. The first period of prediction simulation will be January 2011. It means the first model training will be zone averaging ASST from January 1991 until December 2011, NINO3.4 and locus candidate. The first prediction will be February, March, and April 2011. Prediction of February 2011 rainfall anomaly is based on ASST condition in December 2010 because the ASST of January 2011 is not available in January 2011. All three predictions will be based on December 2010. The predictions are based on lag-2, lag-3, and lag-4 concepts. We do split data in two parts, but not only once. We do it repeatedly. This study is simulation-based research. In the next simulation period, one datum on each series will be added, and the regression will be recalculated. The last simulation period will be October 2019, so that we calculate 106 regressions for this study.

The verification process in this study is using contingency tables to make it to be simpler [16]. The table will only consist of 3 categories (Below Normal, Normal, and Above Normal). If the prediction is same as the observation, it will be declared as hit, vice versa. Binary classification can explain something intuitively easier to be understood [17]. Only two simple terms are used, hit or miss. We use what used at operational level so if the forecasters read this article, it will be far easier for them to understand. R Statistics software is used for processing in this study [18]. Package raster is also used to enable raster calculation [19]. All scripts are available to be discussed through personal e-mail.

3. Result and Discussion
The median of categorical gridded rainfall anomaly of East Java is used due to its simplicity. If the principal component analysis is used, there will be lot explanations coming with it. In time we use median, all we can explain is that half of the data are above it. The rest of data are below the median. It shows datum in the middle. Figure 2 shows how the median data change over time.

The median shows that in famous El Nino years (1997 and 2015), the median is going down. Vice versa, the median is going up in famous La Nina years (1998 and 2010). These facts bring no debate that East Java rainfall anomalies have correlation with ENSO index. However, there are more noises that are not explained by ENSO index are shown in Figure 2. It brings a possibility to have a challenger for an alternative SST-based index.
Figure 2. East Java median categorical monthly rainfall anomalies

The median data above become both model builder and model verification in the following calculation. Correlation with all ASST grids available is shown in Figure 3. There are positive correlation patterns in the near maritime continent if lag-2 is used. This pattern confirms Tobler’s Law. It appears in the area near East Java. Uniquely, its correlation values seem reversed to the NINO3.4 index. There are also other alternatives if we only see the correlation. Some areas in the Atlantic show a similar pattern, but it is farther from East Java compared to NINO3.4. So, we decide that the locus candidate will be the region of a box with an extent of 90E-150E and 25S-15N. In the subsequent discussion, the index will be called the near maritime continent (NMC) index.

Bootstrap analyses with 1000 resample are then done to ensure that these correlation values’ average is better than NINO3.4’s average. For making it simple, we do use absolute function after averaging in the bootstrap. This sign-changing will not change the game. The closest one will still be observed. Figure 4 shows that the NMC correlation is more potent than NINO3.4. The highest difference is shown in lag-3.

Figure 3. Correlation plot between ASST and the median data: lag-2 (top-left), lag-3 (top right), and lag-4 (bottom)
Verification of predictions using linear regression shows that NMC has a better verification value than NINO3.4. For lag-2, verification of NMC is 61 hits from 106 or 57.5%, and NINO3.4 is only 41 (38.7%). For lag-3 and lag-4, NMC’s hits are 56 (52.8%) and 53 (50.0%) meanwhile NINO3.4’s hits are 49 (46.2%) and 54 (50.9%). Verification of NINO3.4 is better than NMC in lag-3. If we look in more detail, NMC’s hits are decreasing with an increase of lag, contrary to NINO3.4’s hits are increasing. The mechanism behind it can be theorized that the NINO3.4 signal needs more time to disturb the East Java rainfall anomaly pattern.

Both hits can be grouped into months. Figure 5 describes radar charts of monthly verifications of two indices. Numbers 1-12 are represented January until December. The radar axis shows the value of hits. We find that NMC-based prediction does better in April and October in lag-2. For lag-4, NINO3.4-based prediction is better in May, November, and December.

4. Conclusion
After selecting locus candidate using correlation plots, a region with extent 90E-150E and 25S-15N is called the near maritime continent (NMC) index. NMC can defeat NINO3.4 in predicting East Java rainfall anomaly in lag-2 and lag-3 using simple linear regression. It should be admitted that NINO3.4
is still a better predictor in lag-4. We encourage other derivative works to investigate other possibilities in simulation-based verification for producing better monthly rainfall anomalies.

Acknowledgments
This project is one of five projects dedicated to participating in ICTMAS called ASRAB. All authors declare that there is no conflict of interest in doing this project. We would like thanks all entities that had supported us.

References
[1] Supari, F Tangang, E Salimun, E Aldrian, A Sopaheluwakan and L Juneng 2018 Clim. Dyn. 51, 2559–80
[2] Iskandar I, D O Lestrai and M Nur 2019 Makara J. of Sci. 23 217-22
[3] Capotondi A et al 2015 Bull. Am. Meteorol. Soc. 96 921–938
[4] Rustiana S, Trismidianto and Satyawardhana H 2019 IOP Conf. Ser.: Earth Environ. Sci. 303 012006
[5] Hendon H H 2003 J. of Clim. 16 1775–90
[6] Aldrian E and D R Susanto 2003 Int. J. Climatol. 23 1435–52
[7] Tobler W 1970 Economic Geography 46 234–40.
[8] Wang X, W Tan and C Wang 2017 Clim. Dyn. (10.1007/s00382-017-3769-8)
[9] Samanta D, N H Saji, J Dachao, T Vishnu, G Malay, S A Rao and M Deshpande 2018 Sci. Rep. 8 17694 (10.1038/s41598-018-35735-3)
[10] A Kurniawan, E Makmur and Supari 2021 Pros. Sem. Nas. Geomatika 2020, R Ati et al (Bogor: Badan Informasi Geospasial) p 263-72
[11] Kang H M and F Yusof 2012 Int. J. Contemp. Math. Sciences 7 9 –22
[12] Reynolds R W, N A Rayner, T M Smith, D C Stokes and W Wang 2002 J. Climate 15 1609-25.
[13] Kug J S, F F Jin and S I An 2009 J. Climate 22 1499–1515
[14] Yeh S W, J S Kug, B Dewitte, M H Kwon, B Kirtman and F F Jin 2009 Nature 461 (10.1038/nature08316)
[15] R Fredyan, A Kurniawan, A Naba and Abdurrouf 2021 Pros. Sem. Nas. Geomatika 2020, R Ati et al (Bogor: Badan Informasi Geospasial) p 87-96
[16] Muharsyah R 2017 J. Met. dan Geo. 18 33-44
[17] Mehr A D, Nourani V, Hrnjica B and Molajou A 2017 J. of Hydro. 555 397-406
[18] R Core Team 2020 R: A language and environment for statistical computing (Vienna:R Foundation)
[19] Hijmans R J et al 2020 Geographic Data Analysis and Modeling (CRAN/raster)