RESEARCH PUBLICATION NO. 2/2018

Statistical Approach Towards Subseasonal Prediction over the Maritime Continent

By
Norlaila Ismail and Changhyun Yoo

Perpustakaan Negara Malaysia Cataloguing-in Publication Data

Published and printed by

Malaysian Meteorological Department
Jalan Sultan
46667 Petaling Jaya
Selangor Darul Ehsan
MALAYSIA

Contents

STATISTICAL APPROACH TOWARDS SUBSEASONAL PREDICTION OVER THE MARITIME CONTINENT

Norlaila Ismail and Changhyun Yoo

Abstract

Weather and climate prediction over the Maritime Continent (MC) remains as a challenge due to its complex mix of islands and seas as well as its mountainous topography. In general, the weather and climate of the region are contributed by climates in various temporal and spatial scales. Among which are the interannual variability associated with the El-Niño Southern Oscillation (ENSO) and the stratospheric Quasi-Biennial Oscillation (QBO). The Madden-Julian Oscillation (MJO) also has the capability of modulating intraseasonal convection activity over the MC region. Previous studies have shown successes in generating a future prediction using statistical methods. Some of these works employed climate variability as the source of predictability in their models. Motivated by this, we construct simple and multiple linear least-square regression models using climate indices mentioned above to forecast the outgoing longwave radiation (OLR) and surface air temperature (T2m) anomalies over the MC region. The prediction skill is measured by the correlation coefficients between forecasts and the observed values. Results show that when simple regression model is built using individual index separately, ENSO has the highest temporal correlation coefficients. However, it is the multiple regression models using a combination of all the climate indices with the inclusion of the persistence term that produce the most promising forecast skill at all the lags especially at first lead week with r = 0.74 and r = 0.77 for OLR and T2m respectively on the annual time scale. To add, multiple regression models with persistence exhibit satisfying correlation in all seasons except for DFJ especially after lead week-3 for OLR. For T2m, good correlation is obtained in MAM and JJA. Further evaluation using the root mean squared error skill score (RMSESS) support the potential use of utilizing the multiple regression models. Results suggest that the persistence term elevates the prediction skill for both OLR and T2m.

1. Introduction

1.1 Maritime Continent

One of the interesting feature possessed by the Maritime Continent (MC) lies on its location and its composition that comprises of a complex blend of islands, peninsula and water bodies. Thus, its weather and climate are heavily governed by the interaction between the lands and seas. The examples of the interactions are the land-sea breeze and monsoon activity on smaller and bigger scales respectively. The importance of the Maritime Continent as the biggest energy contributor to the global atmospheric circulation was highlighted through the work of Ramage (1968). It was pointed out that among the three tropical continental regions which serve as convective centres, it is the Maritime Continent that produces the highest amount of latent heat. Latent heat of condensation is released when water vapour is converted to moisture during condensation in the process of convection. A huge amount of energy is therefore discharged to the atmosphere originating from this region especially during boreal winter.

The Maritime Continent is greatly affected by the monsoon system. Its two main monsoons are summer and winter monsoons corresponding to the boreal summer and boreal winter respectively with two transition periods in between them. Normally, the region will be associated with the wet condition during the northern hemisphere winter and dry condition during northern hemisphere summer. Aside from the monsoon system, the weather and climate variability of the Maritime Continents are also influenced by interannual and intraseasonal phenomena such as the El-Niño Southern Oscillation (ENSO), Madden-Julian Oscillation (MJO) and Quasi-Biennial Oscillation (QBO). Various studies have shown that all the above-mentioned oscillations have the capability of modulating the weather and climate globally. For this reason, it is natural to have the assumption that they could contribute to stronger and more direct effect on the Maritime Continent.

1.2 Climate Variability

1.2.1 El-Niño Southern Oscillation (ENSO)

On an interannual time scale, the most prominent form of natural phenomenon that influences the variability in the Maritime Continent is the ENSO. The ENSO is a coupled climate variability between the atmosphere and the ocean which occurs over the tropical Pacific Ocean. It is characterized by sea-surface temperature anomalies and linked to the changes in atmospheric circulation. The phases of the ENSO can be divided into three phases i.e the neutral, warm (El-Niño) and cold (La-Niña) phases. Climatologically or during the neutral phase of ENSO, the sea-level pressure over the eastern Pacific is much higher than the sea-level pressure over the western Pacific. This difference in east-west pressure gradient initiates easterly wind also known as the trade wind, to blow from South America to the Western Pacific. The air converges over the Western Pacific, moves upward, and sink over South America, completing a Walker circulation. During the El-Niño, the trade wind weakens and pull the convergence zone over the central Pacific instead, causing drought to develop over the Maritime Continent.

Numerous research has been done between the associations of the ENSO with the Maritime Continent. In a broad sense, most of the findings agree that El-Niño tends to suppress the rainfall while La-Niña favors more rainfall over the Maritime Continent. Even though there is a contradiction of the final conclusion between the work by Haylock and McBride (2001) and Tangang and Juneng (2004), both of their work acknowledge that seasonality and geography can impose variation of ENSO affect towards the region. Haylock and McBride (2001), found a high correlation between Indonesian rainfall with ENSO during the monsoon transitions period and no relationship during the winter season. In contrast, Tangang and Juneng (2004) shown that ENSO correlates well with rainfall during fall and winter seasons.

1.2.2 Quasi-Biennial Oscillation (QBO)

Another interannual oscillation that affects the Maritime Continent is the QBO. The QBO is a dominant modulator in the tropical stratosphere. It is a phenomenon whereby the zonal wind in the lower equatorial stratosphere alternate from westerly to easterly with an approximately two year period. Zonal stratospheric winds from the upper stratosphere will propagate to the lower stratosphere as time passes. The QBO westerly winds generally travel downward faster than the easterly wind but easterly winds are much stronger than its counterpart. One of the most widely accepted theories of the QBO was proposed through the work of Holton and Lindzen (1972). They postulated that the vertical propagation of tropical Kelvin and Rossby-gravity waves are responsible for providing westerly and easterly momentum to the upper atmosphere. While propagating vertically upward, both waves will meet a critical point where they will break, deposited their momentum in the zonal mean flow and eventually make up the QBO.

Since its location and maximum amplitude occur at the equator and has been cited to have teleconnection with higher latitude atmospheric phenomenon such as the polar vortex, it is natural to assume that the QBO has a stronger and direct influence towards the Maritime Continent as they are collocated at a similar location. Plausible ways of the QBO to affect the Maritime Continent are by influencing the growth of convective clouds through the changes in static stability or by the modification of the zonal wind shear at the upper troposphere and lower stratosphere (Collimore et al. 2003). Deeper convective clouds could be expected during the east QBO (EQBO) phase and smaller convective clouds during the west QBO (WQBO) phase. The former case tends to boost the tropopause height through the upward motion in the mean meridional circulation, while the latter has the tendency to prohibit vertical development of the tropopause height.

1.2.3 Madden-Julian Oscillation (MJO)

On an intraseasonal time scale, the largest contributor to tropical variability is the MJO. In contrast to the ENSO and QBO which are classified as stationary oscillations in which the former mainly reside over the tropical Pacific and the latter is locked over the equator, MJO migrates from one location to another. One of its prominent characteristics is the deep convection anomalies which usually originates over the Indian Ocean and slowly propagates towards the Pacific Ocean. As it propagates eastward, it causes fluctuation on the sea level pressure (SLP), sea surface temperature (SST), wind structure and rainfall. Often time, the Maritime Continent seems to hinder the strength of the MJO convective cell as it passes over the region. This is possibly caused by the lack of moisture and energy over the region due to strong diurnal convection, poor moisture convergence caused by topography and limited surface evaporation (Zhang 2005).

A common method of measuring the strength and location of the MJO is by utilizing the Real-Time Multivariate MJO Index (RMM Index) constructed by Wheeler and Hendon (2004). This index classified the phases of MJO based on the geography and the amplitude of the convection envelope with a total of eight phases. Each phase represents the particular location of the convective cells. For instance, phase two and three, phase four and five and phase six and seven indicate that the active convection cell is located over the Indian Ocean, Maritime Continent and Western Pacific respectively. Last but not least, the remaining two phases represent the location over West Africa. The magnitude can be interpreted as the distance of one point from the centre of the phase diagram of the RMM Index.

Association of the MJO with the Maritime Continent was demonstrated by Tangang et al. (2008) where they investigated the possible role of atmospheric variables in causing an extreme flooding case that took place over the southern peninsula Malaysia. They suggested that the presence of the MJO over the Indian Ocean promote strong easterly over the Maritime Continent. The easterly wind enhances the southern peninsular Malaysian rainfall by two ways, one it inhibits the formation of the Borneo vortex. Normally, when the Borneo vortex is present, most of the rainfall will be concentrated over the western part of East Malaysia. Second, the easterly winds strengthen the effect of the northeasterly wind. As the cold and dry northeasterly wind originating from the Siberian High get heats up as it passes over the warm tropical sea, it also merged with the warm and moist easterly from the Pacific Ocean leading to a more intense formation of convective clouds over southern peninsular Malaysia.

Hidayat and Kizu (2010) investigated the pathways of MJO in affecting the rainfall variability over Indonesia. From their composite analyses, they found that phases of the MJO determine the amount of Indonesian rainfall. Different MJO phases will modulate different rainfall amount to a particular region. A diverse pattern of wet and dry phase can be observed over Indonesia when the MJO is in phase 1-4 and phase 5-8 respectively. A similar variation of rainfall and 850-hPa winds during boreal winter over the Maritime Continent at various MJO phases has also been documented in a study by Xavier et al. (2014). Their study concluded that the probability of rainfall increased for MJO phase 2-4 and the opposite for MJO phase 6-8 over the Southeast Asian (SE) region.

The further relationship between the climate indices are shown in Figure 1 and Figure 2. Figure 1 depicts the OLR pattern when it is regressing with each of the climate indexes. The positive values indicate that the index and the OLR have the tendency to move in the same direction. The opposite can be said for the negative values. For instance, in panel (a), it can be seen that the Maritime Continent and ENSO index has a positive relationship. This implies that when the value of ENSO index increases, the values of the OLR over the Maritime Continent increases as well. The relationship pattern for OMI 1 and OMI 2 are shown in panel (b) and (c). Association of the climate index can also be expressed using a composite map. An example for this is Figure 2 where its show the pattern of the OLR for different QBO phases.

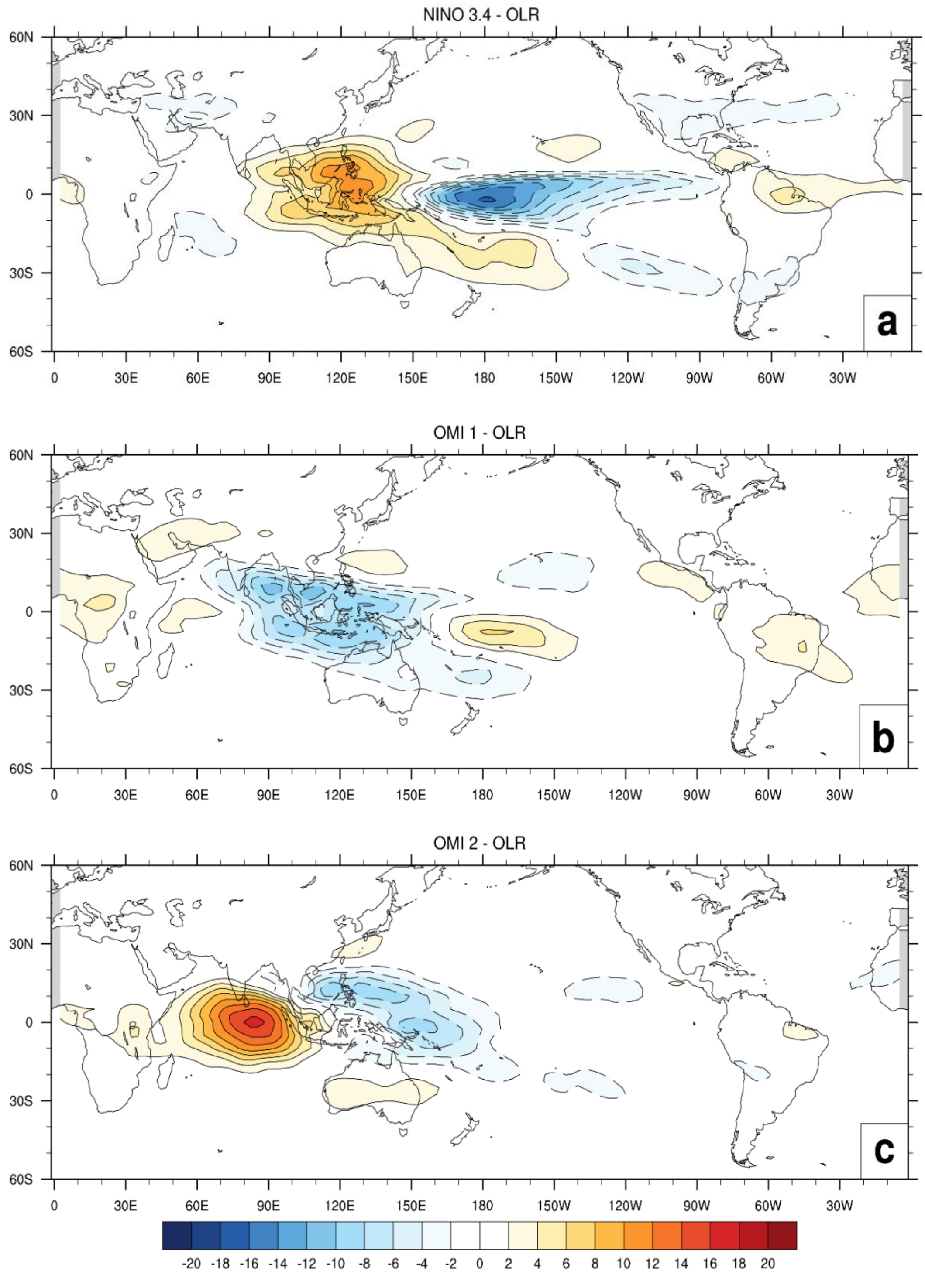

Figure 1 Regression map between the OLR with (a) ENSO Index (b) MJO – OMI 1 and (c) MJO – OMI 2

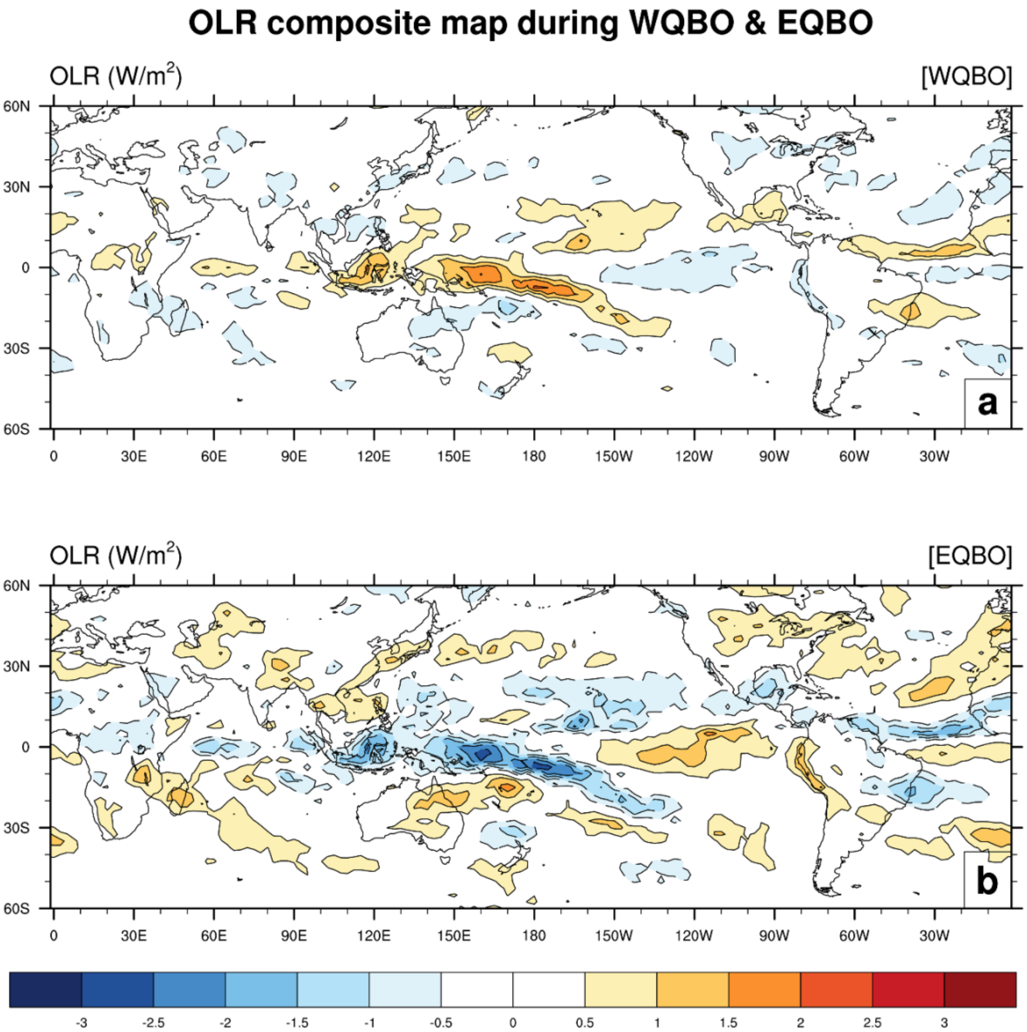

Figure 2 Composite OLR anomalies for the WQBO and the EQBO.

2. Objective

The primary objective of the study is to investigate the usefulness of using simple and multiple regression methods to forecast the OLR and surface temperature over the Maritime Continent. Provided below are the justification of using the OLR and surface temperature as forecast variables in this study.

The usage of OLR as a proxy for tropical convection is not new and has been practised in multiple studies such as in (Wheeler and Hendon, 2009). Negative (positive) OLR anomalies and their eastward propagation are in good agreement with positive (negative) rainfall anomalies and eastward migration of convective anomalies based on the work of (Hidayat and Kizu, 2010) which analyzed Indonesian station rainfall data. The OLR is more relevantly to be associated with big scale convection produced by large scale climate variability rather than the localized convection. Since we are dealing with the ENSO, QBO and MJO it is appropriate to use the OLR as another measure of precipitation. Examples of past studies that have established the influence of these natural oscillations with deep tropical convection in the Maritime Continent are the ENSO (Juneng and Tangang, 2005), MJO (Hidayat and Kizu, 2010) and the QBO (Collimore et al., 2003).

The rationale of using the surface temperature is as the aftermath of the Malaysian heatwave episode happening in March 2016, in which has caused six cases of heat exhaustion and one death (Malaysian Ministry of Health, 2016). Moreover, a strong El-Niño episode like 1997/98 could also impose changes on Maritime Continent surface temperature. The linkage of El-Niño to the rise in surface temperatures in Malaysia has been studied by Tangang et al. (2007), whereby it was found that the interannual variability of Malaysian temperature is largely dominated by the ENSO and that all region experience uniform warming during an El-Niño event. Additionally, the MJO has also been shown to have a connection with surface temperature variability in North America (Rodney et al., 2013; Johnson et al. 2014). However, little work has been done to a model surface temperature in the Maritime Continent using the MJO as a predictor. In his work, Collimore et al. (2003) have shown that the QBO exert an impact on tropical tropospheric temperature but disregard its impact on surface temperature. Hence, we seek to investigate the usefulness of the QBO as a predictor of surface air temperature.

The second objective is to examine the potential of using climate indices of ENSO, MJO, QBO and autocorrelation of the predictands as the predictors in the regression models.

3. Methodology

3.1 Data and Pre-processing

In this study, we used climate indices of ENSO, MJO and QBO as the predictors for our statistical model. The indices were chosen as predictors based on the findings from the studies mentioned in the introduction as well as data availability. Additionally, Outgoing Longwave Radiation (OLR), taken as a proxy of convection and 2-metre surface temperature (T2m) were appointed as the predictand fields. Briefly, the predictor is an independent variable that could provide information to forecast the predictand. The predictand on the other note is the variable or parameter that need to be predicted, which is dependent on the predictor.

3.1.1 The Predictors

To construct the ENSO index, we used the sea surface temperature (SST) Version 1.1 of Hadley Center Global Ice and Sea Surface Temperature (HadISST1) provided by the United Kingdom Meteorological Office (UKMO) and is available in 1.0^{o} x 1.0^{o} resolution (Rayner et al. 2003). The data are available at (http://www.metoffice.gov.uk/hadobs/hadisst/). The monthly SST data were interpolated to daily time series since the focus of the present study is for sub-seasonal prediction. In order to characterize the ENSO variability, we calculated the Niño 3.4 index by taking the area-averaged of the Pacific Ocean SST anomalies over the region of 5^{o}S-5^{o}N and 170^{o}W-120^{o}W.

Outgoing longwave radiation (OLR)-based MJO index (OMI) has been chosen to represent the intraseasonal variability in this study as opposed to the commonly used Real-time Multivariate MJO index (RMM) by Wheeler and Hendon (2004). Based on a study by Kiladis et al. (2014), OMI is well correlated with the RMM but has more capability at tracking MJO convective signal. The OMI index was extracted from NOAA Earth System Research Laboratory (NOAA) and accessible through (http://www.esrl.noaa.gov/psd/mjo/mjoindex/). The OMI was formulated by projecting the 20-90 days band-pass filtered OLR onto two leading EOFs of the 30-90-day eastward-filtered OLR.

Currently, Medium-Range Weather Forecast Reanalysis (ERA-INTERIM) dataset (Dee et al. 2011) is considered to be the most accurate reanalysis product available. Therefore, we utilized the daily zonal mean zonal wind from ERA-INTERIM to calculate the QBO index. The QBO index was developed by calculating the area-averaged of zonal mean zonal wind anomalies at 50 hPa between 10^0S-10^0N.

3.1.2 The Predictands

Outgoing longwave radiation (OLR) was obtained from NOAA Earth System Research Laboratory (NOAA). It is a gridded daily OLR data and has a spatial resolution of 2.5^0 x 2.5^0 global grid (Liebmann and Smith. 1996). Another atmospheric variable that was chosen as the predictand was the daily surface temperature at 2-metre height (T2m). Similar to the zonal mean zonal wind, this dataset was taken from ERA-Interim in 1.5^0 x 1.5^0 grids. OLR and T2m were also used as one of the model predictors and included as the persistence term.

The data between the time periods of 1979 - 2012 were used in the analysis of this paper. All daily mean anomalies were obtained by removing the climatological daily mean of the entire analyzed period for both predictors and predictands. Concisely, an anomaly is the deviation of certain variables from their long term average mean. Specifically for this study, we calculate the area-average of both OLR and the T2m anomalies as we are interested to capture the variation of both parameters over the Maritime Continent in the entire focus period. In contrast to the Maritime Continent portrayed in Ramage (1968) which include Malay Peninsula, Borneo, Java, Sulawesi and New Guinea among other islands, we chose a smaller domain between 15°S-8°N and 95°E-130°E, illustrated in Figure 3 as our Maritime Continent focusing over Malaysia and Indonesia for the purpose of this paper.

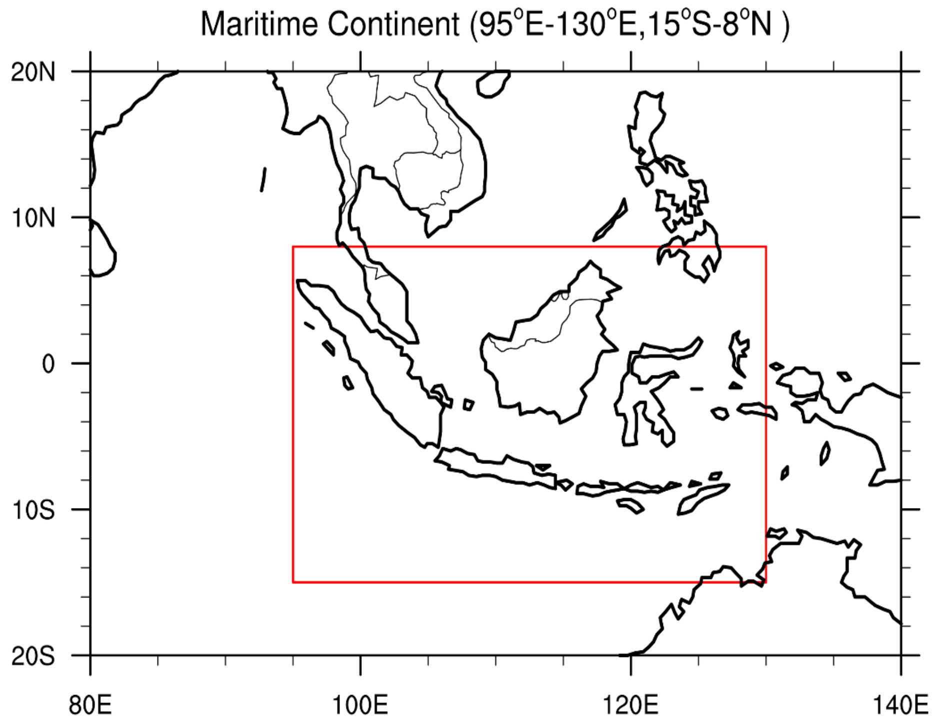

Figure 3 The Maritime Continent (MC) define as 95°E-130°E, 15°S-8°N.

3.2 Statistical Modelling

3.2.1 Linear and Multiple Regression Analysis

The simple and multiple least-square linear regression methods were implemented in the development of the statistical models in this paper. These statistical models are classified as classic regression method as the prognosis are generated without any input from numerical weather prediction (NWP) and completely based upon statistical techniques (Wilks 2011).

For a simple linear regression model, the ultimate goal is to seek a one to one relationship between the dependent and independent variable. In the case where the association of one dependent variable to various independent variables is sought, the method is known as the multiple linear regression methods. Caution should be taken when selecting multiple variables to act as the predictors in the regression models. This is because the inclusion of more than one predictor does not necessarily contribute to a more accurate outcome. The statistical model for simple regression (1) and multiple regression (2) are

$$\hat{y}(t+1) = \alpha + \beta x + \varepsilon \qquad (1)$$

$$\hat{y}(t+1) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (2)$$

where $\hat{y}$ is the predicted value of y, α is the regression slope, β is the coefficient and ε is the residual term. The residual term could be used as an indicator to gauge the performance of the regression model.

3.2.2 Statistical Model Description

In this current study, as mentioned before, we generate simple and multiple regression models to predict the OLR and surface temperature over the Maritime Continent. We started by creating simple regression model using each climate index as the predictor field. Therefore, we have four simple regression models built using ENSO, MJO-OMI 1, MJO-OMI 2 and QBO respectively for six different lead times.

We divided our 34 years data into two non-overlapping subsets known as the training period (1979 – 2005) and the validation period (2006 – 2012). The training period is used to determine the regression coefficient while the validation period is used to verify the forecast skill. This partition is made to reduce the chances of inflation of skill by the model (Jolliffe and Stephenson 2003). The regression coefficient is responsible for storing the information between the predictor and the predictand which is subsequently used to build the regression model.

Next, we combined all the climate indices together as the predictors and repeated the same procedures to construct new regression model. Since more than one predictors are used in generating this forecast model, it is more befitting to call it as a multiple regression models instead of the simple regression model. The generic equation for simple and multiple regression models is described by equation (3) and (4) respectively

$$\hat{y}(t+1) = \beta_0 + \beta x \tag{3}$$

$$\hat{y}\,(t+1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \tag{4}$$

where $\hat{y}\,(\mathrm{t}+1)$ is the forecasted predictand; β_0 is the y-intercept; $\beta_{1\,,}\,\beta_{2,}\,\beta_3$ and β_4 are the regression coefficients corresponding to $x_1, x_{2,}\, x_3$ and x_4 which is ENSO, MJO-OMI 1, MJO-OMI 2 and QBO index respectively; β is the specific climate index; and t+1 = 1-,2-,3-,4-,5-,6-week which represent the predictors lead time.

We later incorporate the persistence term in all of the models above. The persistence term can be defined as the current observation of the predictand (at t = 0) and is added in hope to increase the predictive capability of our models. The equation for the models with the inclusion of the persistence term is similar to (3) and (4) but with a new coefficient and variable to represent the persistence. The model built with just one climate index and persistence as predictors is given as (5) and the combination of all predictors and persistence is given as (6)

$$\hat{y}(t+1) = \beta_0 + \beta x + \beta_5 x_5 \tag{5}$$

$$\hat{y}\,(t+1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \tag{6}$$

Similar to the description given for equation (3) and (4), $\hat{y}\,(t+1)$ and β_0 is the predictand to be forecasted and y-intercept respectively. While $\beta_{1\,,}\,\beta_{2,}\,\beta_3$ and β_4 are the regression coefficients corresponding to ENSO, MJO-OMI 1, MJO-OMI 2 and QBO index; β is the specific climate index. The new addition to the equation are β_5 and x_5 representing the persistence term.

Using all of the regression models described above we calculate their annual and seasonal correlation coefficient and root mean square error skill score (RMSESS). The seasons are specified by dividing a year into four seasons; December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON) for boreal winter, spring, summer and fall accordingly.

4. Results and Discussions

We chose two metrics in order to evaluate the performance of our forecasts which are the correlation coefficient and the RMSE skill score. The correlation coefficient is broadly used as a verification metric in the atmospheric science field. It measures the linear relationship between the variables and as in our case, it represents the linear association between the forecasts and the observations. MSE on the other hand, measure the accuracy of the forecasts and observations.

4.1 Correlation Coefficients

We started the study by constructing simple regression model for each of our selected predictors which are ENSO, MJO-OMI 1, MJO-OMI 2 and QBO index separately using equation (3). Next, we are interested to investigate the performance of the regression model in the case where all of the predictors are accumulated together. For this purpose, we used equation (4) to build multiple regression models by combining all of the climate indices to generate the forecasts. As discussed earlier, the validation period is set to be independent of the training period to avoid any added skill caused by the overlapping of both periods.

Figure 4 illustrates the results of the correlation generated from the models mentioned above. Four simple regression models are represented by purple, green, light blue and yellow each corresponding to individual predictor namely, the ENSO, MJO-OMI 1, MJO-OMI 2 and QBO index respectively. The dark blue line, on the other hand, is the correlation corresponding to the combined indices model. As shown in Figure 4 (a), among the forecasts made with only single climate index as the predictor, the highest correlation coefficient is generated by the ENSO index at all lead times. For ENSO, the maximum correlation appears at lead week 3 and week 4 with $r = 0.52$, while the minimum value is $r = 0.49$ occurring at first and last lead week.

Correlation coefficients produce by the ENSO and the QBO show almost constant variation, while the MJO components display fluctuation from lead week-1 to week-6. Small variation by the ENSO and the QBO can be justified due to their nature as low-frequency climate oscillations while the rapid change in the MJO components are owing to their short time scale. Except for the ENSO, other single predictor models produce correlation less than r = 0.40 for all weekly lead time. However, it is the model with multiple predictors that produce the highest correlation coefficients than the other four simple regression models, with r = 0.65 at lead week-1, and hovering at around r = 0.50 for the rest of the lead weeks.

Depicted in Figure 4 (b) are the correlation coefficients obtained using equations (5) and (6) where we incorporate the persistence term in the regression models. Pronounce increase of the correlation coefficients can be seen in all of the models especially in first three lead weeks. As before, the best correlation coefficients come from the combined climate indices model. The correlation peak at r = 0.74 in week-1 and gradually decreasing to its lowest value in week-6 at r = 0.50. Although the highest correlation comes from the combined climate indices model, the biggest improvement of the correlation coefficients are experienced by models developed using the QBO, MJO-OMI 1 and MJO-OMI 2. With the addition of the persistence term, the correlation coefficients for these three models are relatively higher than the correlations before the inclusion of persistence term at all lead weeks.

We repeated the same calculation for 2-metre surface temperature. Prior to the addition of the persistence term, all models exhibit low correlations between the forecasts and the observations. Forecasts made by multiple indices model and ENSO record modest correlations while MJO-OMI 1, MJO-OMI 2 and QBO have weak and negative correlations. However, when the persistence is included into the models, it enhances every model correlation at all lead times. All models share similar highest correlation which occurs in week-1 and steadily levelling off until the end of week-6. The overlapping of the correlation values in Figure 5 (b) suggests that the persistence has a substantial effect on temperature prediction over the Maritime Continent and the persistence tends to persist for few weeks. This characteristic could justify the high correlations through all six weeks when persistence is added to the models.

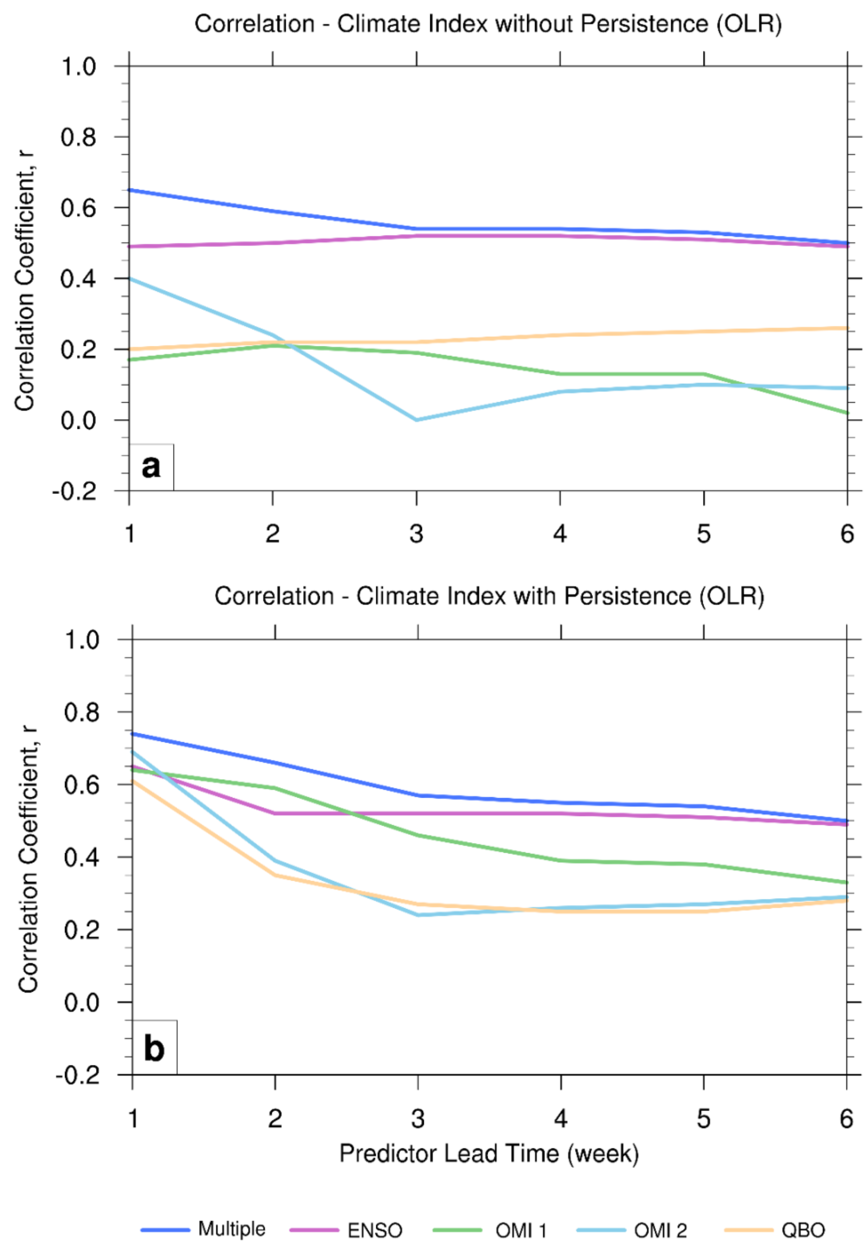

Figure 4 Correlation coefficients calculated in the validation period, 2006 – 2012. (a) Except for the blue line (combined climate indices), all other correlation coefficients are obtained using one climate index as the predictor. (b) Same as (a) but with the inclusion of the persistence term. The predictand use for (a) and (b) is the OLR.

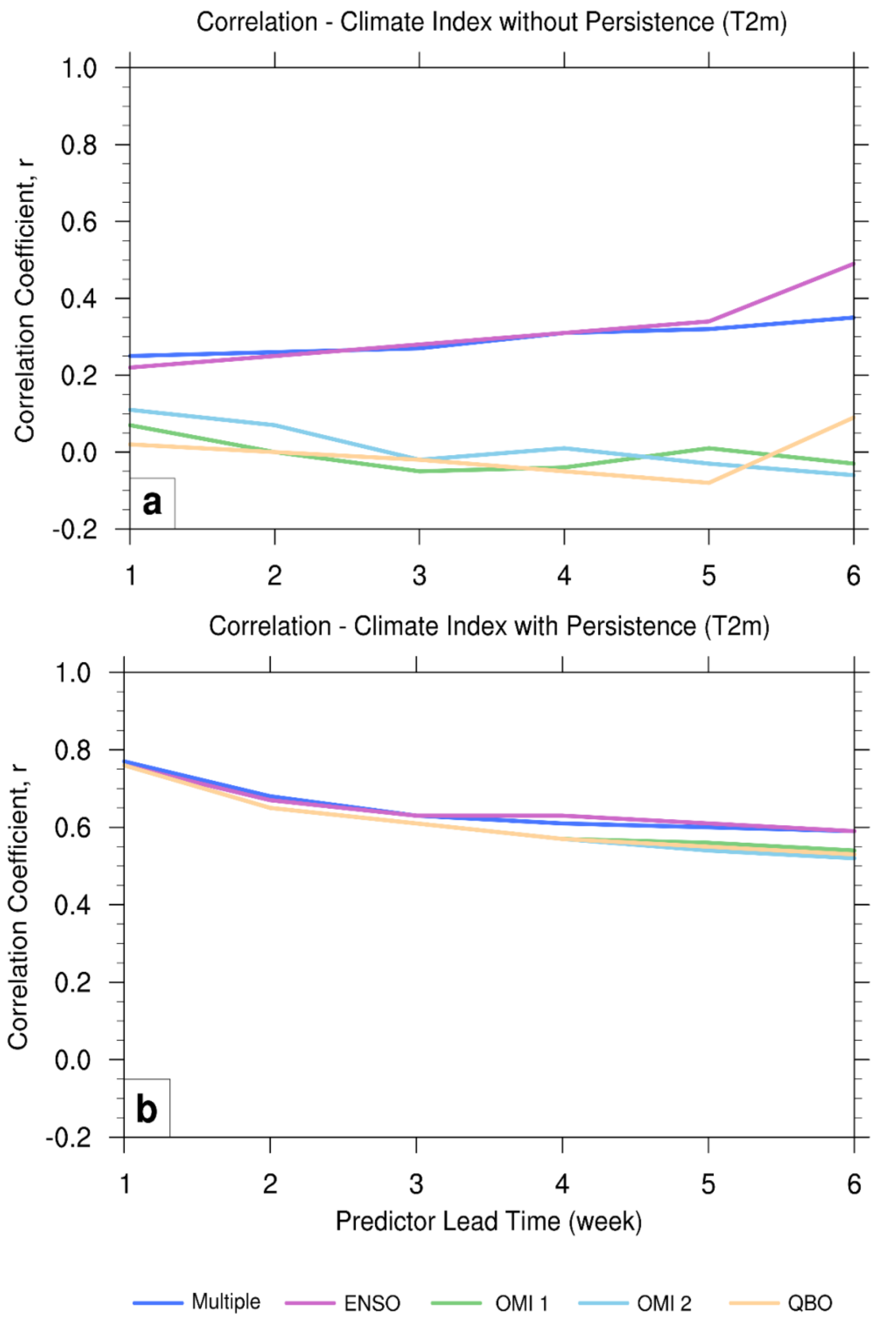

Figure 5 As in Figure 7 but the predictand use for (a) and (b) is the T2m.

Supported by the good performance shown by the model that uses all climate indices and persistence term as predictors for OLR and T2m on the annual time scale, we proceed the study by taking seasonality into account. Figure 6 implies that the persistence assists the increment of the correlations for OLR in all the seasons except for DJF starting at lead week-3 until the end of the lead week. Overall, the correlations with the absence of the persistence term are weaker than the correlations with the persistence. Highest correlations for the case with persistence are visible in SON and DJF. In SON, its correlations invariably surpass the correlations of the model without the persistence. For DJF, the model with persistence has higher correlations for the first two lead weeks, declining to its lowest r = 0.58 at lead week-4 before slowly rising until week-6. Model with persistence shows modest correlations for JJA while MAM experienced a declining trend throughout the lead times.

The variation of the Maritime Continent OLR probably due to its intimate linked with the life cycle of ENSO. The ENSO will normally start to develop in JJA, continuing to evolve during SON until it eventually peaks in DJF. The close relationship between the OLR and ENSO is clearly shown in Figure 6 where the seasonal correlations are strong in JJA, SON and DJF. This finding is supported by Tangang and Juneng (2004) which also acknowledged that the ENSO has good correlations with rainfall during SON and DJF. Meanwhile, in MAM, the ENSO is usually experiencing its decaying phase, hence the correlation between the OLR and the ENSO is weak during this season. As the strength of the ENSO is decreasing, the role of the local factors such as the land-sea breeze and the orographic convection begins to be more dominant. Not to mention, MAM is an inter-monsoon season in Malaysia where it is generally dry which results in a low rainfall variance. Therefore, the predictability for the OLR in MAM is small and it is portrayed in poor correlation values in seasonal correlation in MAM.

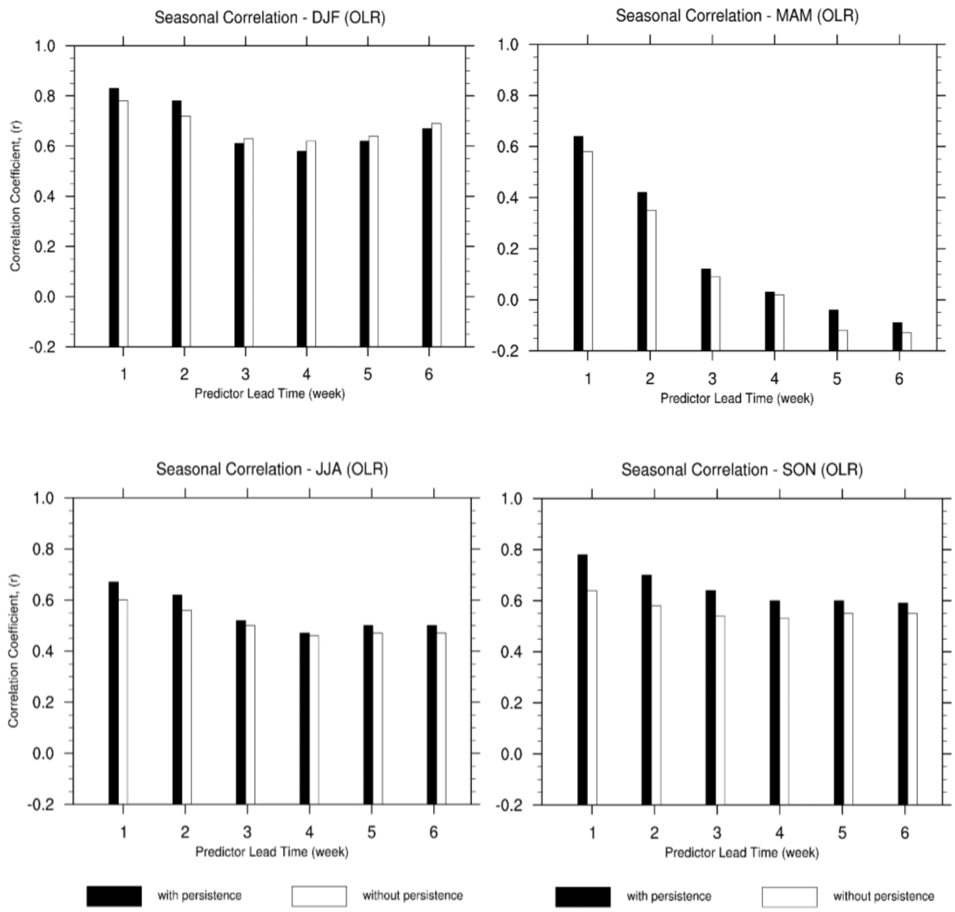

Figure 6 Seasonal correlation coefficients for six different weekly lead times using OLR as the predictand calculated in the validation period (2006 – 2012). The solid bar is the correlation coefficient obtained with model that include the persistence term. Empty bar is the correlation coefficients generated by model without the persistence term.

From Figure 7, we can see the influence of seasonality on the correlations when the surface temperature is taken as the predictand. Seasonal correlations, especially for DJF and MAM, are generally higher than its annual correlation at all lead times, in the case where persistence is not included. Consistent with Figure 6, the influence of the persistence term is visible through all the seasons. Except for DJF and MAM, the present of the persistence in the model provide correlations that are constantly greater than the model without persistence. Model in MAM and JJA work particularly well for T2m. Using the Student's t-test, correlations for MAM are significant at 0.10 significance level for all lead times. Also, correlations for JJA and DJF are significant at the same significance level up to lead week-3 and week-2 respectively (not shown).

The correlation magnitudes for temperature in this current study are in resemblance to the work that discussed the trend of temperature variability from Tangang et al. (2007). Through their seasonal EOF analysis, it is found that the largest explained variance (first mode) of the temperature takes place in winter. The amplitude of the power spectrum of this first mode peaks at two to five year range, suggesting that this particular mode is governed by interannual variability such as ENSO. Their study classified the seasons as follow October-November-December (OND) and January-February-March (JFM) as the early and late stages of winter while April-May-June (AMJ) and July-August-September (JAS) are considered as the early and late stages of the summer monsoon in Malaysia.

Concisely, the correlation between the anomalous surface temperature and El-Niño events in their study exhibit weak correlation during JAS but it continues to gain strength as the El-Niño continue to mature in the following months. The highest correlation is recorded during JFM. Similar results are reflected in our model that excludes the persistence where the correlations are weak in JJA and SON. They postulated that these low correlations could be because it is the early stage of El-Niño development. During the early development of the El-Niño, the temperature is less influenced by the event. Furthermore, higher correlations in DJF and MAM, correspond to OND and JFM in Tangang et al. (2007) are said to be related to the peak of the El-Niño and the continuous warming in the western North Pacific.

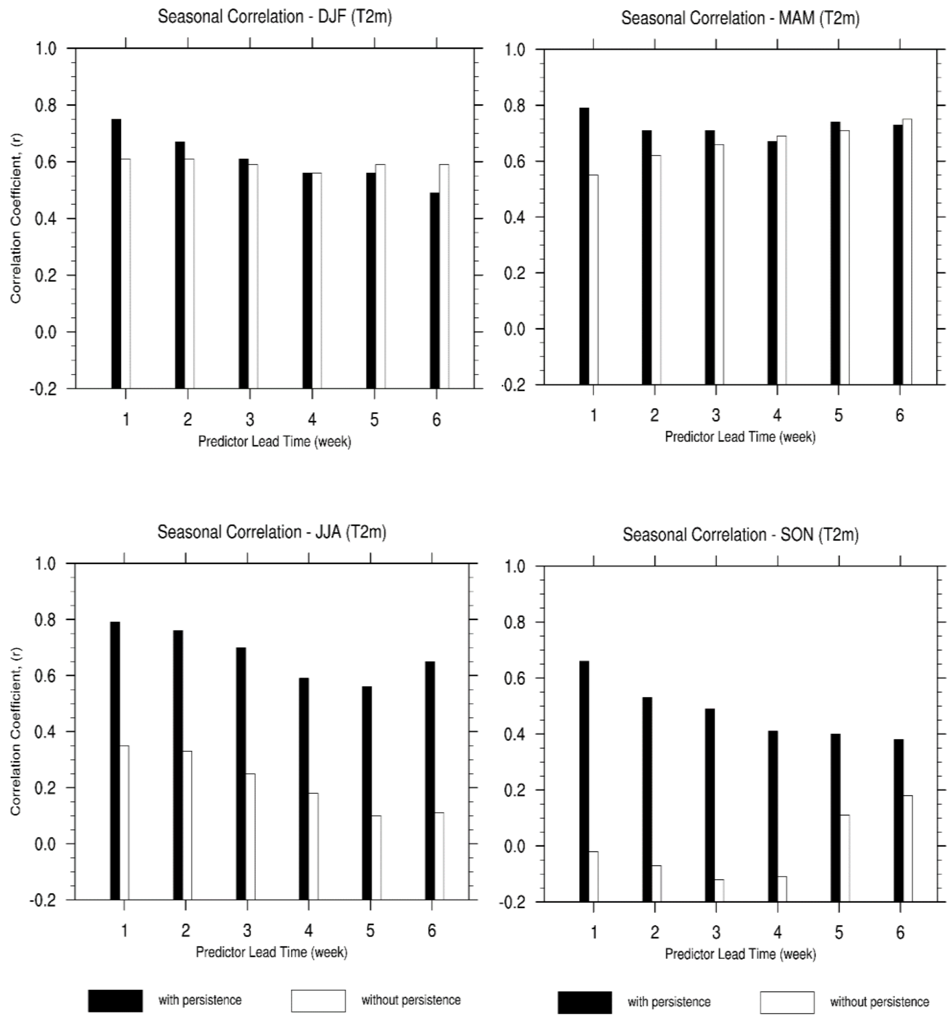

Figure 7 Same as in Figure 6 but using T2m as the predictand.

It is hard to definitively pinpoint the individual influence of the QBO and the MJO or even for the ENSO, purely depending on the season alone. Fortunately, the ENSO tends to be phase-locked in winter and normally persists for the duration of two to five years. Hence, we could to some extent, deduce the results by linking it to the ENSO. Whereas this is not achievable for the QBO and the MJO. Previously, Collimore et al. (2003) study reveal that deeper convective clouds are likely to develop during the EQBO phase. In addition, Baldwin et al. (2001) stated that the phase of the QBO tends to transition from one phase to another during late spring and summer, i.e WQBO to EQBO and vice versa. Yet, without a proper separation of

EQBO and WQBO, that information are rather not useful. Likewise, alluded to earlier studies, MJO too has strong phase dependence. Phase 4-6 has a higher impact on Maritime Continent rainfall and phase 3, 4, 7 and 8 impose influence in Canadian wintertime surface air temperature (SAT).

4.2 RMSE Skill Score

As briefly mentioned in preceding sections, RMSE is the measure of accuracy. In other words, it could also be called as the measure of error as it calculates forecasts deviation from the observation values. Smaller RMSE implies smaller forecast errors. The benefit of using RMSE rather the MSE is that the RMSE will maintain the same unit as the variable which later will aid the comparison of the magnitude of the forecast error. Since we are using the climatology as the reference forecast in this present study, we are aiming for positive RMSESS for our forecast model to have better performance than the climatology.

Figure 8 shows the RMSESS for forecast model with and without persistence term using OLR as the predictand. The solid bar represents the skill score produced by the model which include persistence term while the empty (white bar) for the model without persistence. When we exclude the persistence term from the model (white bar), we can see that the model does have some skill over the climatology as they are able to yield positive RMSESS. This is especially true for DJF as it gives RMSESS = 0.43 at the first lead week and maintains a good score until the end of the lead week with RMSESS = 0.23. RMSESS is weak for JJA and SON with skill score lingering at 0.15 to 0.10 but nonetheless signify a better performance than climatology. The worst score for this case comes from MAM. Even though the score at the first week is 0.20, it experiences a rapid decline throughout the lead week. By lead week 3, the forecast generated is equal to the climatology and for the rest of the week, the scores are negative implying poor forecast performance.

In the occasion where the persistence is included (solid bar), the RMSESS depict a considerable increase from the case of without persistence. In most cases, the highest score can be seen at week-1 and steadily decline by week-6. Overall, there is skill improvement in all the seasons over the whole time range with skill score advantage up to 0.20, with the exception of DJF. We see that in DJF, the model with persistence has higher scores in the first two lead weeks with an increase of 0.06 and 0.05 respectively from its counterpart. However, from lead week 3 onwards, the RMSESS of the model without persistence surpass the RMSESS of the model with persistence. This result suggests that the better forecast for week 3 ahead could be obtained even without taking the persistence term into account.

Again, we repeat the same calculation of the RMSESS for the surface temperature for both models and the result is demonstrated as in Figure 9. Different from the OLR that indicate that model without persistence does have the potential to be used to generate seasonal forecasts, the same implication could not be applied to the case of T2m. Referring to the empty bar (white bar), except in MAM which has skill score ranging from 0.09 to 0.16 in all weekly lead, the scores for the other seasons are mainly equal to the climatology or at some cases the scores reflect that the model has poorer skill as to when the climatology are used to make the prediction. Without persistence, the worst score is seen in DJF where the score in all lead week is negative.

Compared to the model without persistence, the RMSESS of the model with persistence is significantly higher. Over this period of six lead weeks, the pattern of the RMSESS is uneven across the four seasons. Skill score in DJF and SON show a decreasing pattern, while the other two seasons show fluctuation in the skill scores. The most dramatic increase of the RMSESS is in DJF, where the score is enhanced more than twice than that of without persistence. Other seasons also indicate the promising performance of this particular model.

This result supports the practicality of including the persistence term, as it has successfully proved that it helps boost the skill score in both OLR and T2m.

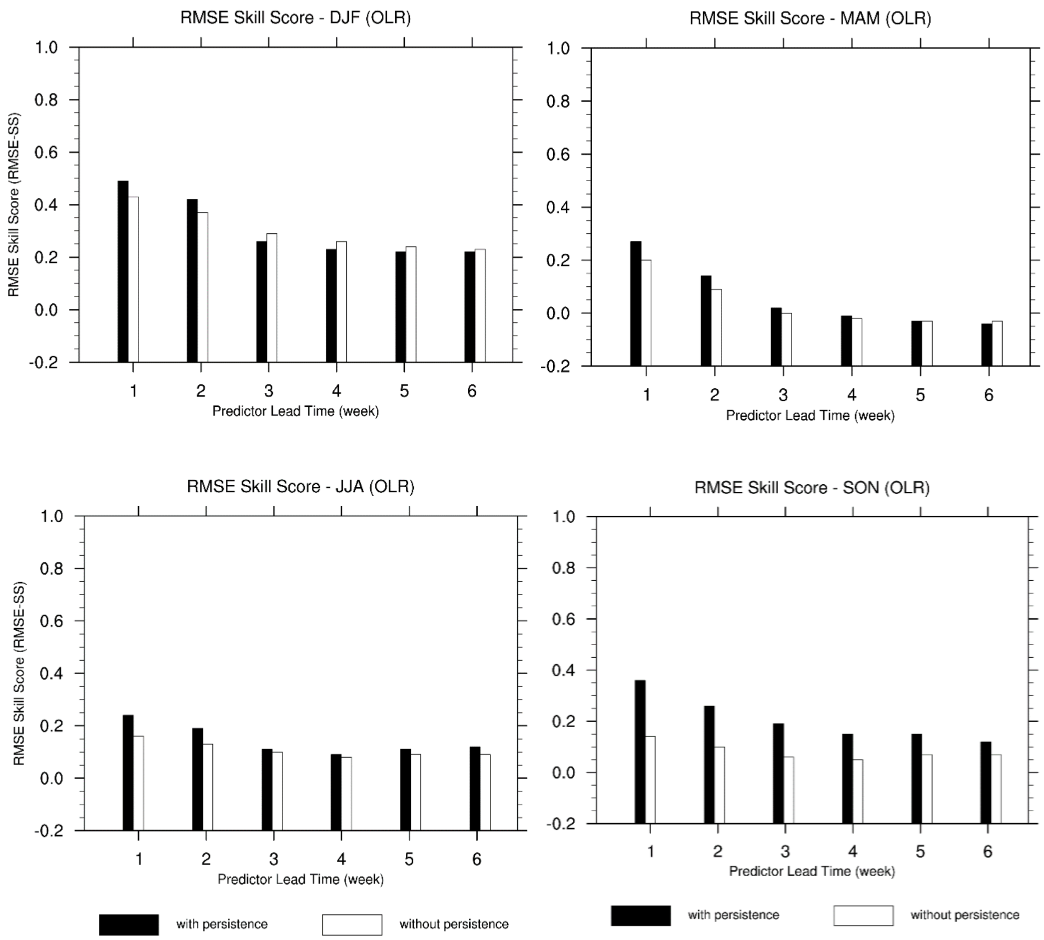

Figure 8 Seasonal RMSE skill score for six different weekly lead times using OLR as the predictand calculated in the validation period (2006 – 2012). The solid bar is the skill score obtained with model that include the persistence term. Empty bar is the skill score generated by model without the persistence term.

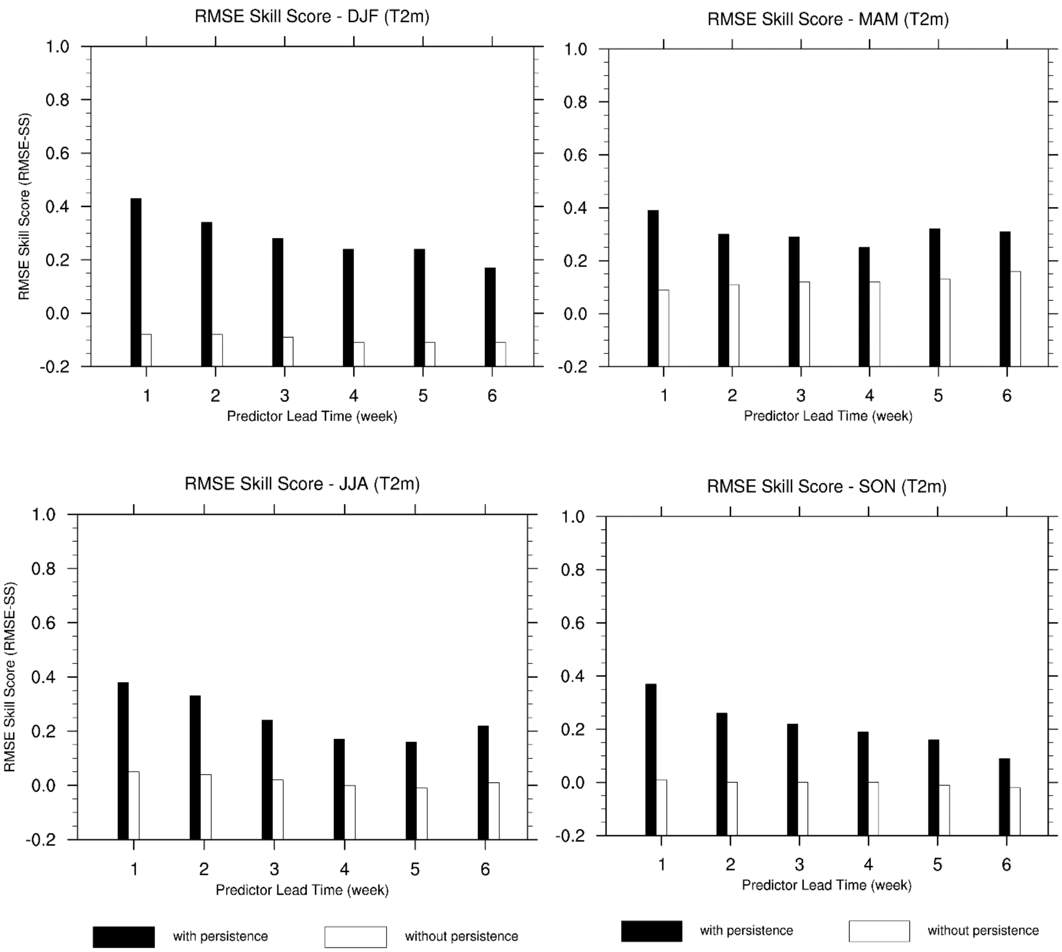

Figure 9 Same as in Figure 8 but using T2m as the predictand.

5. Conclusions

This present study used simple and multiple regression methods to forecast the OLR and surface air temperature over the Maritime Continent using the possible predictability provided by main climate indices, namely the ENSO, the QBO and the MJO. The selection of the predictors are based on the successfulness of preceding works that had utilized the predictors either individually or by taking the combination of them. The works employing the pure statistical model over Maritime Continent is rather limited, thus motivated by this, we are interested to investigate the possible prediction opportunity given the information by the ENSO, QBO, MJO as well as the persistence term.

At the first part of the study, we constructed regression model using single predictor to assess their respective skill. For both predictands, the model developed using the ENSO has shown the best and consistent performance than the other three predictors. Despite the good performance from the ENSO model, the performance from the model with a combination of all available predictors outdo the ENSO model. Due to this, we proceed the next part of the study using the combination model instead of the simple model. The evaluation of the model is done by calculating the correlation coefficient and RMSE skill score on the annual and seasonal time scale.

Throughout the study, the persistence term (autocorrelation) show that it has its own distinct signature. In a way, it facilitates the correlation and skill scores values in most cases discussed above. The persistence term helps the model in the sense where it takes account of the current conditions of a certain variable and subsequently provides that information for the forecast calculation. However, the persistence is less effective to be used in DJF for lead above week-3 as evident in the correlations for OLR and T2M. Perhaps, the existence of the occasional cold surges from the Siberian High disturbs the efficiency of the persistence term during the boreal winter. Moreover, as the persistence is defined as the degree of similarity of a past value of variables to their future values, it inevitably decreases as the number of lag increases.

As has been noted, better results could be attained if specific cases, such as strong El-Niño and MJO events and the particular phase of the MJO (phase 3, 4 and 5) and QBO are considered in this study. We optimistically believe that adding specific phase of the MJO could reduce the residual of our multiple regression models (not shown). Nonetheless, even without this definitive conditions, the inclusion of the MJO components, having the ability to supply prediction skill of 15 -20 days, not to forget the influence of QBO, in our multivariable regression model is justified with the improvement of the annual and seasonal correlations and RMSE skill scores. However, whether or not this model has any practicality to be integrated into operational weather forecasting requires extra justifications with an appreciable amount of refinement and adjustment in the future work.

REFERENCES

Baldwin, M. P., and Coauthors, 2001: The quasi-biennial oscillation. *Reviews of Geophysics*, **39,** 179-229.

Collimore, C. C., D. W. Martin, M. H. Hitchman, A. Huesmann, and D. E. Waliser, 2003: On the relationship between the QBO and tropical deep convection. *Journal of Climate*, **16,** 2552-2568.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137,** 553-597.

Fox, J., 1997: *Applied Regression Analysis, Linear Models, and Related Methods.* SAGE Publications. **597** pages.

Haylock, M., and J. McBride, 2001: Spatial coherence and predictability of Indonesian wet season rainfall. *Journal of Climate*, **14,** 3882-3887.

Hidayat, R., and S. Kizu, 2010: Influence of the Madden–Julian Oscillation on Indonesian rainfall variability in austral summer. *International Journal of Climatology*, **30,** 1816-1825.

Holton, J. R., and R. S. Lindzen, 1972: An updated theory for the quasi-biennial cycle of the tropical stratosphere. *Journal of the Atmospheric Sciences*, **29,** 1076-1080.

Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2014: Skillful Wintertime North American Temperature Forecasts out to 4 Weeks Based on the State of ENSO and the MJO. *Weather and Forecasting*, **29,** 23-38.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* J. Wiley, **231** pages.

Juneng, L., and Tangang, F.T. 2005: Evolution of ENSO-related rainfall anomalies in Southeast Asia region and its relationship with atmosphere-ocean variations in Indo-Pacific sector. *Climate Dynamics,* **25:** 337. doi:10.1007/s00382-005-00316.

Kiladis, G. N., and Coauthors, 2014: A Comparison of OLR and Circulation-Based Indices for Tracking the MJO. *Monthly Weather Review*, **142,** 1697-1715.

Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, **77**, 1275–1277.

Malaysian Ministry of Health, 2016: Heat exhaustion and heat stroke cases in Malaysia. Press Statement. Retrieved from www.moh.gov.my

Ramage, C. S., 1968: Role of a tropical "Maritime Continent" in the atmospheric circulation. *Monthly Weather Review*, **96,** 365-370.

Rayner, N., and Coauthors, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, **108:** 4407, doi: 1.1029/2002JD002670.

Rodney, M., H. Lin, and J. Derome, 2013: Subseasonal prediction of wintertime North American surface air temperature during strong MJO Events. *Monthly Weather Review*, **141,** 2897-2909.

Tangang, F. T., and L. Juneng, 2004: Mechanisms of Malaysian Rainfall Anomalies. *Journal of Climate*, **17,** 3616-3622.

Tangang, F. T., L. Juneng, and S. Ahmad, 2007: Trend and interannual variability of temperature in Malaysia: 1961–2002. *Theoretical and Applied Climatology*, **89,** 127-141.

Tangang, F. T., and Coauthors, 2008: On the roles of the northeast cold surge, the Borneo vortex, the Madden-Julian Oscillation, and the Indian Ocean Dipole during the extreme 2006/2007 flood in southern Peninsular Malaysia. *Geophysical Research Letters*, **35:** L14S07, doi: 10.1029/2008GL033429.

Taylor J. R., 1997: *An introduction to error analysis. The study of uncertainties in physical measurements*. Second edition. University Science Books. **327** pages.

Wheeler, M. C., and H. H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, **132,** 1917-1932.

Wheeler, M. C., and H. H. Hendon, 2004: Impacts of the Madden-Julian Oscillation on Australian rainfall and circulation. *Journal of Climate,* **22,** 1484.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* Elsevier, **676** pages.

Xavier, P., R. Rahmat, W. K. Cheong, and E. Wallace, 2014: Influence of Madden-Julian Oscillation on Southeast Asia rainfall extremes: Observations and predictability. *Geophysical Research Letters*, **41,** 4406-4412.

Yao, W., H. Lin, and J. Derome, 2011: Submonthly forecasting of winter surface air temperature in North America based on organized tropical convection. *Atmosphere-Ocean*, **49,** 51-60.

Yim, S.-Y., B. Wang, W. Xing, and M.-M. Lu, 2014: Prediction of Meiyu rainfall in Taiwan by multi-lead physical–empirical models. *Climate Dynamics*, **44,** 3033-3042.

Zhang, C., 2005: Madden-Julian oscillation. *Reviews of Geophysics,* **43,** RG2003, doi: 10.1029/2004RG000158.