Level and source of predictability of seasonal rainfall anomalies in Malaysia using canonical correlation analysis

Liew Juneng and Fredolin T. Tangang*

Marine Science Program, School of Environmental and Natural Resource Sciences, Faculty of Science and Technology, National University of Malaysia, Malaysia

ABSTRACT: This study examines the level and origin of seasonal forecast skills of rainfall anomalies in Malaysia. The forecast models are based on an empirical technique known as the canonical correlation analysis (CCA). The CCA technique searches for maximally correlated coupled patterns between sets of predictor and predictand matrices. The predictive skills of five predictor fields, namely station rainfall in preceding seasons (i.e. the predictand itself), local sea surface temperature (SST) over the western Pacific sector, quasi-global SST, sea level pressure, and northern hemisphere 700 hPa geopotential height, are investigated. The sequence of four consecutive 3-month predictor periods is used to capture evolutionary features in the predictor fields. The predictor-predictand setup is designed such that the predictor fields are followed by a lead-time ranging from 0 to 9 months and then by a single predictand period of 3 months' duration. The skills are estimated in hindcast mode using the one-year-out version of the cross-validation technique. Skill estimates are expressed as temporal correlation coefficients between forecasts and observed values.

A series of experiments with different predictor combinations reveal that the model with quasi-global SST alone produces most favourable forecast skills. The forecast skills of this model generally outperform the persistence forecast. Moreover, the model also has higher forecast skills in the East Malaysian region compared to those in Peninsular Malaysia. The forecast skill peaks during the late Northern Hemisphere winter season (January–February–March (JFM)) with a secondary maximum during the early summer season (May–June–July, (MJJ)). The average forecast skills are between 0.3–0.5 for up to 5 months' lead-time in the East Malaysian region and this can be considered very useful for the appropriate users. In the Peninsular Malaysia region, the forecast skills are generally weak, although some stations registered modest skills even at a 5-month lead-time. For both prediction periods, the source of predictability originates from anomalous SST conditions associated with the El Niño-Southern Oscillation (ENSO) phenomenon. Generally, during a La Niña (an El Niño) event, regions in northern Borneo experience excess (deficit) rainfall during the JFM period. Similar conditions are experienced during the MJJ period except that the impact tends to be more widespread throughout the country. Interestingly, the origin of predictability during the JFM period can be traced to typical ENSO events, while ENSO events of longer duration are responsible for the MJJ period. Copyright © 2007 Royal Meteorological Society

KEY WORDS Malaysian rainfall; predictability; canonical correlation analysis; ENSO; Indian Ocean dipole

Received 11 November 2006; Revised 27 June 2007; Accepted 23 July 2007

1. Introduction

Large-scale climate events, such as the El Niño-Southern Oscillation (ENSO), exert an appreciable impact on the socio-economic well being of Malaysia. A prolonged drought associated with an El Niño frequently causes severe water supply crises, disrupts agricultural activities, and destroys rain-fed crops. It also induces various other environmental hazards such as forest fires and haze episodes (e.g. Nichol, 1997, 1998; Gutman et al., 2000; Page et al., 2002; Matsueda et al., 2002; Fuller et al., 2004). During severe haze episodes, other related problems emerge such as a dramatic increase in the number of people suffering from various respiratory illnesses due to deteriorating air quality; huge decreases in tourist arrivals; flight cancellations; and increased risk of ship collisions in the Straits of Malacca due to poor visibility. All these impacts can be translated into losses that amount to hundreds of millions of dollars. Equally damaging is the occurrence of severe floods in association with La Niña. During periods of intense winter monsoon, the occurrence of widespread flood in low-lying areas, particularly in the east coast of Peninsular Malaysia, often results in massive evacuation and destruction of public and private properties, damage to crops, and even leads to the loss of lives. Hence, the importance and need for accurate and useful climate prediction schemes are clear. The skillful forecast of seasonal climate with appreciable lead-time provides useful information for decision makers to better manage resources, establish mitigation plans, and enhance response strategies related to flood and drought disasters.

The Malaysian climate is basically dominated by the Southeast Asian monsoon with a cycle of two opposite
regimes, i.e. the summer monsoon [locally known as the southwest monsoon (SWM)] and the winter monsoon [locally known as the northeast monsoon (NEM)]. The SWM commences in late May and ends in late September, while the NEM arrives around November and retreats in March of the following year. The country receives substantial rainfall year round but the amount peaks during the NEM period. Despite its cyclic nature, there exists significant variability of the monsoon rainfall on interannual time scales (Tangang, 2001; Tangang and Juneng, 2004; Juneng and Tangang, 2005). Floods associated with anomalously intense monsoon seasons and drought associated with delayed or failed monsoon are recurrent events in Malaysia. In the Indo-Pacific region, through the disruption of general circulation patterns, ENSO is considered the strongest climate modulator (Ropelewski and Halpert, 1987). The ENSO–Malaysian rainfall relationship has been demonstrated in numerous studies (e.g. Tangang, 2001; Tangang and Juneng, 2004; Juneng and Tangang, 2005). These studies indicate that a significant proportion of anomalous rainfall variability in Malaysia can be attributed to variation associated with ENSO. Hence, the variability of Malaysian rainfall should have appreciable predictability based on ENSO evolution.

In this paper, we investigate this issue by employing a linear statistical modelling technique called canonical correlation analysis (CCA). The success of using the CCA utilizing the signature of ENSO evolution as predictors in predicting anomalous rainfall in various regions has been reported in various studies (e.g. Barnston, 1994; He and Barnston, 1996; Shabbar and Barnston, 1996; Yu et al., 1997; Johansson et al., 1998). However, despite ENSO’s expected influence on anomalous rainfall in Malaysia, the issue of its predictability has yet to be investigated. Hence, the aim of this paper is to investigate the origin and source of predictability of seasonal rainfall anomalies in Malaysia based on the CCA technique, and subsequently use the results as a basis towards the establishment of an operational long-lead forecast scheme for anomalous rainfall in the region.

2. Data and methods

2.1. Predictand
Seasonal rainfall, taken as the running average of three consecutive months’ total value, is used as the predictand variable. Precipitation data, provided by the Malaysian Meteorological Department (MMD), spans a period of about 50 years (1951–2000). A total of 14 meteorological stations (Figure 1) were considered in this study. Eight of the stations are located in Peninsular Malaysia, while the rest are in East Malaysia (part of northern Borneo).

2.2. Potential predictor candidates
In order to search for an optimum model, the predictive skills of several models using a single predictor variable or combination of several potential predictors are investigated. These predictor variables include the quasi-global sea surface temperature (QSST), regional sea surface temperature (RSST), Northern Hemisphere 700-hPa geopotential heights (H700), as well as the quasi-global sea level pressure (QSLP) and the rainfall itself in the preceding seasons (RAINFALL). Both QSST and QSLP cover the same region i.e. from latitude 45.5°S–45.5°N and longitude 24.5°E–75.5°W, while the RSST extends from latitude 9.5°S–20.5°N and longitude 89.5°–139.5°E. The planetary scale QSST, QSLP, and H700 may be taken as predictors that describe, among other things, ENSO evolution. Since anomalous rainfall in Malaysia is largely modulated by ENSO (Tangang, 2001; Tangang and Juneng, 2004; Juneng and Tangang, 2005), these predictors may provide the predictive skills associated with the phenomenon. These predictor variables may also carry signatures of other phenomenon such as the Indian Ocean dipole (IOD). Several studies indicate a strong influence of IOD events on rainfall and temperature fluctuations in the surrounding region (e.g. Saji and Yamagata, 2003; Ashok et al., 2001, 2004). In Malaysia, surface air temperature fluctuations appear to be partially modulated by IOD events (Tangang et al., 2007). Both QSST and H700 have been used successfully by He and Barnston (1996) for predicting seasonal rainfall variation at tropical Pacific island stations in the equatorial Pacific. The QSLP is usually used to represent the atmospheric component of ENSO and is well known to be highly skillful in predicting sea surface temperature (SST) anomalies associated with ENSO (Barnston and Ropelewski, 1992; Tangang et al., 1997, 1998). The RSST is also included in the set of possible predictors. This is because the regional atmosphere–ocean interaction in the western North Pacific region plays an active role in modulating the anomalous rainfall in the southern Philippines and northern Borneo during winter (Tangang and Juneng, 2004; Juneng and Tangang, 2005). However, since this regional atmosphere–ocean interaction is phase-locked to boreal winter seasons, its evolution may not be as relevant in predicting anomalous rainfall for other seasons. On the other hand, the inclusion of RSST
may provide additional regional information, which may be excluded from the heavy spatially smoothed QSST. The last predictor variable is the predictand itself (seasonal rainfall) from the preceding four seasons. The use of preceding rainfall as predictor may incorporate some useful autocorrelation information in the model (He and Barnston, 1996). For the combination of predictors, the inter-field weighting technique (Barnston, 1994) is used.

2.3. Data pre-processing

QSPL and H700 data is taken from the reanalyis product of NCEP/NCAR and come in 2.5° × 2.5° grids. QSST is a blend of Version 2 of NOAA optimum interpolated sea surface temperature (OISST) (Reynolds and Smith, 1994) and Version 1.1 of the Hadley Centre Global Ice and Sea Surface Temperature (HADiSST) (Rayner et al., 2003). The OISST is available in 2.0° × 2.0° grids from the Climate Diagnostic Centre (CDC), while the HADiSST is provided by the United Kingdom Meteorological Office (UKMO) in 1.0° × 1.0° grids. All QSPL, H700, and OISST data are available at near real time from the CDC website and are suitable for operational forecasting. Nevertheless, OISST dates back to only December 1981 and hence QSST data prior to this date is taken from HADiSST and blended together using the extended empirical orthogonal function (EOF) technique described in Smith et al. (1996). QSST is then spatially averaged to a 10° × 10° grid, while RSST remains in 2.0° × 2.0° resolution grids.

In searching for the optimal predictive model, the CCA technique searches for the optimum linear combination between the matrices of the predictand and predictors that produces the maximum correlation. The essential equations used in this study are those used in Barnett and Preisendorfer (1987). Extensive mathematical basis of CCA can also be found in Graham et al. (1987). To capture the predictive signatures associated with ENSO evolution, each predictor variable is stacked sequentially according to season over a total period of 1 year. Standardization and inter-field weighting were performed prior to subsequent analyses. The stacking of predictor variables introduces a large number of highly correlated time series in the predictor fields and hence pre-orthogonalisation using standard EOF analysis is employed. An EOF mode is included in the model if its variance is significant at at least 95% level. The significance level is estimated using the Monte Carlo technique (e.g. Preisendorfer and Barnett, 1977; Overland and Preisendorfer, 1982). Depending on the season, for both predictor and predictand fields the retained EOF modes account for between 60 and 75% of total variance. The entire procedure follows the sequence highlighted in Barnston (1994; Figure 2).

2.4. Forecast validation

Artificial skills associated with overfitting of random variations are a main problem of all empirically based prediction schemes. Models with large artificial skills tend to perform much better in training sets but relatively more poorly in forecast periods. To avoid or minimize over-fitting, an exhaustive version of cross-validation (Michaelcens, 1987) procedures is employed to estimate the model skill. In this technique, each of the 50 years from 1950 to 2000 is held out as a forecast target, while the rest of the data sequences are used for model fitting (e.g. Barnston and Ropelewski, 1992; Barnston, 1994; He and Barnston, 1996; Yu et al., 1997). The withdrawn year plays no part in any of the pre-processing procedures. All the pre-processing steps (standardization, inter-field weighting, and EOF analysis) are then applied repeatedly on the training set to fit the model. The withdrawn predictor field is then projected onto the resultant CCA predictor loading patterns and the corresponding forecasts are generated and verified against the observed data of the withheld year. This procedure is repeated until each year is held out for forecast validation.

In this study, the cross-validated correlation coefficients between forecast and observed values at each station are used as a statistical measurement for model skill evaluation. Additionally, the station’s correlation skills are averaged into two regions, one averaged over East Malaysia and the other over Peninsular Malaysia. Juneng and Tangang (2005) indicated that the two regions appear to have different anomalous rainfall patterns associated with ENSO events during the NEM period and hence the regions may have different responses to large-scale forcings.

2.5. Lead-time

To ensure the usefulness of the prediction, a sufficient and appropriate lead-time must be placed between the predictor and the predictand period. The lead-time is defined as the time skip between the last month of the predictor season and the first month of the predictand season (Figure 2). In this experiment, the forecast skills for 0- to 9-month lead-times are evaluated.

3. Results and discussion

3.1. Persistence forecast

Persistence forecasts are often used as benchmark forecasts to which the CCA forecast skills levels are compared. In the persistence forecast, the predictand itself is used as the predictor field, but only the data from the fourth predictor season (i.e. the most recent predictor season) is used. In comparison, the seasonal rainfall predictor

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Target</th>
<th>Lead Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAS OND (JFMA)</td>
<td>JFM</td>
<td>0-month</td>
</tr>
<tr>
<td>JAS OND (JFMA)</td>
<td>JFM</td>
<td>3-month</td>
</tr>
<tr>
<td>(JAS OND)</td>
<td>JFM</td>
<td>6-month</td>
</tr>
</tbody>
</table>

Figure 2. Schematic representation of the timing of the predictand–predictor periods for 0, 3, and 6 months lead-time forecasts.
consists of the predictand data during the preceding four predictor seasons. In both cases, all time series are normalized prior to further processing. For direct comparison and consistency, the persistence forecast skill is also evaluated using the leave-one-out cross-validation scheme. Figure 3 shows the lead-time – forecast season cross-section of the cross-validated persistence skill averaged for the two different regions of East Malaysia and Peninsular Malaysia. The persistence forecast skills for the East Malaysian region develop during the transitional season of Aug–Sept–Oct (ASO) and peak at \( \sim 0.35 \) during the early NEM period for shorter lead-times (1–2 months). During the late NEM (January–April), averaged persistence skills reach \( \sim 0.30 \) at 3- to 4-month lead times. On the basis of the lead-time definition and the generally tilted skill pattern in Figure 3, it is recognized that the persistence forecast skill is largely contributed by rainfall anomalies during the transitional period of around August–October. Rainfall anomalies during the late NEM period do not contribute to the predictability of the rainfall of the subsequent seasons as there is apparently no notable persistence skill for the SWM season. On the other hand, it is clear that there is no significant predictability based on persistence forecast in Peninsular Malaysia for any seasons or lead-times. This suggests that there is a significant difference in the anomalous rainfall behaviour over the two regions. Overall, it can be concluded that the persistence model is inadequate as an operational forecast model of anomalous rainfalls over the two regions.

3.2. Experiments with single predictor fields

Five models using a single predictor variable were considered and compared with one another. These include the preceding four seasons’ rainfall anomalies, QSST, LSST, QSLP, and H700. To facilitate easy interpretation and comparison among the models, only the forecast skill performed at 1-, 5-, and 9-month lead-times that correspond to short-lead, medium-lead, and long-lead forecasts are shown (Figure 4). Also shown in the figure is the persistence forecast skills to be used as a reference level for other models. Figure 4 clearly depicts the seasonal variations of the forecast skill, individually for the two portions of Malaysia. Generally, the cross-validated forecast skills are relatively higher in the East Malaysian region, particularly during the NEM period for both short- and medium-lead forecasts. In Peninsular Malaysia, the skill scores are relatively higher during the late NEM and the following monsoon transitional period (February–May). For both regions, visual comparison suggests that models with any predictor (except rainfall) outperform persistence forecasts for most target seasons. However, consistent with the persistence skills in Figure 3, for a 1-month lead-time prediction in the East Malaysian region, persistence forecasts can be a tough competitor for other models during the early NEM period. The skills of rainfall predictor and persistence appear to be comparable, although in some periods the rainfall predictor appears to be slightly more skillful. This implies that the inclusion of additional rainfall information for three seasons prior to the fourth season does not significantly increase forecast skills. He and Barnston (1996) showed that, for islands in the tropical Pacific, rainfall predictors can be much better than persistence forecasts particularly for islands in the North Pacific region during a period of February to May for 1-month lead-time prediction (Figure 2(a)). However, for islands in the South Pacific, persistence skills and that of rainfall predictors are comparable (their Figure 2(c) and (d)). Hence, the performance of rainfall predictors depends on region, season, as well as lead-time.

Overall, the QSST and QSLP models produce the most skillful prediction throughout the target seasons, although the skill difference becomes more apparent as lead-time increases. Generally, the H700 and RSST models produce lower skills than that of the QSST or QSLP, although in certain seasons the skills are higher. For H700, this is consistent with those reported by He and Barnston (1996). In the East Malaysian region, QSST skills peak during DJF with values exceeding 0.40 for both 1-month and 5-month lead-times. A secondary peak with an average skill of \( \sim 0.30 \) is apparent during the relatively dry period of June–July–August (JJA) for a 1-month lead-time and during May to June for 5- and 9-month
lead-times. A minimum skill score is attained during the boreal spring and this is consistent with the spring predictability barrier problem as noted by other authors (e.g. He and Barnston, 1996; Yu et al., 1997). This skill seasonality appears to be consistent with those reported in Yu et al. (1997). Interestingly, the skills in East Malaysia are less sensitive to increasing lead-time, particularly during the MJJ period in which the skills remain level at ~0.30 for up to 9 months’ lead-time. During the NEM period (DJF), the model produces skills of ~0.40 for up to 5 months’ lead-time, although it decreases gradually to ~0.25 for 9-months’ lead-time.

In Peninsular Malaysia, the QSST model outperforms other models, although the skills are relatively lower compared to those in the East Malaysian region. Notably, there are two periods i.e. February to April and May to June, in which the skills of the QSST model attained values >0.20. For the second half of the year, no models show useable skills. Thus, there is a notable difference of the QSST skill structure between regions in Peninsular Malaysia and East Malaysia. This difference supports the earlier contention of different behaviour of anomalous rainfalls over the two regions.

3.3. Experiments with combined predictor fields

The skills of various combinations of predictors are also examined. Since QSST is one of the most skillful predictors, the combined model takes QSST as the first predictor. Other predictor candidates are augmented one by one into the model with the implementation of the inter-field weighting technique and other CCA procedures highlighted in previous sections.

Figure 5 shows the skills for various combinations of predictors in the East Malaysia and Peninsular Malaysia regions. For comparison, the skill of the QSST model is also plotted. Interestingly, the most notable result is the inferior skill of the QSST + RAINFALL model. For all lead-times, this model produces significantly lower skills than other models. This may suggest that the incorporation of the rainfall predictor into the model increases the noise level particularly during seasons other than the NEM period. This could be due to the heavily
weighted rainfall field when it is joined with other predictor fields via the inter-field weighting process. On the other hand, the augmentation of other predictor fields with QSST generally does not significantly increase the predictive skills. In fact, the combination of the two best single predictor variables (i.e. QSST + QSLP) produces lower skills during some periods. Similarly, for the QSST + H700 combination, the skills are comparatively similar to those of QSST itself except in the case of a 5-month lead-time during October–November in East Malaysia in which there is a notable increase in skill. For both regions and almost all seasons and lead-times, the skills of QSST + RSST are slightly lower than those of the QSST. Combining all the three major predictors i.e. QSST + H700 + QSLP does not significantly increase the skill either.

Incorporation of other predictor fields increases the computation load steeply but generally does not significantly enhance the predictive skill and in some cases the performances are lower. On the other hand, QSST alone produces forecast skills that are comparable to those reported by He and Barnston (1996) for islands in the tropical Pacific region. Hence, the QSST model is selected as the optimum model in this study and further discussion is based entirely on this model. Yu et al. (1997) also used SST as a single predictor in their CCA model.

Figure 6 shows the lead-time target season cross section for the average correlation skill scores for the QSST model for both East Malaysia and Peninsular Malaysia. The differences in the skill structure between the two regions are apparent. In the East Malaysian region, the skills during late NEM and MJJ periods are less sensitive to lead-time compared to other seasons. Within the NEM period, there are changes in East Malaysia’s skill structure related to the dissipation as a function of lead-time. During January–February, the model attains an average skill of 0.3 for up to 8-month lead-time prediction. However, during early period of the NEM (November–December), the model only attains the same level of skill at 5-month lead-time prediction. These results indicate that, for the East Malaysian region, the long-lead forecast is feasible for the late stage of the NEM season, while for the early stage a forecast is only feasible at medium-lead.

In Peninsular Malaysia, the skills are generally much weaker compared to those over the East Malaysian region. The model does not produce useful skills during...
the second half of the year. However, during the first half of the year the model attains marginal skills particularly during the months of February and May in which the skills are greater than 0.2 for almost all lead-times. This result shows that a forecast is only marginally feasible in Peninsular Malaysia regions during certain seasons of the year, and it is an open question whether forecasts of this skill level should be issued operationally.

3.4. The spatial distribution of the model skill scores

Figure 7 shows the spatial distribution of skills for individual stations during the JFM period. Consistent with Figure 5, the higher skills concentrate in the East Malaysian region. Two stations, i.e. Kota Kinabalu and Labuan, registered skills greater than 0.5 even for predictions at 5-month lead-time (Figure 7(b)). A skill score of 0.3 is considered statistically significant at the 95% level and 0.3 is also considered to be the minimum level of skill that has any practical utility of societal benefit. Generally, for all lead-times, the skills over East Malaysia decrease southwestward, with Kuching station showing no apparently useable skills for all lead-times. In comparison, the skills in Peninsular Malaysia are relatively low with no apparent useable skills. Surprisingly, the most affected stations during the NEM season over the northeast coast of the Peninsular Malaysia, i.e. Kota Bharu, Kuala Terengganu and Kuantan, show no apparent useable prediction skills. These stations are located in the wettest area within Peninsular Malaysia during the NEM season. Figure 8 shows the time series plot for the standardized values of anomalous rainfall and the corresponding predicted values for Kota Kinabalu and Labuan stations at 5-month lead-time. Generally, for both stations, the model predicted the peaks and troughs associated with La Niña and El Niño events, particularly during the 1982/83 and 1997/98 El Niños and the 1999/2000 La Niña.

Figure 9 shows the spatial skill distribution during the MJJ period. For both regions, the prediction produces a secondary peak during this period (Figure 5). Also, during this period, the skills show relatively less sensitivity...
to the increment of lead-times. In comparison to the skill distribution during the JFM period, modest skill scores are evenly distributed across the country with slightly higher skill scores attained for longer lead-times (5 and 9 months). For 9-month lead-time predictions, both Kota Kinabalu and Kuala Lumpur stations attain skills of 0.43 and 0.47 respectively. Figure 10 depicts the time series of observed and forecast values for these two stations at 9-month lead-time prediction. It is immediately apparent that the time series in Figure 10 do not represent the typical interannual variation of ENSO as in the case of JFM. As described later, the relationship is related to a very different aspect of ENSO than illustrated for JFM. Generally, for both stations, the model captured large fluctuations in the observed rainfall reasonably well after the mid-1970s period. For the Kuala Lumpur station, the large rainfall deficit event during the 1989–1990 period and also the subsequent event of large amounts of rainfall in 1991–1992 were correctly forecasted 9 months ahead of their occurrence. However, during the pre-mid-1970s period, the observed rainfall time series show relatively small amplitudes. This suggests a significant shift between pre- and post-mid 1970s periods of the rainfall pattern. Interestingly, the model correctly captured the shift as the forecast amplitudes are relatively lower prior to the mid-1970s but higher after the period.

3.5. Sources of predictability during Jan–Feb–Mar (JFM)

In addition to the forecast, the CCA procedure produces useful diagnostic outputs that can help identify the physical processes that contribute to the observed forecasting skill and the identification of the origin of the predictability in the model (Graham et al., 1987). The first two CCA modes are used in the forecast model for 1-month lead-time forecast of the JFM rainfall anomalies. The first and second CCA mode accounted for about 33 and 24% of the predictand total variance respectively. Figure 11 displays the sequence of the mode 1 predictor loading patterns for the QSST in forecasting JFM rainfall anomalies at 1-month lead-time. These maps actually represent the extended EOF loading patterns that are highly correlated to the predictand loading pattern shown in Figure 12(a).

The loading patterns basically depict the 1-year evolution of anomalous SST in the Indo-Pacific region from the DJF period of the preceding year (i.e. DJF(−2)) to SON(−1)). The magnitude of the loadings indicates the relative importance of the SST anomalies in predicting the JFM(0) rainfall. The sequence begins from a decaying El Niño condition at DJF(−2) with an initial cooling observed at about 160–170°E in the central Pacific. The
cold SST subsequently spreads eastward and by JJA(−1) the entire eastern-central equatorial Pacific region experiences anomalously cold SST. Also, during the JJA(−1) period, negative SST anomalies begin to establish in the southern Indian Ocean centered at about 15°S and 90°E. Subsequently, the anomalous SST expands and together with warmer SST located west of Sumatera, a dipole pattern is established in the equatorial Indian Ocean during the SON(−1) indicating a negative mode of the IOD (e.g. Saji and Yamagata, 2003). At this stage, the condition of the anomalous SST in the Indo-Pacific region reflects that of a nearly developed La Niña event with the co-occurrence of a negative IOD event in the Indian Ocean. Hence, the temporal sequence of QSST loading patterns depicts a typical development and evolution of a La Niña event that begins with a decaying previous El Niño event. Equally, by reversing the polarity of the loadings, the sequence can represent a development of an El Niño event from a decaying La Niña event.

The corresponding predictand canonical pattern (usually referred to as the $h$-map), which consists of predictand loading values for the first canonical mode, is shown in Figure 12(a) with all stations except Kuala Lumpur showing positive loadings. Moreover, consistent with the spatial distribution of the forecast skill, large positive values are concentrated in the East Malaysian region. This implies that, in conjunction with a La Niña (an El Niño) event, areas across the country experience excess (deficit) rainfall during the late NEM period. Again, the East Malaysian region is affected more than Peninsular Malaysia.

The canonical correlation coefficient of CCA mode 1 for the 1-month lead forecast of JFM is 0.78. Figure 12(b) shows the predictor amplitude time series plotted together with the Niño3.4 index in which the two time series correlate strongly ($r = 0.79$). The power spectra of the predictor amplitude also indicate periodicity of a 3- to 5-year cycle (Figure 12(c)). These diagnostics, together with the evolution of the SST loading patterns, imply that the source of the predictability during JFM in which the most useable skills are located over East Malaysia, originate from ENSO. Most of these signatures are reflected in the first canonical mode as the canonical correlation in the second mode is weak (0.42) and only marginally contribute to the observed predictability.

As for JFM, the first two CCA modes are used in the forecast model for 1-month lead-time forecast of the MJJ rainfall anomalies. The first and second CCA mode accounted for 22 and 17% of the total variance respectively. Figure 13 displays the sequence of four SST loading patterns for the leading canonical mode for the prediction of MJJ rainfall anomalies at a 1-month lead-time. The sequence begins 13 months prior to the predictand season and thus span the periods of AMJ(−1), JAS(−1), OND(−1), and JFM(0). As in Figure 11, the magnitude of the loadings indicates the relative importance of SST anomalies in predicting anomalous rainfall during the MJJ period. The patterns clearly depict an anomalous SST pattern associated with La Niña events. In each period, the ‘tongue’ of negative loading extends westward from the eastern equatorial Pacific region to the date line. It is also flanked by the positive loadings in a boomerang-shaped pattern. Also, the negative loadings dominate almost the entire Indian Ocean throughout the sequence. Interestingly, in contrast to the corresponding loading pattern for predicting the JFM season (Figure 11), the patterns hardly change throughout the sequence. The minor changes are confined to the southeastern Indian Ocean and the Java Sea regions. Thus, the sequence indicates a nearly ‘steady state’ condition of the ENSO episode during the 1-year
evolution period. This suggests that the ENSO events depicted in the sequence of SST loading patterns are those of longer than a 1-year duration. However, this could also be related to the fact that the 1-year predictor time window coincides, approximately, with the life cycle of a typical ENSO episode. The predictor time window for JFM forecasts at 1-month lead did not match the seasons representing the typical ENSO episode life cycle, but rather included the last half of one cycle and the first half of the next cycle. Nevertheless, He and Barnston (1996) highlighted that, for the tropical Pacific Islands, the source of predictability of anomalous rainfall during the JJA period was associated with ENSO events of longer duration. This suggests that ENSO events of longer duration tend to affect anomalous rainfall across the country during the MJJ period. However, as shown in Figure 8, the forecasting skills were relatively lower and thus implied the weaker impact of ENSO of longer duration on the Malaysian anomalous rainfall. In contrast, the typical shorter duration ENSO events can have a profound impact on rainfall anomalies in Malaysia during the late NEM season. However, as indicated, the affected area is confined to the East Malaysian region.

The corresponding canonical predictand pattern for prediction of MJJ anomalous rainfall is shown in Figure 14. As for JFM, all stations except Kuala Lumpur show positive loadings. Moreover, most stations on the west coast of East Malaysia have larger loading amplitudes. In contrast to the pattern for JFM (Figure 12(a)), it is also shown that the stations located in the west coast and northern Peninsular Malaysia registered relatively higher loading values compared to those on the east coast of Peninsular Malaysia. Hence, this implies that stations on the west coast of East Malaysia and those in the west coast and northern parts of Peninsular Malaysia experience excess rainfall (except Kuala Lumpur) during the MJJ period in the presence of a longer duration of a La Niña event.

The canonical correlation coefficient of CCA mode 1 for the 1-month lead forecast of MJJ is 0.68. The corresponding predictor canonical component time series for the leading mode is shown in Figure 14(b). However, the Niño3.4 index is not plotted in the figure as the two time series do not correlate significantly. Two significant periodicities are shown i.e. ~3.5 and 6.5 years for the time series (Figure 14(c)). Interestingly, the characteristics of the time series during pre- and post-mid-1970s are markedly different. The fluctuations during the pre-mid-1970s period appear to have smaller amplitudes, while those during the post-mid-1970s have larger amplitudes. These characteristics are also apparent in the observed time series of anomalous rainfall as well as in the CCA forecasted rainfall anomalies (Figure 10). This suggests that SST evolutions after the 1970s tend to have relatively larger variability and affect the anomalous rainfall in Malaysia during MJJ more than prior to the mid 1970s. This major change after the mid-1970s is likely to be associated with the major climate regime shift during the mid-1970s (Trenberth, 1990). Several authors (e.g. An and Wang, 2000; Wang and An, 2001) have reported the profound impacts of ENSO episodes of stronger and longer duration after the mid-1970s.

In contrast to the JFM prediction, the second CCA mode contributed slightly higher to the predictability of the MJJ rainfall anomalies, although the SST predictor pattern is not physically clear. The canonical correlation coefficient for this mode is 0.52. This mode describes the weakening aspect of a cold event and the initial development of a weak warm episode (not shown), which are not accounted for in the first canonical mode. The anomalous rainfall pattern associated with this SST pattern depicts suppressed rainfall in the west coast of Peninsular Malaysia (not shown).

4. Summary

The predictability of seasonal rainfall anomalies in Malaysia is examined using the CCA model. The data
pre-processing and model building procedures followed that of Barnston (1994). The model performances were evaluated for up to 9-month lead-time based on cross-validated correlation skills between the observed and predicted values. In determining the optimal model, several possible predictors and combinations of predictors were evaluated and compared. The QSST was found to be the most skillful single field predictor. While He and Barnston (1996) reported a skill increment with the addition of northern hemisphere 700-hPa geopotential height as a secondary predictor field, current investigations revealed that the inclusion of other potential predictors combined with QSST does not lead to a significant improvement of model skill. In some cases, the inclusion of other predictors leads to a drop in cross-validated forecast skills. Hence, the QSST model remained as the optimum one on which the subsequent analyses were based.

In general, the skill of the CCA forecast is markedly better than that of the persistence forecast for short, medium, and long-lead forecasts. However, the skill structures for East Malaysia and Peninsular Malaysia are markedly different. Generally, the model produces relatively higher skills for stations in East Malaysia compared to those in Peninsular Malaysia. In East Malaysia, a prediction for the late NEM (i.e. JFM) period is found to be the most skillful, followed by a prediction for the early period of SWM (i.e. MJJ). In Peninsular Malaysia, the model showed modest skills during the MJJ period but the skills during the JFM period were generally low (below 0.30). Hence, during the late NEM period, the modest skills are concentrated in East Malaysia with station correlation values ranging from 0.36 to 0.54 for a prediction of 5-month lead-time, applying to stations in the northern part of Borneo (Figure 12(b)). Some stations also recorded skills greater than 0.3 for a long-lead forecast of 9 months (Figure 12(c)). This spatial distribution of skills is very much consistent with the spatial distribution of anomalous rainfall associated with ENSO highlighted in Juneng and Tangang (2005). Juneng and Tangang (2005) related the evolution of the anomalous SST for a period that spans from JJA to the following MAM season to the evolution of ENSO-related precipitation in the Southeast Asian region. On the basis of this analysis, it was found that the areas mostly affected by ENSO during DJF and MAM are confined to northern Borneo and the southern Philippines (Juneng and Tangang, 2005). It was also revealed that the establishment of the anomalous SST dipole in the western North Pacific region during DJF and MAM periods is related to an anomalous anti-cyclonic/cyclonic circulation over the southern Philippines and northern Borneo. It is through this regional atmosphere–ocean interaction that the ENSO–precipitation relationship over the southern Philippines and northern Borneo is maintained (Hsu et al., 2001; Wang et al., 2003; Tangang and Juneng, 2004; Juneng and Tangang, 2005); thus modest prediction skills over the region are present during the JFM period. However, as noted earlier, the inclusion of anomalous regional SST in the western North Pacific region did not enhance the model skills. This is because the anomalous SST dipole pattern in the region emerges only during the DJF and MAM periods, whereas the predictor period for predicting JFM ends in the SON period before the pattern appears. The skills, as shown by the predictor canonical pattern (Figure 11), originate from the evolution of anomalous SST associated with an ENSO event for a predictor period that spans from DJF(−2) to SON(−1). The IOD pattern during SON(−1) also indicates that the skills may partially originate from anomalous conditions associated with IOD. Furthermore, as shown in Juneng and Tangang (2005), such evolution is subsequently followed by a regional atmosphere–ocean interaction over the western North Pacific. Thus, the signatures of such atmosphere–ocean interaction may occur partly after the end of the predictor sequence.

In contrast, the forecast skills during the early period of the SWM season (MJJ) were relatively low but distributed widely across the country. Also, the skills appear to be relatively insensitive to lead-times. Interestingly, these skills also originated from anomalous SST associated with ENSO events. However, as shown by the ‘steady state’ of ENSO conditions in Figure 13, these were ENSO episodes of longer than a year duration, while forecast skills during the JFM period originated from typical ENSO events that grow during the middle of the calendar year. Hence, the typical ENSO event tends to affect regions in East Malaysia during the late NEM period, while the impact of longer duration of ENSO event is mostly felt throughout the country during the MJJ period. The lower sensitivity of MJJ forecast skills to the lead-time could be related to the steady-state characteristic of the ENSO as shown in the predictor field. Interestingly, the model correctly captured the regime shift during the mid-1970s.

Overall, the CCA model produces forecast skills comparable to those reported by He and Barnston (1996) and Yu et al. (1997) for islands in the tropical Pacific region. Although the model produces modest but usable skills only for certain seasons and regions, these can be useful for purposes such as mitigation and response strategies to anticipated ENSO-related flood and drought disasters. Future improvements of this model can focus on understanding why skills in some adjacent stations can be markedly different. Roles of local features such as topography and proximity to water may be important. Since Peninsular Malaysia and East Malaysia appear to have different levels and sources of predictability, building a separate CCA model for each region will be considered in our next investigation.

Acknowledgment

The authors are grateful to various agencies – MMD, CDC-NCEP, UKMO for providing various datasets. This research was funded by the Malaysian Ministry of Science and Technology IRPA Grant 08-02-0012-EA215.

Copyright © 2007 Royal Meteorological Society

DOI: 10.1002/joc
References


