Predictable patterns of the May–June rainfall anomaly over East Asia

Wen Xing1, Bin Wang2,3, So-Young Yim4, and Kyung-Ja Ha5

1Physical Oceanography Laboratory/CIMST, Ocean University of China and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, 2International Pacific Research Center and Department of Atmospheric Science, School of Ocean and Earth Science and Technology, University of Hawai`i at Mānoa, Honolulu, Hawaii, USA, 3Earth System Modeling Center, Nanjing University of Information Science and Technology, Nanjing, China, 4Korea Meteorological Administration, Seoul, South Korea, 5Division of Earth Environmental System, Pusan National University, Busan, South Korea

Abstract During early summer (May–June, MJ), East Asia (EA) subtropical front is a defining feature of Asian monsoon, which produces the most prominent precipitation band in the global subtropics. Here we show that dynamical prediction of early summer EA (20°N–45°N, 100°E–130°E) rainfall made by four coupled climate models’ ensemble hindcast (1979–2010) yields only a moderate skill and cannot be used to estimate predictability. The present study uses an alternative, empirical orthogonal function (EOF)-based physical-empirical (P-E) model approach to predict rainfall anomaly pattern and estimate its potential predictability. The first three leading modes are physically meaningful and can be, respectively, attributed to (a) the interaction between the anomalous western North Pacific subtropical high and underlying Indo-Pacific warm ocean, (b) the forcing associated with North Pacific sea surface temperature (SST) anomaly, and (c) the development of equatorial central Pacific SST anomalies. A suite of P-E models is established to forecast the first three leading principal components. All predictors are 0 month ahead of May, so the prediction here is named as a 0 month lead prediction. The cross-validated hindcast results demonstrate that these modes may be predicted with significant temporal correlation skills (0.48–0.72). Using the predicted principal components and the corresponding EOF patterns, the total MJ rainfall anomaly was hindcasted for the period of 1979–2015. The time-mean pattern correlation coefficient (PCC) score reaches 0.38, which is significantly higher than dynamical models’ multimodel ensemble skill (0.21). The estimated potential maximum attainable PCC is around 0.65, suggesting that the dynamical prediction models may have large rooms to improve. Limitations and future work are discussed.

1. Introduction

East Asian summer monsoon (EASM) rainfall has a profound influence on the lives of billions of people. Tremendous efforts and significant progress have been made in understanding the sources of variability of the EASM rainfall over the past two decades. The mechanisms that have been identified to influence EASM variations include El Niño–Southern Oscillation (ENSO) teleconnection during ENSO development phase [Zhang et al., 1996], the positive feedback between the western North Pacific (WNP) subtropical high and warm pool sea SST [Wang et al., 2000; Wang and Zhang, 2002; Lau et al., 2004; Lau and Nath, 2006; Kosaka et al., 2012; Wang et al., 2013; Xiang et al., 2013], the North Atlantic Oscillation (NAO) through Atlantic–Eurasian teleconnection [Lu et al., 2006; Z. Wu et al., 2009], the northern Eurasian snow cover during the previous winter and spring [Ogi et al., 2004; Yim et al., 2010], the Tibetan Plateau snow cover and anomalous thermal conditions [Zhang et al., 2004; Wang et al., 2008a], the spring arctic sea ice [B. Y. Wu et al., 2009], and the enhancement (weakening) of the thermal contrast between the Asian continent and the western North Pacific [Zhou and Zou, 2010; Zhao et al., 2012]. The question is how to test and translate these to useful prediction tools?

Traditional seasonal forecast of EASM deals with June–July–August mean rainfall anomalies, which may not be the best strategy because the EASM rainy season is typically from May to August [Wang and LinHo, 2002] and pronounced differences exist between the early summer (May–June, or MJ) and peak summer (July–August, or JA): both climatological mean states and the principal modes of interannual variability exhibit distinct spatial and temporal structures [Wang et al., 2009a; Li and Zhou, 2011; Qin et al., 2014; Oh and Ha, 2015]. Besides, the seasonal marches from May to June and from July to August are both relatively gradual [Wang et al., 2009a]. For these reasons, we are motivated to study separately the early and peak summer rainfall variability and predictability.
During MJ the East Asia (EA) subtropical front is a defining feature of Asian monsoon circulation, and it produces the most prominent precipitation zone in the global subtropics. The observed long-term mean precipitation in MJ is characterized by a salient rainband extending about 9000 km from the Bay of Bengal northeastward via Indo-China, southern China, all the way to east of Japan (Figure 1a). This prominent rainband is as associated with the EA subtropical front [e.g., Tao and Chen, 1987]. This early summer rainband reflects the EASM onset over the South China Sea in mid-May and the intense Meiyu/Baiu in mid-June [Ding, 1992; Chu et al., 2012]. Asian monsoon rainy season commences from the southeastern Bay of Bengal and rapidly expands northeastward through the Indochina peninsula (in early May) to the South China Sea (mid-May) and east of Taiwan (late May), establishing the EA subtropical rainband [Wang and LinHo, 2002]. In MJ, Indochina, southern China, Taiwan, and Okinawa all reach their yearly peak or one of the peaks of the local rainy seasons [Chen, 1983; Yim et al., 2014]. Therefore, prediction of the MJ precipitation is important for prediction of the total amount of summer rainfall of the EASM. Previous study has already predicted area-averaged rainfall over southern China during MJ [Yim et al., 2014]. However, predicting the spatial structure of the rainfall anomalies is more difficult but more valuable for end users.

The present work focuses on the MJ rainfall variability along the western portion of EA subtropical front that covers important densely populated land areas (20°–45°N, 100°–130°E). As shown in Figure 1, the climatological boreal summer precipitation amount in this region is large. It is also the region where the precipitation...
has large bias in climate models’ simulations [Wang et al., 2009b, Figure 1b]. Therefore, understanding the physical mechanisms responsible for precipitation anomalies in this domain and establishing prediction model are of vital importance for society.

The present work aims to develop physical understanding of the sources of the MJ rainfall predictability over EA and further establish a suite of physically based empirical or physical-empirical (P-E) prediction models to estimate the lower bound of the predictability. A detailed introduction about P-E prediction models will be given in section 3.2.

2. Dynamical Models’ Performance in Hindcast of MJ EA Rainfall

To assess the performance of rainfall prediction by numerical models, we used retrospective forecasts of four advanced atmosphere-ocean coupled models with initial conditions on various dates in May, including the National Centers for Environmental Prediction Climate Forecast System version 2 [Saha et al., 2014], the Australia Bureau of Meteorology Predictive Ocean Atmosphere Model for Australia version 2.4 [Hudson et al., 2011], the Geophysical Fluid Dynamics Laboratory Climate Model version 2.1 [Delworth et al., 2006], and the Frontier Research Center for Global Change Scale Interaction Experiment-F model [Luo et al., 2005]. Each model includes ensemble forecast with an ensemble size ranging from 10 to 40. It has been demonstrated that the multimodel ensemble (MME) prediction has considerably higher skill than individual model prediction [Lee et al., 2010]. To obtain higher skill, we used the MME prediction by simply averaging the four coupled models’ ensemble mean anomalies after removing their own climatology.

The selected models are notoriously deficient in simulating MJ precipitation climatology (Figure 1b). The models underestimate the strength of the subtropical rainband by about 40–60% while overestimating rainfall amounts over the western-central Pacific intertropical convergence zone by more than 50%. Given this large model bias, it is of interest to find out their prediction skills for MJ rainfall prediction.

Figure 2 shows the evaluation results for the four models’ MME hindcast experiments. The skill is measured by pattern correlation coefficient (PCC) between the observed and model-predicted MJ rainfall anomaly patterns over EA for each year from 1979 to 2010. The 32 year averaged PCC score is only 0.21. Is this low skill due to intrinsic limit of the predictability or due to models’ deficiencies? To address these questions, we will adopt a nondynamical model approach (i.e., physics-based empirical model), which will be introduced in section 3.2. It should be pointed out that the MME prediction results will not be used when establishing the physics-based empirical model. The performance of MME prediction given in Figures 1b and 2 is only for the purpose of showing the limitation of dynamical models on predicting MJ EA rainfall.

3. Methodology and Data

3.1. Data Used

Several observed data sets are used in this study, including (a) monthly mean precipitation from the Global Precipitation Climatology Project (GPCP) version 2.2 data sets [Huffman et al., 2011], (b) Climate Prediction Center Merged Analysis Of Precipitation (CMAP) [Xie and Arkin, 1997], (c) monthly mean SST from the
National Oceanic and Atmospheric Administration Extended Reconstructed SST (ERSST) version 3b [Smith et al., 2008], and (d) monthly mean circulation data from ERA-Interim [Dee et al., 2011] for the period of 1979–2015.

3.2. Empirical Orthogonal Function-Based Anomaly Pattern Prediction Using P-E Models

Objective prediction of precipitation anomalies in a given domain has been relied on either dynamical models or pure statistical models such as canonical correlation analysis [Barnett and Preisendorfer, 1987; Barnston, 1994] and empirical orthogonal function (EOF)-based partial least square method [Xing et al., 2016a]. Since it has been demonstrated that the dynamical models have limited capability in reproducing observed EOF patterns in section 5.1, we adopted an EOF-based physical-empirical model approach [Yim et al., 2015; Xing et al., 2016b; Xing and Wang, 2016]. A comparison of prediction skill between dynamical models and EOF-based physical-empirical model will be given in Figure 6.

The first step is to perform EOF analysis, which is the most convenient way to derive frequently observed patterns and to reconstruct the total variation of a geophysical field. The second step is to explore the physical processes that reproduce the pattern. Understanding the origin of the EOF patterns is more important than statistical test of their separability, especially when the sample size is limited and the separation of modes is difficult. If the EOF patterns are physical meaningful, we will use it as potentially predictable patterns. The third step is to predict the principal components by establishing a set of P-E prediction models. If the principal component (PC) can be predicted with significant skill, the corresponding EOF pattern may be considered as predictable mode. The last step is to predict the precipitation anomaly pattern by using the predictable modes, i.e., use the observed EOF patterns and the corresponding predicted principal components to reconstruct the total anomaly pattern.

A P-E prediction model is established based on understanding of the physical processes linking the predictors and the predictand. Statistical tests are used as an auxiliary tool to maximize the predictors-predictand correlation in training periods and to confirm their significance and ascertain mutual independence among the predictors [Wang et al., 2015].

Different from statistical approaches that fish predictors through variety of fields, the P-E model approach detects predictors by focusing on only two fields that reflect ocean and land surface anomalous conditions, i.e., the SST and 2 m temperature over land and the sea level pressure (SLP). The 2 m air temperature over land can, to a large extent, reflect snow cover anomalies. The SLP, albeit interacting with atmospheric heating, is often forced by surface thermal anomalies [Lindzen and Nigam, 1987], and persistent SLP anomalies often induce atmospheric anomalous heating. The P-E approach also searches only two types of lower boundary anomalies: (a) persistent signals from winter (January and February, JF) to spring (March and April, MA) (JFMA) and (b) tendency signals from JF to MA (IMA minus JF). The persistent signals normally reflect positive feedback processes associated with the local atmosphere-ocean or atmosphere-land interaction, which may help maintain the lower boundary anomalies. The tendency predictors denote changes from winter to spring and frequently tip-off the direction of subsequent evolution.

Note that the physically based searching principle requires examination of only four correlation maps (i.e., two fields and two types of anomalies) This makes the selection procedure not only objective and reproducible but also easy to apply to other climate prediction problems. In the selection of predictors from the aforementioned maps, we emphasize understanding of the processes that explain the lead-lag relationships between predictors and the MJ rainfall patterns.

Stepwise multiple linear regression method is used to establish P-E models for prediction. Prior to the regression, all variables are normalized by removing their means and divided by their corresponding standard deviation, which allows direct comparison of the relative contribution of each predictor by examining the normalized regression coefficient. The stepwise regression procedure identifies statistically important predictors at each step. Each selected predictor has significant contribution to increasing the regressed variance by a standard F test [Panofsky and Brier, 1968]. A 95% statistical significance level is used as a criterion to select new predictor at each step. Once in the model, a predictor can only be removed if its significance level falls below 95% by the addition/removal of another variable. To circumvent overfitting, the number of predictors is required to be less or equal to three (i.e., less than 10% of the sample size 37).
3.3. Cross-Validated Reforecast

To test the hindcast experiment skills, we used cross-validation method [Michaelsen, 1987] to make a retrospective forecast. To lessen overfitting problem, we leave 3 years of data out progressively centered on a forecast target year for the period of 1979–2015, then train the model by using the data of the remaining years and finally apply the model to forecast the three target years. A detailed procedure is described in section 5.1.


EOF analysis is used to identify the leading patterns of rainfall variability over EA (20°N–45°N, 100°E–130°E) during May–June for the 37 year period of 1979–2015 using GPCP rainfall data. The first three leading EOF modes of the MJ EA rainfall variability using GPCP data account for 26.7%, 13.1%, and 11.0% of the total variance.

Figure 3. (a) The spatial pattern and (b) the corresponding principal component of the first EOF mode (EOF1) derived from MJ precipitation over East Asia (20°N–45°N, 100°E–130°E) for the period of 1979–2015. Wind vectors in Figure 3a show the correlation coefficients with respect to the PC1. The correlation maps of (c) anomalous May–June (MJ) mean and (d) JFMA mean SST over ocean/T2m over land (color shading), and SLP (contours) with reference to the PC1. (e) The correlation maps of MA-minus–JF SST (ocean), 2 m air temperature (T2M, land) anomalies with respect to the time series of the PC1. The areas exceeding 95% confidence level for SST and T2M are dotted. The rectangular regions outline where the predictors are defined. The GPCP V2 and ERSST V3b data were used, and area weighting is applied in the EOF analysis.
precipitation variance, respectively. These EOF patterns and the corresponding PCs are shown in plots (a) and (b) of Figures 3–5, respectively.

Are these three EOF patterns physically meaningful? How do they link to lower boundary anomalies? What are physically consequential predictors for each PC? In the following sections, these questions are addressed for each pattern. We shall show that each of the three EOF patterns of the precipitation anomaly is physically meaningful and associated with distinct lower boundary temperature and SLP anomalies.

4.1. EOF1: Western Pacific Subtropical High-SST Dipole Coupled Mode

EOF1 is characterized by enhanced rainfall over southern China that represents an enhanced subtropical frontal rainfall in MJ and suppressed rainfall north of it (Figure 3a). The PC1 time series indicates that (a) this pattern often occurs after a peak El Niño, such as 1983, 1992, 1995, 1998, 2005, and 2010, and (b) there was a prominent decadal shift around the mid-1990s which was consistent with previous studies [Kwon et al., 2005, 2007]: The MJ mean rainfall amount increased significantly after the mid-1990s, and the periodicity also changed accordingly from 5–6 years to 2–3 years (Figure 3b). Since the central-Pacific El Niño (or El Niño Modoki) happened more frequently after the 1994 than before [Ashok et al., 2007; Yeh et al., 2009] and its impacts on the EASM are significantly different from the EP-type El Niño [Feng et al., 2011; Weng et al., 2007; Feng and
Chen, 2014], the decadal shift around 1993–1994 may be associated with the secular change in the ENSO properties.

The EOF1 is associated with a strong, large-scale anomalous high over the Philippine Sea and the North Pacific that persists from the previous winter (JF) to MJ (Figures 3c and 3d). This is the key circulation system that determines the EOF1 rainfall pattern. What mechanism drives the EOF1? The key process maintaining the western Pacific subtropical high (WPSH) anomaly is the positive thermodynamic feedback between the WPSH anomaly and the cooling to the southeast of the WPSH anomaly [Wang et al., 2000; Lau et al., 2004]. The positive SST anomalies over the Indian Ocean also persist from winter to spring, and these anomalies can enhance the WPSH through Kelvin wave-induced Ekman divergence [Xie et al., 2009]. However, one should note that the persistence of the northern Indian Ocean warming is also a result of the subsidence and easterly anomalies associated with the WPSH (Figure 3d) [also Du et al., 2009]. The northern Indian Ocean warming can in turn enhance the WPSH [Chowdary et al., 2011] by increasing precipitation heating whose equatorial component can generate anomalous (Kelvin wave) easterly and associated negative shear vorticity. Thus, the WPSH-SST dipole interaction over the entire Indo-Pacific warm pool maintains both the SST anomalies and the WPSH anomalies [Wang et al., 2013]. For this reason, we refer EOF1 as WPSH-SST dipole coupled mode. This mode has been recognized as responsible for the origins of the first EOF mode.
of EA summer rainfall [Wang et al., 2009a; Xing et al., 2016a] and the second EOF mode of the entire Asian summer monsoon rainfall [8. Wang et al., 2014].

The predictors for PC1 are shown in Figures 3d and 3e. There are two pronounced persistent predictors (Figure 3d): (a) the equatorial zonal tripolar SST anomalies that are often seen during a decaying El Niño phase and (b) the meridional tripolar SST anomalies over North Atlantic associated with a negative North Atlantic Oscillation (NAO). To depict the zonal tripolar SST anomaly, a predictor named JFMA EQSST is defined by the SST anomalies (SSTAs) averaged over the equatorial eastern Pacific (5°S–10°N, 120°W–80°W) and maritime continent (10°S–20°N, 80°E–140°E) minus the SST anomalies averaged over the central Pacific (5°N–15°N, 160°E–160°W). To quantify the North Atlantic meridional tripolar SSTA, a predictor JFMA NAT is defined by the SSTA averaged over the tropical (0–15°N, 60°W–15°W) and high-latitude (40°N–55°N, 60°W–30°W) Atlantic minus that averaged over the midlatitude Atlantic (25°N–35°N, 80°W–45°W). Figure 3e indicates that significant opposite tendency signals exist between the WNP SST and Eurasian 2 m air temperature (T2m). To depict this Eurasian-Pacific thermal tendency contrast, a predictor MA-JF EU-WPT is defined by the MA-JF tendency in 2 m air temperature averaged over Eurasia (EU) (40°N–60°N, 100°W–120°E) minus that in SSTA averaged over WNP (15°N–30°N, 120°E–180°E).

The three predictors represent three different processes that affect subsequent MJ EA rainfall. The predictor JFMA EQSST reflects SSTAs during a decaying ENSO that sets up a favorable condition for persistence of the WPSH anomaly. The warming to the west of the Philippine Sea and the cooling to its southeast imply an enhancement of Philippine Sea high pressure (anticyclone) [Wang et al., 2000]. Numerical experiments have shown that the positive thermodynamic feedback between the Philippine Sea anticyclonic anomaly and underlying Indo-WP SST dipole anomalies can maintain both the Philippine Sea anticyclone and the SST dipole through spring to early summer [Lau et al., 2004; Wang et al., 2013; Xiang et al., 2013]. Thus, the JFMA EQSST is a precursor for the enhanced MJ Philippine Sea anticyclone (Figure 3d).

The second predictor JFMA NAT reflects spring North Atlantic Oscillation (NAO) affecting MJ EA rainfall. Previous studies found that the North Atlantic SST anomalies can have a positive feedback with the local winds, which maintains themselves from winter into MJ [Z. Wu et al., 2009] (also Figure 3d). The NAO related SST anomalies may affect EASM through (a) excitation of the Eurasian wave train extending to northeast Asia and consequently influencing EA subtropical frontal rainfall [Z. Wu et al., 2009; Yim et al., 2013], (b) alternation of Arctic oscillation that changes the EA westerly jet stream and thus the strength of the WNPSH [Gong et al., 2011], and (c) excitation of westward propagating tropical Rossby waves to affect eastern Pacific trade winds and generate SST anomalies over the equatorial central Pacific, which in turn impact the strength of the Philippine Sea subtropical high [Wang et al., 2013]. Therefore, the SST anomalies during winter and early spring over the North Atlantic can be a physically meaningful predictor.

The third predictor MA-JF EU-WPT reflects the Asia-Pacific thermal contrast tendency, which presages development of the pressure gradient between EA low and WP high [Zhou and Zou, 2010]. The EU warming tendency also reflects reduced snow anomalies over central Asia, which is consistent with the results of Yim et al. [2010].

4.2. EOF2: North Pacific SST Mode

The EOF2 features an enhanced Meiyu over the lower reach of Yangtze River valley and western Japan, representing a northward advance of the normal subtropical front zone in MJ. The increased rainfall is associated with high SLP anomaly that extends from North Pacific to East China Sea. The corresponding PC2 shows a dominant 5 year oscillation but has no significant correlation with ENSO (Figure 4b). In fact, the equatorial SSTA across the Indo-Pacific region is insignificant from the previous winter to spring except over the WP where weak positive SSTAs persist (Figures 4c and 4d). However, EOF2 is associated with pronounced SSTA and SLP anomaly (SLPA) over the North Pacific (NP) (Figure 4c), similar to a pattern associated with a negative phase of Pacific Decadal Oscillation (PDO) [Mantua and Hare, 2002]. We speculate that the key system that shifts Meiyu northward, namely, the high SLP anomaly over the East China Sea and south of Japan, is a westward extension of the North Pacific high-pressure anomaly; thus, anomalous North Pacific SST is probably the origin of the EOF2 pattern. For this reason, the EOF2 is called north Pacific SST mode.

Figure 4d shows that significant persistent precursory signals for PC2 include (a) North Pacific SST and (b) a pronounced warming over the southeastern IO and maritime continent. Thus, two JFMA persistent predictors
are selected: (a) SSTA in the North Pacific (NP SST) defined by JFMA mean SST anomalies averaged over domain A (30°N–50°N, 150°W–130°W) minus that averaged over the domain B (30°N–45°N, 150°E–180°E) and (b) the IO SST anomalies defined by the JFMA mean SSTA averaged over the southeastern Indian ocean (45°S–10°S, 90°E–120°E).

The persistent NP SST and SLPA imply a positive feedback between the anomalous PDO high SLP and underlying SST dipole through reduced evaporation cooling in the west and enhanced coastal upwelling-induced cooling in the east. Thus, the PDO-like SSTA foreshadows the further development/maintenance of a North Pacific high SLP anomaly whose westward extension may lead to the increased SLP to the south of Japan and East China Sea (Figure 4c). As shown in Figure 4a, the anticyclonic circulation anomaly over East China Sea is a major circulation anomaly that produces the rainfall anomaly pattern of EOF2.

The persistent southeastern IO warming is a large-scale manifestation of the warming over and north of the maritime continent (Figure 4d), which may be associated with an early retreated EA winter monsoon and a weak Australian summer monsoon. The southeastern IO warming has been also recognized in Oh and Ha [2016]. The latter induces ocean warming to the west and north of Australia, which causes increased rainfall over the maritime continent in MJ (not shown). The intensified condensational heat over the maritime continent helps to establish the cyclonic circulation anomaly over the South China Sea through equatorial Rossby wave response and to reduce the rainfall over the southern coast of China (Figure 4a).

Figure 4e shows that the most evident tendency precursor is the SST cooling tendency over the northern Indian Ocean (NIO). Thus, the third predictor, MA-JF NIO SST, is defined by the SST tendency (MA minus JF means) averaged over the NIO (0°–15°N, 50°E–110°E). The cooling tendency in NIO foreshadows reduced rainfall there in MJ, which favors a cyclonic anomaly over the South China Sea through excitation of anomalous (Kelvin wave) westerly and associated cyclonic shear vorticity [Xie et al., 2009], thus reducing the rainfall along the southern coast of China, a feature of EOF2 (Figure 4a).

### 4.3. EOF3: Equatorial Central Pacific (ECP) Mode

EOF3 has large loading over South Korea, Taiwan, and the East China Sea (Figure 5a) but has little weight over mainland China. This rainfall pattern is associated with a strong anomalous high over Japan and enhanced southerly monsoon along the sea board of EA (Figures 5a and 5d). The corresponding PC3 shows an irregular year-to-year fluctuation but has no significant correlation with ENSO index (Figure 5b). Note, however, the EOF3 concurs with a notable cooling in the equatorial central Pacific (ECP) (Figures 5c and 5d). Previous studies have shown that ECP SST can affect EASM through teleconnection [e.g., Yim et al., 2008; Li et al., 2010]. For simplicity, we name EOF3 the ECP mode.

There are two precursors for development of the ECP cooling. The first is a PDO-like SST tendency pattern in the northeast Pacific (NEP) defined by the SST tendency (MA minus JF) between Area 1 and Area 2, where Area 1 is (20°N–35°N, 180°–150°W) and Area 2 includes (5°N–20°N, 160°W–130°W) and (20°N–35°N, 140°W–120°W) (Figure 5e). This tendency predictor is referred to NEP SST, which is consistent with the anticyclonic anomaly over the NEP in the previous spring (Figure 5e). The second is a persistent predictor, JFMA CPT, which is defined by the JFMA mean SSTA averaged over the equatorial-south central Pacific (20°S–5°N, 175°E–155°W) (Figure 5d).

The third predictor features a significant persistent high SLP anomaly over northeast Asia (northeast China and southeast Siberia) from JF to MA (Figure 5d) and the associated warming tendency (Figure 5e). It foreshadows the anomalous high over Japan, which is a key system to produce the EOF3 rainfall anomaly pattern. It seems that during MJ the high SLP anomaly over northeast Asia shifts slightly southeastward to form the anomalous high over Japan due to the effects of land warming and relatively ocean cooling toward summer. Thus, the third predictor named JFMA NEAP is defined as the JFMA mean SLP anomalies averaged over (40°N–60°N, 115°E–140°E). The warming tendency over northeast Asia (Figure 5e) is not selected because of its dependence of NEAP.

### 5. Hindcast With P-E Models and Potential Predictability

In the previous section, we have shown that the first three EOF patterns have different origins and predictors. Prediction of the early summer EA rainfall over the entire domain includes two steps. First, each PC is predicted. Second, the forecast field of the EA rainfall anomaly is reconstructed by using the sum of the
three observed spatial EOF patterns multiplied by their corresponding predicted PCs. It is worthy to mention that all predictors can be obtained 5 day before 1 May because the values for the last 5 days of April can be obtained from weather forecast, so the prediction here may be called a 0 month lead prediction.

5.1. Prediction of Principal Components With P-E Models

A suite of P-E models for prediction of each PC was established through stepwise multilinear regression method as introduced in section 3.2. To test the hindcast experiment skills, we applied cross-validation method to each P-E model [Michaelsen, 1987] to make a retrospective forecast. In the cross-validation method, the “training” and “prediction” samples at each step are strictly separated. At each cross-validation step, 3 years were taken out as “target prediction year” and the remaining years are used as training period. The EOF patterns and PCs are derived from “training period.” The PCs of the target years are predicted by using the multiregression equation derived from the corresponding training period.

Figure 6. The corresponding PCs of the first three EOF modes (a–c) in observation (OBS), cross-validated physical-empirical prediction (EmpM) by using three 0 month lead predictors for 1979–2015 and multimodel ensemble (MME) dynamical prediction. The numbers within the parenthesis in the figure legend indicate the PCC between the observed and predicted PC.
The predicted PCs are shown in Figures 6a–6c. For comparison the corresponding observed PCs are also shown in each panel. The cross-validated correlation skills between observation and prediction are 0.72, 0.56, and 0.48, respectively. The correlation coefficients are all significant at 99% level, which means that, to a large extent, the first three EOF modes can be regarded as "predictable." Figure 6 also shows the prediction skills of MME. The dynamical models can only capture EOF1. The correlation coefficient between observed PC1 and MME predicted PC1 is only 0.47, which is lower than PEM prediction skill (0.72). The prediction skills of MME for PC2 and PC3 are only 0.07 and 0.14, respectively. So the MME cannot capture EOF2 and EOF3. Therefore, the dynamical models have limited capability in reproducing observed EOF.

5.2. Prediction of Anomaly Pattern and the Potential Predictability

Given the fact that the first three EOF modes can be potentially predicted by the P-E models, we may estimate the potential maximum attainable forecast skill for the May–June EASM rainfall by assuming that the first three modes can be predicted perfectly. The potentially attainable forecast skill can be obtained from the correlation between the observed total field and the reconstructed predictable part (i.e., the first three predictable modes). For detailed formulation the reader may refer to Lee et al. [2013] and D. Wang et al. [2014].

Figure 7 shows the temporal evolution of the potentially maximum attainable pattern correlation coefficient (PCC) skill. It fluctuates from year to year with a skill in between 0.3 and 0.9 and the 35 year averaged PCC skill around 0.65. The red line in Figure 7 shows the pattern correlation coefficient (PCC) skill for each year obtained by using 3 year out cross-validated P-E models. The long-term mean of the PCC skill is 0.38, which shows large year-to-year variation with high skills (over 0.7) in the years of 1981, 1985, and 1998 and low skills (below ~0.2) in the years of 1990, 1993, 2000, and 2009.

6. Concluding Remarks

6.1. Conclusion

We have identified three predictable EOF modes of EA rainfall variability during May and June for the period of 1979–2015 (Figures 3–5). The EOF1 often occurs during ENSO decay phase and is sustained by the positive thermodynamic feedback between the western Pacific subtropical high (WPSH) and underlying dipole SST anomalies over the Indo-Pacific warm ocean. The EOF2 is mainly associated with the SST anomalies over the North Pacific Ocean, and the EOF3 is primarily induced by the equatorial central Pacific (ECP) SST anomalies. The factors that determine EA early summer rainfall variability include (a) positive thermodynamic feedback between the anomalous WPSH and underlying ocean, (b) the atmosphere-ocean interaction involved in North Pacific SST anomalies, and (c) the development of equatorial central Pacific SSTA.

To better understand the source of the potential predictability and to improve prediction skills, a suite of physical-empirical (P-E) models is established for prediction of first three leading principal components. The physically meaningful and statistically robust predictors are selected based on atmospheric lower
boundary anomalies that reflect persistence from winter to spring (January to April) and the tendency from winter (January–February) to spring (March–April). The major precursory conditions before May include (1) the following persistent signals from January to April: the equatorial Indo-Pacific tri-polar SSTA, the equatorial central Pacific SSTA, the tri-polar North Atlantic SSTA associated with NAO anomalies, the SSTA over the North Pacific, and the SSTA over the southeastern IO and the maritime continent; and (2) the following tendency precursors from winter to early spring: the Eurasia-WNP thermal contrast tendency, the northern Indian Ocean SST tendency, and the North Pacific dipolar SST tendency.

By examining the normalized regression coefficient of each P-E model, we can directly compare the relative contribution of each predictor. Among the above precursors, the Eurasian–Pacific thermal tendency contrast has large contribution to the first MJ EA rainfall mode. On the other hand, ENSO is the most important precursor for the EASM prediction during JA [Xing et al., 2016a]. This justifies the necessity to predict EA rainfall in MJ and JA separately. The predictors “JFMA IO SST” associated with air-sea interaction over warm pool region and “MA-minus-JF ECP SST” associated with development of ECP SSTA are the major contributor for EOF2 and EOF3, respectively. Some high-latitude signals (i.e., NAO, Eurasian, and Tibetan Plateau snow cover and arctic sea ice) seem to show less importance for the first three modes.

The cross-validated prediction results demonstrate that the first three modes can be predicted with significant skills ranging from 0.48 to 0.72 (Figure 6). Thus, they are identified as predictable modes. Using these predictable modes, we have made MJ rainfall hindcast over the EASM domain. The cross-validated (by taken 3 year data out) prediction skill (pattern correlation coefficient averaged through the 35 years) is 0.38 (Figure 7), which is substantially higher than the corresponding skill obtained by the current dynamics models’ multimodel ensemble (MME) hindcast (0.21; Figure 2), suggesting that the models have large room to improve.

We also estimated the predictability of the MJ rainfall over the EASM domain and found that the potential maximum attainable pattern correlation coefficient (PCC) skill may fluctuate between 0.3 to 0.9 from year to year and averaged PCC skill is about 0.65 (Figure 7c).

6.2. Discussion

The first three EOFs account for 45% of the total variance of interannual variability. With a short historical record, the EOF2 and EOF3 do not to meet the North et al. [1982] rule for the separability test. However, when data record increases, they may become statistically separable. To test this assertion, we made an EOF analysis of land precipitation for the same EA domain (i.e., the land coverage is the same except no ocean rainfall) during the 56 year period (1960–2015). The results show that the first three EOF patterns of rainfall over land and the corresponding PCs bear close similarity with those shown in Figures 3–5 (figure not shown), suggesting that the first three EOF modes derived from the two data sets are consistent and robust. Since the land only rainfall data set is longer (56 years), the first three EOFs became statistically separable at 95% significance level.

Understanding the physical linkages between the selected predictors and the corresponding PCs holds the key, but it is also challenging for the P-E prediction models. The first three principal components basically involve only the predictors that reflect anomalous surface thermal conditions, and we have detailed explanations regarding how these lower boundary thermal anomalies could lead to ensuing EA MJ precipitation and circulation anomalies through teleconnection and/or local atmosphere-ocean-land interaction. Based on the above discussion, we may conclude that the first three EOF modes are predictable. Therefore, about 45% of the total variance may be predictable for MJ EA rainfall.

With similar methodology one can derive a suite of prediction equations with longer lead time. The PC1 can be predicted at a 2 month lead using similar predictors as used for the forecast made in the end of April. In a previous study, we have revealed the processes that determine the MJ southern China rainfall and established a P-E model for its prediction at various lead times [Yim et al., 2014]. The predictors for the PC1 are generally consistent with those predictions for the MJ southern China rainfall but offer detailed spatial structure of the rainfall anomalies. The southern China rainfall reflects only the EOF1 very well. In the present study we further explored EOF2 and EOF3 and found their sources of predictability for seasonal prediction. However, at a longer lead the physical interpretation of the predictors would be more difficult and the skills would decrease as the lead time increases [Wang et al., 2008b; Lee et al., 2011].
We have examined four state-of-the-art dynamical models’ MME dynamic prediction. The current models have great difficulty in capturing EOF2 and EOF3 (figure not shown). This is likely due to the models’ bias in simulating climatology (Figure 1b), and more importantly, the models’ inability to capture the following processes: the atmospheric responses to SST anomalies outside the equatorial Pacific, the local monsoon-ocean interaction, and especially the impacts of land surface anomalous conditions and land-atmosphere interaction. The P-E model, at this stage, is valuable in detecting the sources of the predictability for the higher EOF modes and may serve as an effective tool for the difficult seasonal rainfall prediction.

In some years, the P-E model shows low prediction skills while the perfect PC prediction yields high PCC such as 1990, 2000, 2009, and 2013. The low skills in these years come from the deficiency of the PC prediction equations. The selected predictors cannot capture the variations in these years; hence, other sources of predictability should be further explored. For instance, the Baiu rainfall variabilities over Japan during June–July months have linkage with the monsoon rainfall over north-central India through teleconnection patterns [Krishnan and Sugi, 2001]. Observations and model simulation experiments show that warm SST anomalies in the equatorial eastern Indian Ocean (EEO) during 2000 played an important role in driving long monsoon breaks over India [Krishnan et al., 2003]. Thus, it would be interesting to see whether the teleconnection between the EEO SST anomalies and the early summer rainfall anomalies over East Asia is mediated through convection anomalies over India. Besides more and more studies that focus on the impacts of Arctic sea ice and aerosol on East Asia climate [Lau and Kim, 2006; Chen et al., 2014], related predictors may also be added to P-E models to improve the prediction skill in the future work. Given that there is still uncertainty among different data sets, further efforts should strive to assess the possible influence of such uncertainty on the seasonal precipitation prediction [North et al., 1982; Beven et al., 2008; Li and Zhou, 2011; Laloy and Vrugt, 2012; B. Wang et al., 2014].

Acknowledgments
This work was jointly supported by APEC Climate Center, the National Research Foundation of Korea through a Global Research Laboratory grant of the Korean Ministry of Education, Science and Technology (2011-0021927). We also acknowledge support from the Atmosphere-Ocean Research Center at University of Hawaii partially supported by Nanjing University of Information Science and Technology. The work was also supported by National Natural Science Foundation of China (41575067). The ERA-Interim data are available at https://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc/. This is the publication no 9905 of the SOEST, partial support by Nanjing University of Information Science and Technology. The work was also supported by National Natural Science Foundation of China (41575067). The ERA-Interim data are available at https://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc/. This is the publication no 9905 of the SOEST, publication no 1234 of IPRC, and publication no 147 of Earth System Modeling Center (ESMC).

References
Beven, K. P., and J. Freer (2008), So just why would a modeller choose to be incoherent?, Hydrology, 354, 15–32.


Zhou, T., and L. W. Zou (2010), Understanding the predictability of East Asian summer monsoon from the reproduction of land-sea thermal contrast change in AMIP-type simulation, J. Clim., 23(22), 6009–6026.