

# Seasonal forecasting of East Asian summer monsoon based on oceanic heat sources

Eungul Lee,<sup>a,b,\*</sup> Thomas N. Chase<sup>a,b</sup> and Balaji Rajagopalan<sup>a,c</sup>

<sup>a</sup> Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado, USA

<sup>b</sup> Department of Geography, University of Colorado, Boulder, Colorado, USA

<sup>c</sup> Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado, USA

**ABSTRACT:** We use the upper-level divergence zone at 150 hPa to define the areas of study for the East Asian summer monsoon (EASM) and to show the advances and retreats of the EASM. We find that the EASM can be subdivided into a northern and southern component with distinctly different driving mechanisms. The northern EASM (NEASM) is affected by heat sources in the tropical oceans related to El Niño events while the southern EASM (SEASM) is affected by the subtropical oceans related to a North Pacific sea surface temperature (SST) dipole mode. A stronger NEASM is related to above-normal western North Pacific (WNP) anticyclonic anomalies, while a stronger SEASM is related to below-normal WNP anticyclonic anomalies. These WNP anticyclonic anomalies are connected to SST anomalies in the tropical and subtropical Pacific during the pre-monsoon season (December~May). We also find that NEASM precipitation can be predicted from regional oceanic heat sources, i.e. SST and ocean heat content, in the tropical Pacific and Indian Oceans during the pre-monsoon season using a linear regression model. SEASM precipitation can be predicted from pre-monsoon SST in the eastern North Pacific. The NEASM forecast model is more skillful than that for the SEASM. Copyright © 2007 Royal Meteorological Society

**KEY WORDS** East Asian summer monsoon; forecasting model; sea surface temperature; ocean heat content; El Niño; Western North Pacific anticyclone; linear regression analysis; upper-level divergence

Received 20 May 2006; Revised 17 March 2007; Accepted 24 March 2007

## 1. Introduction

The East Asian summer monsoon (EASM) involves both tropical and mid-latitude weather systems including cross-equatorial airflow and the *Mei-yu*, *Changma* and *Baiu* frontal zones (e.g., Tao and Chen, 1987; Chang *et al.*, 2000a; Wu and Wang, 2002; Wang and Li, 2004; Lee *et al.*, 2005). To define the moisture sources of the EASM, many studies have focused on the relationships between the monsoon and sea surface temperature (SST) in Pacific and Indian Ocean (e.g., Shen and Lau, 1995; Zhang *et al.*, 1996; Chang *et al.*, 2000a,b; Wang *et al.*, 2000; Wu and Wang, 2002; Wang and Li, 2004; Lee *et al.*, 2005), because SST influences the atmosphere through surface heat fluxes, mainly latent heating (Wu and Newell, 1998).

Shen and Lau (1995), using 1956 ~ 85 data, found the EASM as represented by rainfall over China (but not including the Korean peninsula, Japan, and adjacent marginal seas) was positively correlated with the SST over the tropical central to eastern Pacific and the Indian Ocean in the preceding winter. During the El Niño episode of 1986/1987, the mature phase was in the northern summer and the EASM was intensified

(Zhang *et al.*, 1996). Wu and Wang (2002), using station rainfall data in mainland China, Japan, South Korea, and Taiwan, investigated changes in the inter-annual relationship between EASM rainfall and the preceding winter Niño3 SST anomaly in the late 1970s, concurrent with the Pacific climate shift. They identified that, from pre- to postshift period, the summer rainfall anomaly in eastern north China during the decaying phase of El Niño changed from above to below normal, whereas rainfall in central Japan changed from below normal to normal.

Anomalies in the intensity and spatial scale of the western North Pacific (WNP) anticyclone, which affect the EASM, are related to the Pacific-East Asian (PEA) teleconnection pattern and are the mechanism that links the tropical central Pacific SST anomalies with East Asian climate variations. The PEA teleconnection is a vorticity wave that emanates from tropical central Pacific convection anomalies and propagates westward and poleward against the westerly jet stream (Wang *et al.*, 2000). During the mature phase of an El Niño/Southern Oscillation (ENSO), the PEA pattern consists of an anomalous anticyclonic circulation over the WNP and anomalous cyclonic circulations over both the eastern North Pacific and East Asia (Lee *et al.*, 2005). Thus, this WNP anticyclone plays an important role in the connection between the East Asian monsoon and the

\* Correspondence to: Eungul Lee, CIRES, Campus Box 216, University of Colorado, Boulder, Colorado, CO 80309-0216, USA.  
E-mail: Eungul.lee@colorado.edu

tropical Pacific SSTs (e.g., Chang *et al.*, 2000a,b; Wang *et al.*, 2000; Wang and Li, 2004; Lee *et al.*, 2005).

In our study, we examine rainfall over a more general EASM region encompassing not only eastern China, but the Korean peninsula, Japan, and adjacent marginal seas. We divide the EASM into two regional systems, i.e. the northern EASM (NEASM;  $30^{\circ} \sim 50^{\circ}\text{N}$  and  $110^{\circ} \sim 145^{\circ}\text{E}$ ) and southern EASM (SEASM;  $20^{\circ} \sim 30^{\circ}\text{N}$  and  $110^{\circ} \sim 145^{\circ}\text{E}$ ) for reasons explained in Section 2. Lee *et al.* (2005) showed that the spatial correlation pattern between the NEASM and SST was similar to the structure of SST anomaly during El Niño. However, regional variability and the interactions of the sub-EASMs (i.e. NEASM and SEASM) and the surrounding oceans have not been previously addressed.

Most monsoon studies have focused on SST anomalies; however, ocean heat content (OHC) may be an important index for explaining the relationship between the EASM and the ocean as a heat source. Zhou and Chan (2007) investigated the relationship between the onset date of the South China Sea summer monsoon and ENSO using the ENSO signal represented by OHC. We explore the possibility that it may also be a useful predictor as it generally exhibits less high frequency variability than SST due to the thickness of the layers examined. For instance, interquartile range of the standardized anomalies of 1979 ~ 2003 annual mean OHC in the North Pacific (= 1.31) is less than that of SST (= 1.78). The subsurface thermal structure anomaly showed the El Niño during December 2002 ~ February 2003 more clearly than the corresponding SST anomaly, as it showed not only a positive response in the central and east-central equatorial Pacific but also a negative response in the western Pacific (Bell and Halpert, 2004). Kallummal and Kirtman (2006) also mentioned the importance of subsurface information in efforts to predict ENSO through a comparison of the prediction skill of the Markov models constructed with SST alone and SST and OHC.

Research on the effects of both SST and OHC during the pre-monsoon season, i.e. the preceding winter and spring, on the NEASM and SEASM is rare, even though this monsoon system (including *Mei-yu* in China, *Changma* in Korea, and *Baiu* in Japan) greatly influences both the weather and climate due to prevailing winds and heavy rain. The study of how the sub-EASMs are related to the changes of heat budgets in the surrounding oceans could help to determine the influence of a changing climate on the EASM and it may facilitate better seasonal forecasting.

The major objectives of this study are (1) to determine the relationships between EASM precipitation, NEASM and SEASM precipitation, and the heat sources of the surrounding oceans, which influence monsoon strength; and (2) to develop forecast models for NEASM and SEASM precipitation based on ocean thermal factors during the pre-monsoon season. Section 2 explains the data used. In Section 3, the advances and retreats of the EASM

are shown in terms of upper tropospheric divergence patterns. In Section 4, we examine the relationships between EASM precipitation and the heat sources of the surrounding oceans by means of spatial correlation analysis. A physical mechanism is then proposed in Section 5 to understand the relationships between the sub-EASMs and SST anomalies in the surrounding oceans using the composite analysis of low-level flow. In Section 6, multivariate linear regression is applied to develop models for NEASM and SEASM precipitation. Finally, conclusions and remarks are presented in Section 7.

## 2. Data

The EASM is normally referred to as a subtropical monsoon covering eastern China, the Korean peninsula, Japan, and adjacent marginal seas (e.g., Zhang *et al.*, 1996; Wang and Lin, 2002; Wu and Wang, 2002; Wang and Li, 2004). The region of the EASM in this study is  $20^{\circ} \sim 50^{\circ}\text{N}$  and  $110^{\circ} \sim 145^{\circ}\text{E}$ , divided into the NEASM and SEASM regions (Figure 1(a)). To define the areas for study of the EASM, we consider the regions of onset and withdrawal defined by divergence maxima in the upper troposphere. The sub-EASM regions have different mean divergences at 150 hPa for June through August (JJA). Stronger upper-level divergence occurs in the NEASM region (Figure 1(a)). The sub-regions also have different correlations with the heat sources of the surrounding oceans, as explained in Section 4.

The mean precipitation (mm/day) derived from the Global Precipitation Climatology Project version 2 (GPCP; Adler *et al.*, 2003) is used to calculate EASM precipitation which is area-averaged JJA precipitation over the each EASM region. In order to determine advances and retreats of the EASM, horizontal divergence ( $\text{s}^{-1}$ ) at 150 hPa is calculated with *u*- and *v*-winds obtained from the National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research reanalysis (Kalnay *et al.*, 1996). To understand the physical relationship between the EASM and SST anomalies in the surrounding oceans, wind vector (m/s) and vorticity ( $\text{s}^{-1}$ ) at 850 hPa are calculated with *u*- and *v*-winds from NCEP reanalysis. For oceanic heat sources, we use SST ( $^{\circ}\text{C}$ ) from Hadley centre (Rayner *et al.*, 2003) and OHC ( $10^9 \text{ J/m}^2$ ) from Scripps Institution of Oceanography (White *et al.*, 2001). OHC was calculated by depth in 11 levels (0 ~ 400 m), hence, it can express the characteristic of three-dimensional ocean. For OHC, the optimum interpolation procedure of Gandin (1963) was used to interpolate temperature anomalies from observation locations to the uniform grid (i.e.  $2^{\circ}$  latitude by  $5^{\circ}$  longitude) each month ([http://jedac.ucsd.edu/DATA\\_IMAGES/DATA/ABOUT/about.html](http://jedac.ucsd.edu/DATA_IMAGES/DATA/ABOUT/about.html)). The regional observing system from buoys and ships was expanded to include the entire Pacific Ocean ( $30^{\circ}\text{S} \sim 60^{\circ}\text{N}$ ) in 1980, since OHC data in the equatorial and northern oceans are available, which is used to calculate the potential predictors for the EASM

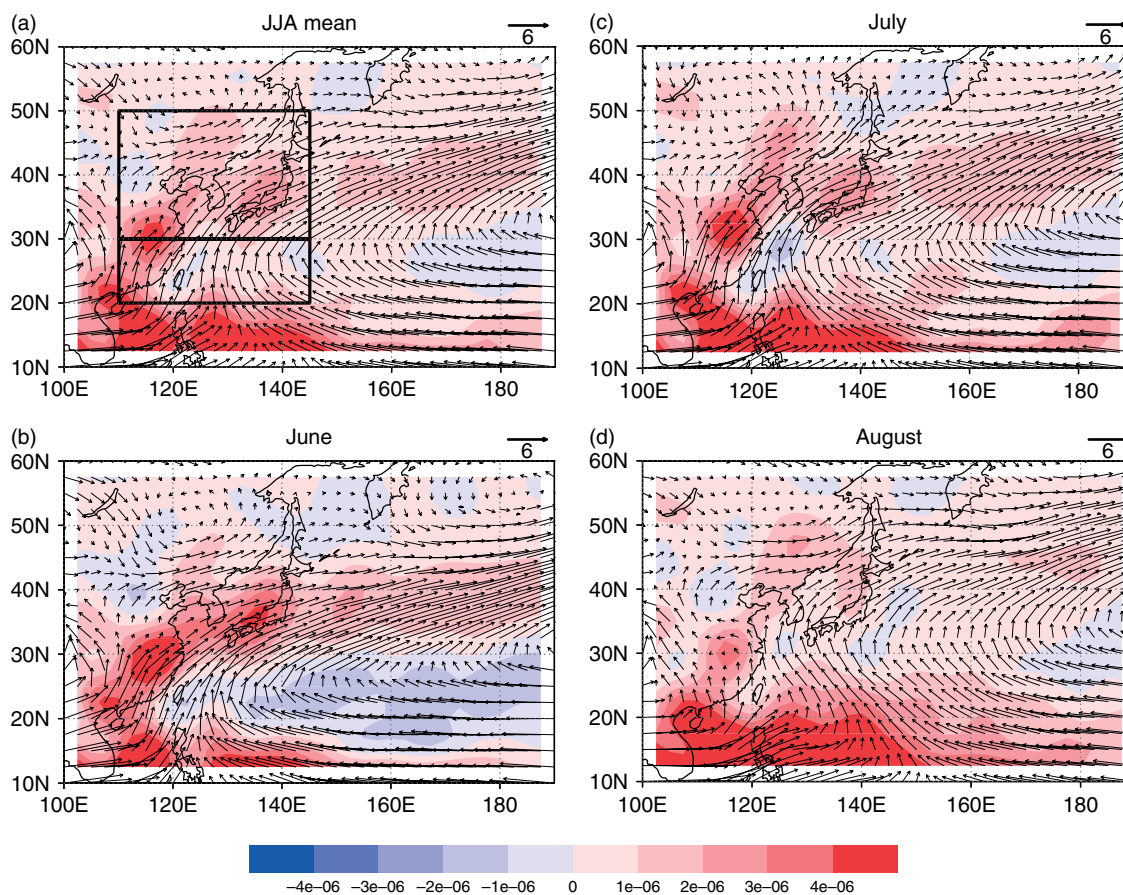


Figure 1. JJA and the monthly mean 150 hPa divergences ( $s^{-1}$ ; shaded) and 850 hPa wind vectors (m/s) for 1979 ~ 2003. The upper and lower boxes in (a) denote the NEASM and SEASM regions, respectively. Positive (red) and negative (blue) values represent upper-level divergences and convergences. This figure is available in colour online at [www.interscience.wiley.com/ijoc](http://www.interscience.wiley.com/ijoc)

models in Section 6. Spatial resolutions (longitude  $\times$  latitude) of these datasets are  $2.5^\circ \times 2.5^\circ$  for GPCP and NCEP,  $1^\circ \times 1^\circ$  for SST, and  $5^\circ \times 2^\circ$  for OHC data.

We examine 25 years from 1979 to 2003. GPCP data started in January 1979, and OHC data is only available through December 2003. Since the mid-1970s, the global average surface temperature has shown very steep increases (e.g., Hansen *et al.*, 1999). The global mean ocean temperature increase for the 0 ~ 300 m layer was  $0.31^\circ\text{C}$ , corresponding to an increase in heat content for this layer of  $\sim 10^{23}$  J between the mid-1950s and mid-1990s (Levitus *et al.*, 2000). This warming climate might influence local monsoonal hydrological cycles in the EASM region. Also, in the late 1970s, the reanalysis incorporated more reliable satellite data, and, in 1980, the regional observing system for oceanic subsurface temperature was expanded to include the entire Pacific Ocean. The period 1979 ~ 2003 is one of increasing temperature and more accurate datasets.

Seasonally averaged values of December through February (DJF) and March through May (MAM) are defined as the pre-monsoon season, and JJA as the EASM season, which are the 3 months with highest precipitation. Ocean thermal factors affecting EASM precipitation are decided by the spatial correlation analysis in Section 4,

and the best predictors of the forecast models are selected by stepwise procedures in Section 6.

### 3. Advances and retreats of the EASM

The onset times of the summer monsoon in East Asia are different in each regional frontal zone, whether *Mei-yu*, *Changma*, or *Baiu*. The annual migrations of the centers of upper-level divergence at 150 hPa (e.g., Chase *et al.*, 2003) are used to define the advances and retreats. The monthly mean divergence is shown in Figure 1 (shaded). During the EASM season, area-averaged divergences at 150 hPa over the NEASM and SEASM regions have significant positive correlations with NEASM precipitation ( $r = 0.87$ ) and SEASM precipitation ( $r = 0.83$ ) at significance level ( $\alpha$ ) = 0.01, indicating its usefulness as a measure of monsoon intensity.

The upper-level divergence zone at 150 hPa, which coincides with the lower-level convergence zone of the summer monsoon, begins in May over southeastern China and Japan. In June, the divergence zone is at its maximum intensity and forms a band that includes regional frontal zones: *Mei-yu* in eastern China, *Changma* in the southern Korean peninsula, and *Baiu* in southern Japan (Figure 1(b)). The divergence zone over East Asia

elongates to the upper-level divergence zone over the northwestern Pacific. In July, the intensity of the divergence gets weaker over Japan (Figure 1(c)). In August, the monsoonal band advances to northeastern China at around 50°N (Figure 1(d)). The upper-level divergence over the northwestern Pacific gets weaker. In September, the upper-level divergence zone begins to retreat to the southeast and is over East Sea (Sea of Japan) and Japan. During the EASM period, the upper-level divergence weakens from June through August. Southwesterly and southeasterly low-level flows in the tropical oceans converge over Philippine Sea, and they flow into the EASM region (see wind vectors in Figure 1(a)). Thus, moisture flowing into the region of deep convection from the surrounding oceans may affect the EASM.

#### 4. Relationship between EASM precipitation and ocean heat sources

In order to determine the effects of the surrounding oceans on the EASM, we calculate correlation coefficients of JJA precipitation over the sub-EASM regions with seasonally averaged ocean thermal factors, i.e. SST and OHC, in each grid over globe. NEASM precipitation is significantly correlated with the principal component 1 (PC 1) NEASM index obtained from the first leading empirical orthogonal function (EOF 1) of JJA precipitation over the 20° ~ 50°N and 100° ~ 180°E ( $r = 0.70$ ). The PC 1 NEASM index is defined as the PC 1 time series multiplied by area-averaged value of the first eigenvector over the NEASM region (Lee *et al.*, 2005). NEASM and SEASM precipitation are also significantly correlated with area-averaged divergences at 150 hPa over the NEASM and SEASM regions, as mentioned in Section 3, thus we use them as an index of each sub-EASM. The correlation patterns of JJA divergences at 150 hPa over the NEASM and SEASM regions with the ocean thermal factors (figure not shown) are consistent with those of JJA precipitation over the NEASM and SEASM regions with the ocean thermal factors, respectively, but the latter is more significant than the former.

##### 4.1. NEASM precipitation

SSTs over the large area of tropical Pacific and Indian Oceans are significantly correlated with NEASM precipitation from December through May at  $\alpha = 0.05$  (Figure 2(a) and (b)). NEASM precipitation is positively correlated with SSTs in the tropical eastern Pacific (TEP) and tropical Indian Ocean (TIO), and negatively correlated with SSTs in the tropical western Pacific (TWP) and subtropical South Pacific. The area of each region is described in Table I. A significant positive correlation ( $r = 0.49$ ) at  $\alpha = 0.05$  between NEASM precipitation and SST in the TEP for December through May supports the strong relationship between the intensity of NEASM precipitation and El Niño events. Consistent with our results, Chang *et al.* (2000a) showed that a wet Yangtze River valley monsoon occurred after warm eastern Pacific SST anomaly during the preceding winter. Our

correlation patterns are also similar to those of NEASM index with SSTs by Lee *et al.* (2005), constructed for 1979 ~ 2001. In our patterns, there is a more significantly positive relationship in the Indian Ocean and a more negative relationship in the South Pacific during the pre-monsoon season than in Lee *et al.* (2005). A significant negative correlation in the TWP extends into the subtropical North and South Pacific during the pre-monsoon season. NEASM precipitation has positive correlations with OHCs in the TEP and TIO, and a negative correlation with OHC in the TWP during the pre-monsoon season (Figure 3(a), (b), and (c)).

##### 4.2. SEASM precipitation

SEASM precipitation is significantly correlated with SSTs in the subtropical Pacific during the pre-monsoon season at  $\alpha = 0.05$  (Figure 2(d) and (e)), but there is little correlation between SEASM precipitation and SST in Indian Ocean. During JJA, the relationships between SEASM precipitation and SSTs in North Pacific and Indian Ocean (Figure 2(f)) are not as strong as those between NEASM precipitation and SSTs. SEASM precipitation has a negative correlation with SST in the eastern North Pacific (ENP), and it has positive correlations with SSTs in the WNP, East and South China and Philippine Seas (CPS), and the subtropical South Pacific. SEASM precipitation is positively correlated with OHCs in the WNP, South China Sea (SCS), and the subtropical South Pacific (Figure 3(d), (e), and (f)). In the ENP, a negative correlation between OHC and SEASM precipitation is weaker than that between SST and SEASM precipitation. For March to August, there is a dipole which is positive in the TIO and negative in the southern Indian Ocean (SIO; Figure 3(e) and (f)). OHC in the SIO is included in the potential predictors for the SEASM precipitation model.

As a robust verification of Figure 2, we perform the reverse correlations that are computed with JJA precipitation at each grid point and area-averaged SSTs over the TEP and ENP during the preceding DJF. The correlation patterns (figure not shown) are consistent with the results of Figure 2(a) and (d).

#### 5. Physical mechanism

In order to understand physically the relationship between the EASM and SST anomalies in the surrounding oceans, we use the composite analysis of mean JJA 850 hPa wind vector and vorticity. The WNP anticyclone plays an important role in the connection between the East Asian monsoon and the tropical SST anomalies (e.g., Chang *et al.*, 2000a,b; Wang *et al.*, 2000; Wang and Li, 2004; Lee *et al.*, 2005). Chang *et al.* (2000a) showed that a wet Yangtze River valley (YRV; i.e. 27° ~ 34°N and 110° ~ 125°E) monsoon was associated with a strong western Pacific subtropical ridge, which occurs after warm eastern Pacific SST anomaly during the preceding winter. In contrast to the YRV results, the reduction of rainfall in

SEASONAL FORECASTING OF EAST ASIAN SUMMER MONSOON

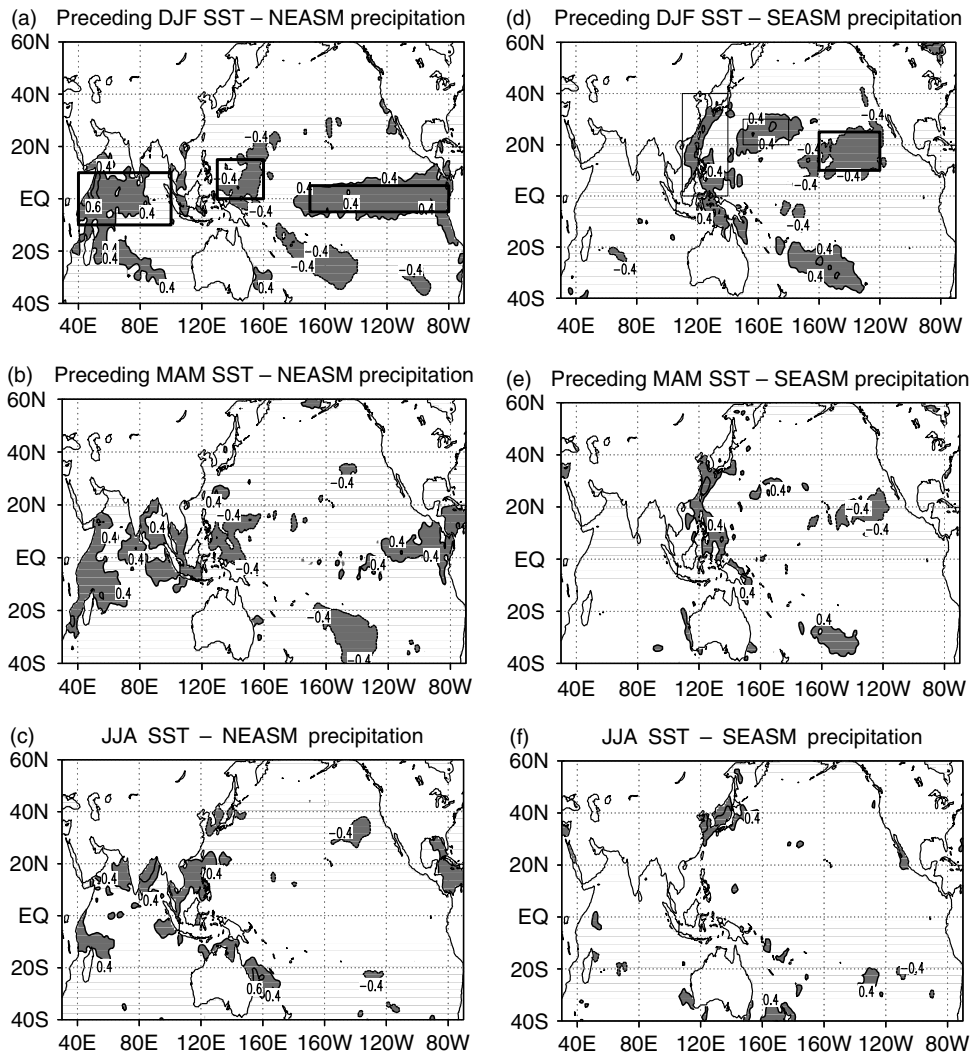


Figure 2. Correlation coefficients of area-averaged JJA precipitation over the NEASM and SEASM regions with seasonal SSTs. Significant regions at the 95% level are shaded. The boxes in (a) and (d) indicate the regions where potential predictors are calculated over. The regions for selected predictors are shown in bold boxes.

Table I. Areas and correlation values of potential predictors of NEASM and SEASM precipitation. Bold represents the selected predictors. \*\* and \* represent that correlations are significant at  $\alpha = 0.01$  and  $0.05$  levels, respectively.

Response variables	Potential predictors	Areas	Correlation values
NEASM precipitation	<b>SST in the TEP</b>	5°S – 5°N, 170° – 80°W	+0.49*
	<b>SST in the TWP</b>	0° – 15°N, 130° – 160°E	–0.48*
	<b>SST in the TIO</b>	10°S – 10°N, 40° – 100°E	+0.50*
	OHC in the TEP	10°S – 5°N, 120° – 80°W	+0.48*
	OHC in the TWP	10°S – 5°N, 140° – 180°E	–0.45*
	<b>OHC in the TIO</b>	20°S – 15°N, 50° – 70°E	+0.54**
SEASM precipitation	<b>SST in the ENP</b>	10° – 25°N, 160° – 120°W	–0.56**
	SST in the WNP	20° – 30°N, 150° – 180°E	+0.54**
	SST in the CPS	0° – 40°N, 110° – 140°E	+0.54**
	OHC in the WNP	20° – 30°N, 150° – 180°E	+0.44*
	OHC in the SCS	5° – 20°N, 110° – 130°E	+0.46*
	OHC in the SIO	10° – 20°S, 70° – 100°E	–0.42*

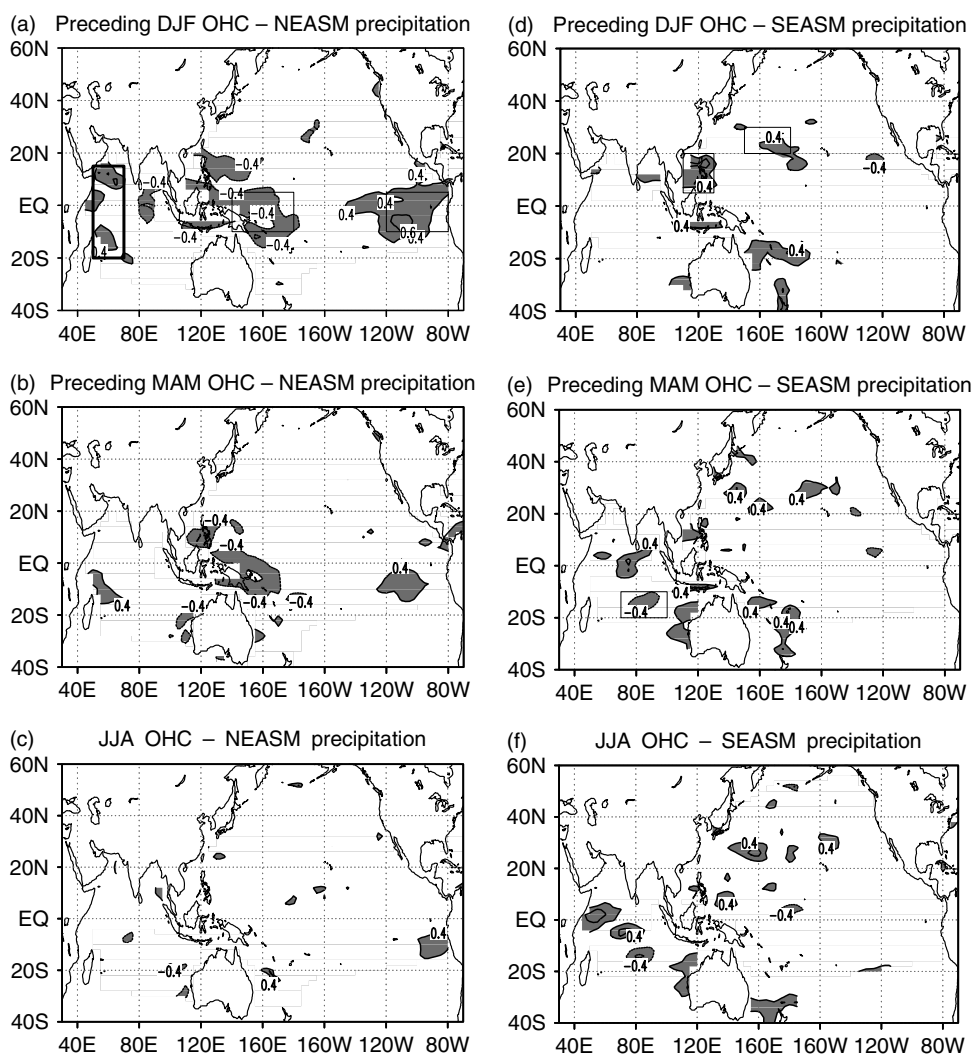


Figure 3. As in Figure 2, but with seasonal OHCs.

the southeastern coastal area of China (i.e.  $20^{\circ} \sim 27^{\circ}\text{N}$  and  $110^{\circ} \sim 125^{\circ}\text{E}$ ) during 1978 ~ 96 was attributed to the stronger subtropical ridge in the WNP (Chang *et al.*, 2000b). Figure 4 shows the composite differences of mean JJA 850 hPa wind vector and vorticity for 5 years of highest and of lowest of NEASM precipitation, SEASM precipitation, and SSTs in the tropical and subtropical Pacific during the pre-monsoon season.

### 5.1. NEASM

For the years of highest SST anomalies in the TEP, the Pacific-East Asian (PEA) teleconnection pattern, which consists of a wave creating an anomalously strong WNP anticyclonic circulation and anomalously strong cyclonic circulation in the NEASM region, creates more monsoon rainfall (Figure 4(b)). The spatial pattern of composite difference for the years of highest SST anomalies in the TEP is similar to that for the years of highest NEASM precipitation in Figure 4(a). This physical relationship is consistent with a positive correlation between NEASM precipitation and SST in the TEP in Figure 2(a) and (b). For the years of highest SST anomalies in the TWP,

there are fewer WNP anticyclonic anomalies due to the PEA teleconnection and also fewer cyclonic anomalies in the NEASM region, which create a weaker NEASM (Figure 4(c)). The spatial pattern of composite difference for the years of highest SST anomalies in the TWP is nearly opposite to that for the years of highest NEASM precipitation in Figure 4(a). Again, this is consistent with a negative correlation between NEASM precipitation and SST in the TWP in Figure 2(a) and (b).

### 5.2. SEASM

SEASM precipitation is predicted by the preceding SST anomalies in the ENP and WNP (Figure 2(d)). Anomalies in these two regions are of opposite sign with the correlation of  $-0.55$ . This SST dipole has been identified as the North Pacific mode, which is linearly independent of ENSO (Deser and Blackmon, 1995). SST anomalies in the ENP are the more statistically significant predictor for SEASM precipitation (Table I), so cold SST anomalies in the WNP are removed from the SEASM precipitation model for reasons of multicollinearity (i.e. the predictors are correlated amongst themselves). However,

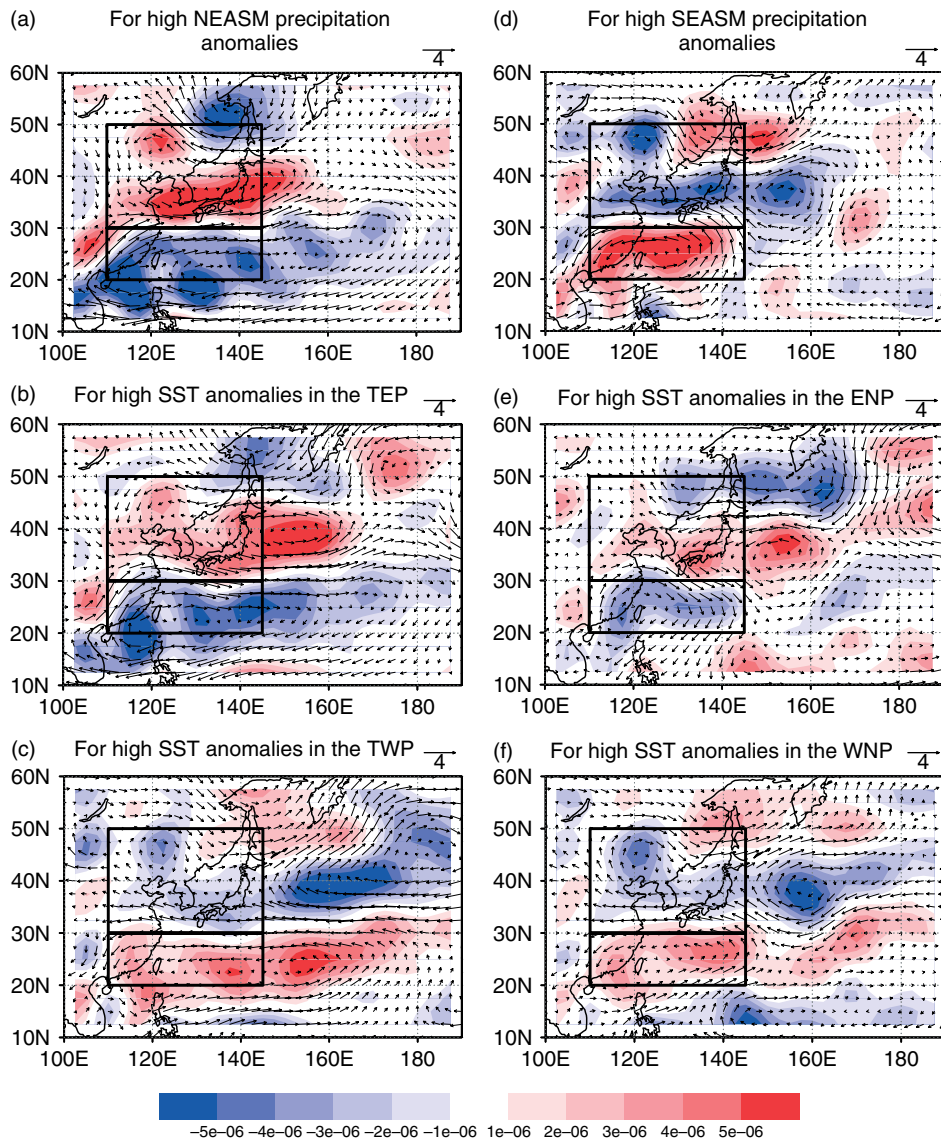


Figure 4. Composite differences of mean JJA 850 hPa wind vectors (m/s) and vorticity ( $s^{-1}$ ; shaded) for the 5 years of highest and of lowest of (a) NEASM precipitation, (b) SST in the TEP, (c) SST in the TWP, (d) SEASM precipitation, (e) SST in the ENP, and (f) SST in the WNP. The upper and lower boxes denote the NEASM and SEASM regions. This figure is available in colour online at [www.interscience.wiley.com/ijoc](http://www.interscience.wiley.com/ijoc)

it is cold SST anomalies in the WNP that are physically responsible for change in monsoon rainfall. For the years of highest SST anomalies in the ENP, cold WNP SST anomalies create a stronger WNP anticyclone (Figure 4(e)). The proximity of the SEASM to the WNP anticyclone causes diminished monsoon rainfall when the anticyclone is stronger or covers a larger area. The spatial pattern of composite difference for the years of highest SST anomalies in the ENP is nearly opposite to that for the years of highest SEASM precipitation in Figure 4(d). This is consistent with a negative correlation between SEASM precipitation and SST in the ENP in Figure 2(d) and (e). For the years of highest SST anomalies in the WNP, there are fewer WNP anticyclonic anomalies and thus fewer anticyclonic anomalies in the SEASM region, creating a stronger monsoon (Figure 4(f)). The spatial pattern of composite difference for the years of highest SST anomalies in the WNP is similar to that for the years

of highest SEASM precipitation in Figure 4(d). Again, this is consistent with a positive correlation between SEASM precipitation and SST in the WNP in Figure 2(d) and (e).

Thus, oceanic heat sources in the tropical and subtropical Pacific can impact the EASM through the intensity of WNP anticyclonic anomalies, which have different effects on the NEASM and SEASM. Physical linkages proposed between SST anomalies and the sub-EASMs are shown in Figure 5.

## 6. Forecast models for NEASM and SEASM precipitation

We use the linear regression framework to develop the forecast models. Regression models in general can be written as

$$Y = f(\mathbf{x}) + e \quad (1)$$

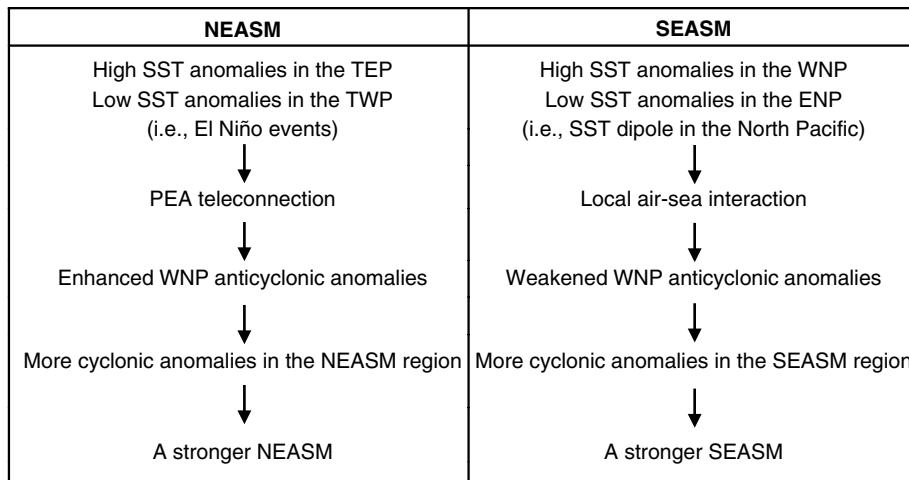


Figure 5. Physical linkages proposed between SST anomalies and the sub-EASMs.

where  $\mathbf{x}$  represents a vector of predictors,  $f$  is the function,  $Y$  is the dependent variable and  $e$  is the error, often assumed to be normally distributed with a mean of zero and variance  $\sigma^2$ . In traditional linear regression,  $f$  is a linear function fitted to the entire data. If there are ' $p$ ' predictor variables then the linear regression model takes the form:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_p x_{pt} + e_t \quad t = 1 \text{ to } N \quad (2)$$

where  $y_t$  is the response variable (JJA precipitation in this paper),  $x_1, \dots, x_p$  are the predictor variables,  $\beta_0, \dots, \beta_p$  are the regression coefficients,  $e_t$  refers to the model error (difference between observed and estimated values), and  $N$  is the total number of observations. The coefficients are estimated by minimizing the sum of squares of the errors (also known as the ordinary least squares method). Key assumptions of this approach are normal distribution of the variables (or transformation to a normal distribution), and uncorrelated and normally distributed errors with a mean of zero. The fitted coefficients and hence, the model, can be used to provide estimates or prediction of the dependent variable,  $y_{pred}$ , at any desired point,  $x_{pred}$ . From linear regression theory, the standard deviation of the estimates,  $\sigma_{pe}$ , can also be obtained (Helsel and Hirsch, 1992). Normal random deviates with mean zero and variance  $\sigma_{pe}^2$  when added to  $y_{pred}$ , provide an ensemble forecast (Singhtrattna *et al.*, 2005). Linear regression approaches are widely used due to well-developed theory and readily available software.

The potential predictors for NEASM and SEASM precipitation are obtained as area-averaged SSTs and OHCs over statistically significant regions during the pre-monsoon season identified and described earlier in Figure 2 and 3 and also listed in Table I. Furthermore, the predictors and response variables are detrended by performing a linear regression with time as the independent variable.

The predictors are correlated among themselves, so a best subset has to be identified. Typically, this is done via the stepwise regression in which different combinations

of predictors are selected and a regression of the form in Equation 2 is fitted and objective score functions such as Mallows'  $C_p$  or predicted residual sums of squares ( $PRESS$ ) are computed and the combination that provides the minimum value is selected (e.g., Helsel and Hirsch, 1992; Neumann *et al.*, 2003).

The Mallows'  $C_p$  statistic is computed as

$$C_p = p + \frac{(n-p)(s_p^2 - \hat{\sigma}^2)}{\hat{\sigma}^2} \quad (3)$$

where  $n$  is the number of observations,  $p$  is the number of explanatory or predictor variables plus one,  $s_p^2$  is the mean square error from the regression model with  $p$  predictor variables, and  $\hat{\sigma}^2$  is the best estimate of the true error also the mean square error from the regression using all the predictors (Helsel and Hirsch, 1992).

The  $PRESS$  is estimated as

$$PRESS = \sum e_i^2 / (1 - h_i) \quad (4)$$

where  $h_i$  is the weight associated with the  $i^{\text{th}}$  data point in estimating itself in the regression – also the diagonal elements of hat matrix (Helsel and Hirsch, 1992).

The combination that gives the least  $C_p$  or  $PRESS$  is selected as the best subset of predictors. Another method often used is the  $F$ -statistic. In this method, an independent variable is introduced into the model. It is kept if it results in a reduced  $F$ -value and if not, it is removed. This process is repeated for all the variables resulting in a best subset. Other measures such as Generalized Cross Validation (GCV) which is similar to  $PRESS$  have also been proposed for best subset selection (e.g., Regonda *et al.*, 2005).

### 6.1. Forecast models

Using the  $C_p$ ,  $PRESS$  and  $F$ -statistic methods, which give the same selection of predictors, SSTs in the TEP, TWP and TIO, and OHC in the TIO (shown in bold boxes in Figure 2(a) and 3(a)) are selected for NEASM

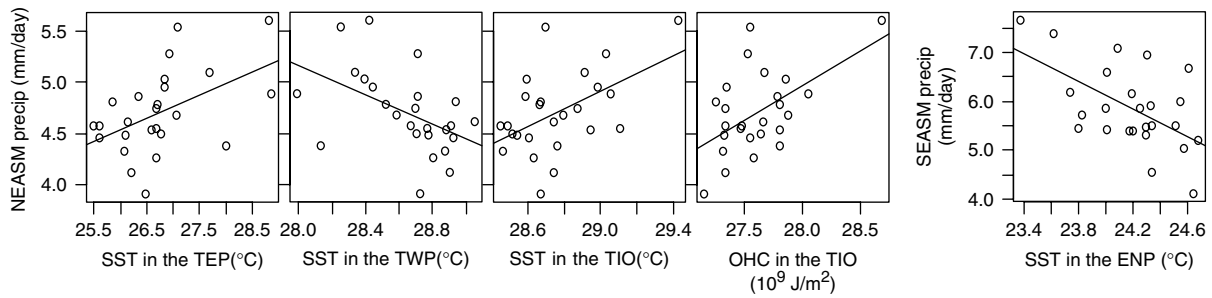


Figure 6. Scatter plots of NEASM precipitation and its selected predictors, i.e. SSTs in the TEP ( $r = +0.49$ ), TWP ( $r = -0.48$ ) and TIO ( $r = +0.50$ ), and OHC in the TIO ( $r = +0.54$ ), and SEASM precipitation and its selected predictor, i.e. SST in the ENP ( $r = -0.56$ ).

precipitation and SST in the ENP (shown in bold box in Figure 2(d)) for SEASM precipitation – these are also indicated in bold in Table I. Figure 6 shows scatter plots of the monsoon precipitation and the selected predictors – strong correlation can be seen.

NEASM precipitation can be predicted from SSTs in the TEP, TWP, and TIO, and OHC in the TIO (Equation 5). SST in the TEP has a negative impact on NEASM precipitation, which shows a significant positive correlation with NEASM precipitation in Figure 6. It may be related to the non-linear relationship between SST in the TEP and NEASM precipitation due to the teleconnections between SST in the TEP and other predictors (i.e. SSTs in the TWP and TIO). The relative importance of the predictors to the overall model is ordered by SST in the TWP ( $-0.48$ ), SST in the TIO ( $0.47$ ), OHC in the TIO ( $0.45$ ), and SST in the TEP ( $-0.44$ ) in terms of the partial correlations in parentheses. The goodness of fit in the overall model shows a significant  $p$ -value at  $\alpha = 0.01$  in terms of  $F$ -statistic; also,  $p$ -values of  $t$ -statistics of all independent variables are significant at  $\alpha = 0.05$ . The correlation and  $R^2$  for the

NEASM precipitation model are 0.72 and 0.52. Adjusted  $R^2$  is 0.42, thus the variance of NEASM precipitation can be explained 42% by the selected predictors in the model.

*NEASM precipitation*

$$= 2.6 - 0.51 \text{ SST in the TEP} - 1.28 \text{ SST in the TWP} + 1.17 \text{ SST in the TIO} + 0.68 \text{ OHC in the TIO} \quad (5)$$

The linear regression model for SEASM precipitation (Equation 6) has one predictor, SST in the ENP. SST and OHC in the WNP are excluded by the stepwise procedures, as they are highly correlated with SST in the ENP, the largest explanatory factor for SEASM precipitation. The overall goodness of fit is statistically significant with the correlation of 0.52 and  $R^2$  of 0.27. Adjusted  $R^2$  is 0.24, thus the predictability of SEASM precipitation is less than that of NEASM precipitation. The higher predictability in the NEASM precipitation model may relate to the strong connection between

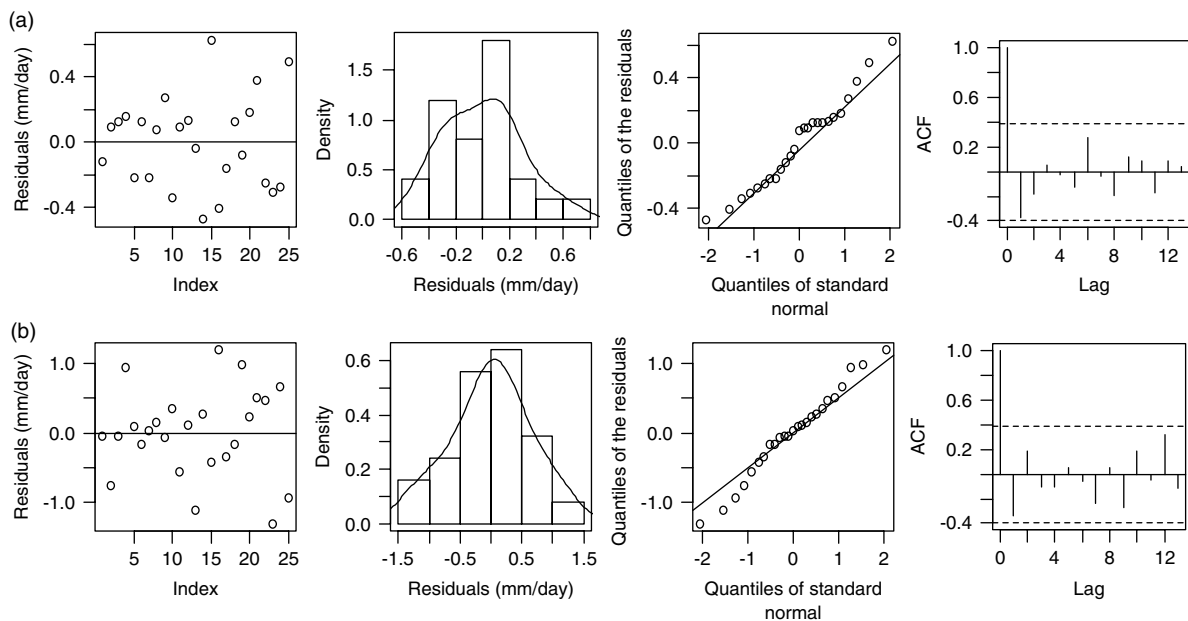


Figure 7. Tests for the normality and autocorrelation of the residuals for the (a) NEASM precipitation and (b) SEASM precipitation models. The dotted lines in the figures of autocorrelation function (ACF) represent the 95% confidence lines.

NEASM precipitation and the heat sources in the TIO that were selected as the predictors for the model. To enhance the predictability of the SEASM precipitation model, other factors such as thermal and land cover indices in the surrounding lands should be considered. Non-linear regression methods also help to enhance the forecasting ability.

#### SEASM precipitation

$$= 34.21 - 1.19SST \text{ in the ENP} \quad (6)$$

#### 6.2. Model results

First, the assumptions of the linear regression model framework are checked. The scatter plot, histogram and probability density function, and quantile-quantile (Q-Q) plot of the residuals of the NEASM precipitation model suggest that they are normally distributed with a mean of zero (Figure 7(a)) also confirmed by Kolmogorov-Smirnov test (von Storch and Zwiers, 2001), and similar observations are found for the residuals of the SEASM precipitation model (Figure 7(b)). Similarly, the residuals show no significant autocorrelation (the last figures in Figure 7(a) and (b)). For the NEASM precipitation model, the lag 1 autocorrelation (the last figure in

Figure 7(a)) is close to the 95% confidence line and the rest are well below the significant lines – indicating they are statistically insignificant. To alter the linear model into one in which the residuals are more independent, *generalized difference procedure* (Pindyck and Rubinfeld, 1981) can be used. The lag 1 correlation was small enough that we felt this modification was not warranted.

To test the predictive capability of the models a cross-validated estimation is performed. There are several measures and ways to evaluate a forecast, but given the short dataset single point cross-validated  $R^2$  is acceptable (e.g., Grantz *et al.*, 2005; Regonda *et al.*, 2006). In this analysis, the observations for a year are dropped and the model is fitted using the remaining observations. The fitted model is then used to estimate (predict) the precipitation value for the dropped year – this is repeated for all the years. The cross-validated estimates and the observed precipitation for the NEASM and SEASM are shown in Figure 8 and 9. The  $R^2$  values of the cross-validated estimates and the observed are 0.21 and 0.16, respectively, for NEASM and SEASM precipitation, which are statistically significant. The root-mean square (RMS) errors for the NEASM and SEASM precipitation models are 0.27 and 0.63 mm/day. They are quite small in comparison to the mean rainfall and

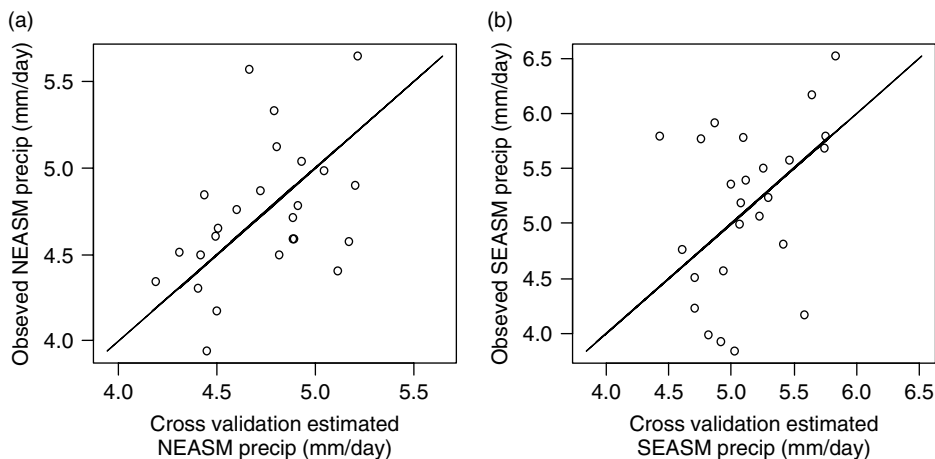


Figure 8. Scatter plots of cross-validated and observed values for (a) NEASM precipitation and (b) SEASM precipitation.

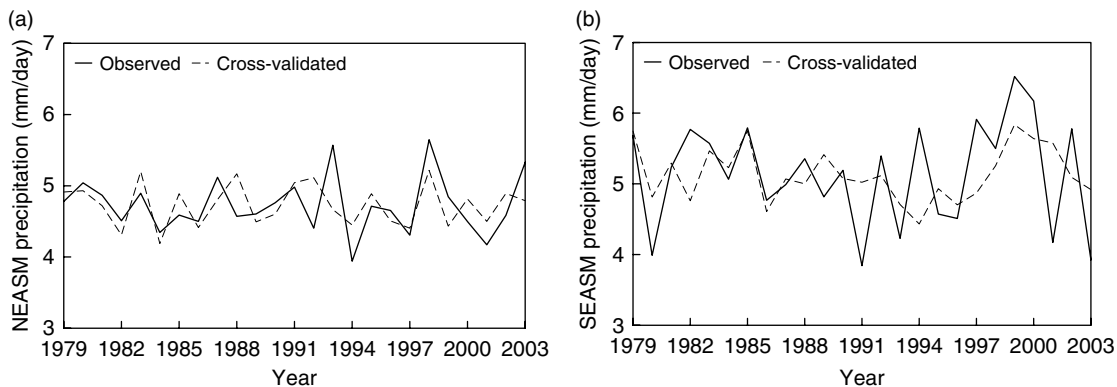


Figure 9. Comparisons of cross-validated values (dotted lines) with observed values (solid lines) for (a) NEASM precipitation and (b) SEASM precipitation.

are comparable to the variances (i.e. mean NEASM and SEASM rainfall is 4.73 and 5.14 mm/day with variances of 0.16 and 0.55 mm/day, respectively). The modeling efforts indicate that there is considerable predictability for the two regional components of the EASM that can be harnessed for resources management and planning.

## 7. Conclusions and remarks

We find that the upper-level divergence zone at 150 hPa forms an East Asian summer monsoonal band including three regional monsoons. The divergence zone shows advances and retreats of the EASM and can be used as an EASM index.

NEASM and SEASM precipitation exhibit different correlation patterns with oceanic heat sources in Pacific and Indian Ocean during the pre-monsoon season. NEASM precipitation has significant correlations with the tropical oceans, which are positive in the tropical eastern Pacific and Indian Oceans and negative in the tropical western Pacific. These correlations support the positive connection between NEASM precipitation and El Niño events. SEASM precipitation has significant correlations with the subtropical oceans, which are positive in the western North Pacific and negative in the eastern North Pacific. This SST dipole mode correlated with SEASM precipitation is the North Pacific SST dipole mode, which is independent of ENSO. The physical relationships between the sub-EASMs and SST anomalies in the tropical and subtropical Pacific are examined. A stronger NEASM is related to the enhanced WNP anticyclonic anomalies and thus more cyclonic anomalies in the NEASM region due to the PEA teleconnection, which is connected to tropical SST anomalies. A stronger SEASM is related to the weakened WNP anticyclonic anomalies and also more cyclonic anomalies in the SEASM region due to the local air–sea interaction, which is also connected to subtropical SST anomalies.

Forecast models of EASM precipitation are developed on the basis of the heat sources in the tropical Pacific and Indian Oceans for the NEASM precipitation model and in the subtropical Pacific for the SEASM precipitation model. Forty-two percent of the variance of NEASM precipitation and 24% of SEASM precipitation variance is explained by the selected predictors in each model. Model results by statistical inspections show the normality and independence of the residuals, and cross-validated estimation verifies good agreements between the estimated values of NEASM and SEASM precipitation and their observed precipitation.

EASM precipitation showed a non-linear relationships with the heat sources in the surrounding oceans, so non-linear regression methods should be applied for further study. More advanced tools other than single-point correlation would help to capture the relations more precisely. Often regressions are not stationary with time and will have to be constantly monitored (e.g., Kumar *et al.*, 1999). Thermal factors and land cover indices in

the surrounding lands could be potential predictors to enhance the predictabilities of the EASM precipitation models. To examine the general relationships between the EASM and oceanic heat sources, we observed an East Asian region that includes three different frontal zones, i.e. *Mei-yu*, *Changma*, and *Baiu*. However, the impact of ENSO on the EASM is usually complex and is often characterized by strong regional features (e.g., Lau and Weng, 2001; Yang and Lau, 2006).

## Acknowledgements

We wish to thank the three anonymous reviewers for valuable suggestions, and Dr. Young-Oh Kwon from Climate and Global Dynamic Division, NCAR, for reading the manuscript. The first author is thankful to CIRES for a graduate student fellowship. Partial support of this work by National Science Foundation via grant ATM-0437538 is also thankfully acknowledged.

## References

- Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P-P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E. 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology* **4**: 1147–1167.
- Bell GD, Halpert M. 2004. State of the climate in 2003. *Bulletin of the American Meteorological Society* **85**: S17–S18.
- Chang C-P, Zhang Y, Li T. 2000a. Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part I: roles of the subtropical ridge. *Journal of Climate* **13**: 4310–4325.
- Chang C-P, Zhang Y, Li T. 2000b. Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part II: meridional structure of the monsoon. *Journal of Climate* **13**: 4326–4340.
- Chase TN, Knaff JA, Pielke RA, Kalnay E. 2003. Changes in global monsoon circulations since 1950. *Natural Hazards* **29**: 229–254.
- Deser C, Blackmon ML. 1995. On the relationship between tropical and North Pacific sea surface temperature variations. *Journal of Climate* **8**: 1677–1680.
- Gandin LS. 1963. Objective analysis of meteorological fields. *Gidrometeor. Izd.*, (Translated by Israel Program for Scientific Translations, Jerusalem, 1965), 242.
- Grantz K, Rajagopalan B, Clark M, Zagana E. 2005. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research* **41**: W10410, Doi:10.1029/2004WR003467.
- Hansen J, Ruedy R, Glascoe J, Sato M. 1999. GISS analysis of surface temperature change. *Journal of Geophysical Research* **104**: 30977–31022.
- Helsel DR, Hirsch RM. 1992. *Statistical Methods in Water Resources*. Elsevier Science BV: Amsterdam.
- Kallummal R, Kirtman BP. 2006. Validity of a linear-stochastic view of ENSO in a CGCM. Center for Ocean-Land-Atmosphere Studies (COLA) Technical Report 216: Calverton, Maryland; 36.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Roplewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**: 437–471.
- Kumar KK, Rajagopalan B, Cane MA. 1999. On the weakening relationship between the Indian monsoon and ENSO. *Science* **284**: 2156–2159.
- Lau K-M, Weng H. 2001. Coherent modes of global SST and summer rainfall over China: an assessment of the regional impacts of the 1997 ~ 98 El Niño. *Journal of Climate* **14**: 1294–1308.
- Lee E-J, Jhun J-G, Park C-K. 2005. Remote connection of the northeast Asian summer rainfall variation revealed by a newly defined monsoon index. *Journal of Climate* **18**: 4381–4393.

- Levitus S, Antonov JI, Boyer TP, Stephens C. 2000. Warming of the world ocean. *Science* **287**: 2225–2229.
- Neumann DW, Rajagopalan B, Zagona EA. 2003. Regression model for daily maximum stream temperature. *Journal of Environmental Engineering* **129**: 667–674.
- Pindyck RS, Rubinfeld DL. 1981. *Econometric Models and Economic Forecasts*, 2nd edn. McGraw-Hill: New York.
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research* **108**(D14): 4407, Doi:10.1029/2002JD002670.
- Regonda SK, Rajagopalan B, Clark M, Zagona E. 2006. A multi-model ensemble forecast framework: application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research* **42**: W09404, Doi:10.1029/2005WR004653.
- Regonda SK, Rajagopalan B, Lall U, Clark M, Moon Y-I. 2005. Local polynomial method for ensemble forecast of time. Special issue on “Nonlinear deterministic dynamics in hydrologic systems: present activities and future challenges”. *Nonlinear Processes in Geophysics* **12**: 397–406.
- Shen S, Lau K-M. 1995. Biennial oscillation associated with the East Asian summer monsoon and tropical sea surface temperature. *Journal of the Meteorological Society of Japan* **73**: 105–124.
- Singhrattana N, Rajagopalan B, Clark M, Krishna KK. 2005. Seasonal forecasting of Thailand summer monsoon rainfall. *International Journal of Climatology* **25**: 649–664.
- Tao SY, Chen LX. 1987. A review of recent research on the East Asian summer monsoon in China. In *Monsoon Meteorology*, Chang CP, Krishnamurti TN (eds). Oxford University Press: New York; 60–92.
- von Storch H, Zwiers FW. 2001. *Statistical Analysis in Climate Research*. Cambridge University Press: Cambridge.
- Wang B, Lin H. 2002. Rainy season of the Asian-Pacific summer monsoon. *Journal of Climate* **15**: 386–398.
- Wang B, Li T. 2004. East Asian monsoon-ENSO interactions. In *East Asian Monsoon*, Chang CP (ed.). World Scientific Publishing: Singapore; 177–212.
- Wang B, Wu R, Fu X. 2000. Pacific-East Asian teleconnection: how does ENSO affect East Asian climate? *Journal of Climate* **13**: 1517–1536.
- White WB, Cayan DR, Dettinger MD, Auad G. 2001. Sources of global warming in upper ocean temperature during El Niño. *Journal of Geophysical Research* **106**: 4349–4367.
- Wu Z-X, Newell RE. 1998. Influence of sea surface temperature on air temperature in the tropics. *Climate Dynamics* **14**: 275–290.
- Wu R, Wang B. 2002. A contrast of the East Asian summer monsoon-ENSO relationship between 1962–77 and 1978–93. *Journal of Climate* **15**: 3266–3279.
- Yang S, Lau WK-M. 2006. Interannual variability of the Asian monsoon. In *The Asian Monsoon*, Wang B (ed.). Praxis Publishing: Chichester; 259–293.
- Zhang R, Sumi A, Kimoto M. 1996. Impact of El Niño on the East Asian Monsoon: a diagnostic study of the ‘86/87 and ‘91/92 events. *Journal of the Meteorological Society of Japan* **74**: 49–62.
- Zhou W, Chan JCL. 2007. ENSO and the South China Sea summer monsoon onset. *International Journal of Climatology* **27**: 157–167.