Exploring the Predictability of 30-Day Extreme Precipitation Occurrence Using a Global SST–SLP Correlation Network

MENGQIAN LU AND UPMANU LALL

Department of Earth and Environmental Engineering, and Columbia Water Center, Columbia University, New York, New York

JAYA KAWALE

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

STEFAN LIESS

Department of Soil, Water, and Climate, University of Minnesota, St. Paul, Minnesota

VIPIN KUMAR

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

(Manuscript received 16 May 2014, in final form 2 November 2015)

ABSTRACT

Correlation networks identified from financial, genomic, ecological, epidemiological, social, and climatic data are being used to provide useful topological insights into the structure of high-dimensional data. Strong convection over the oceans and the atmospheric moisture transport and flow convergence indicated by atmospheric pressure fields may determine where and when extreme precipitation occurs. Here, the spatiotemporal relationship among sea surface temperature (SST), sea level pressure (SLP), and extreme global precipitation is explored using a graph-based approach that uses the concept of reciprocity to generate cluster pairs of locations with similar spatiotemporal patterns at any time lag. A global time-lagged relationship between pentad SST anomalies and pentad SLP anomalies is investigated to understand the linkages and influence of the slowly changing oceanic boundary conditions on the development of the global atmospheric circulation. This study explores the use of this correlation network to predict extreme precipitation globally over the next 30 days, using a logistic principal component regression on the strong global dipoles found between SST and SLP. Predictive skill under cross validation and blind prediction for the occurrence of 30-day precipitation that is higher than the 90th percentile of days in the wet season is indicated for the selected global regions considered.

1. Introduction

The application of complex networks in climate science (Tsonis and Roebber 2004; Tsonis et al. 2006, 2011) is an emerging field that has provided some novel insights into the nature of Earth systems. This paper explores whether extreme precipitation occurrence in some parts of the world may be predictable over the next 30 days, given information about stable, teleconnected patterns of sea surface temperature (SST) and sea level pressure (SLP). Recently developed methods that identify a connected graph from climate data are used to identify a correlation network from pentad SST and SLP data. The strong dipoles from this correlation network are then used as predictors in a regression model to assess where the occurrence of extreme precipitation may be predictable over the next 30 days.

Lorenz (1996) reviewed the progress of estimates of the deterministic predictability of the atmosphere and noted that the estimated error doubling times have actually decreased from 5 days in 1966, based on Charney et al.'s (1966) work, to about 1.5 days in 1995, based on

Corresponding author address: Mengqian Lu, Department of Earth and Environmental Engineering, Columbia University, 918 S. W. Mudd Hall, Mail Code: 4711, 500 West 120th Street, New York, NY 10027.
E-mail: lumengqian.lulu@gmail.com

DOI: 10.1175/JCLI-D-14-00452.1

the ECMWF model. While this seems to suggest that the limits of deterministic predictability based on the growth of small random errors may only be 5 days or so, Lorenz noted that the SST evolves rather slowly and that even atmospheric phenomena such as the quasi-biennial oscillation (QBO) and ocean–atmosphere phenomena such as El Niño–Southern Oscillation (ENSO) may have significant predictability for days to seasons, as long as they stay in the same mode. Given this context, a question that motivates this work is whether a global correlation network of SST and SLP can inform the prediction of precipitation extremes over the next 30 days, at least at some places in the world.

The links between the most important oceanic climate indices [e.g., Atlantic multidecadal Oscillation (AMO), Pacific–North American pattern (PNA), ENSO, North Atlantic Oscillation (NAO), and Madden–Julian oscillation (MJO)] and precipitation at various space and time scales have been examined by several authors (Ropelewski and Halpert 1986, 1987; Leathers et al. 1991; Huntington et al. 2004; Trigo et al. 2004; Sutton and Hodson 2005; Kim et al. 2006; Cañón et al. 2007; Jeong et al. 2008). Indeed, the pioneering studies by Walker (1923, 1924, 1928) and Walker and Bliss (1930, 1932, 1937), which first documented ENSO on a global scale, were motivated by attempts to understand and predict variations in Indian monsoon rainfall.

Given these potential sources of predictability, various studies have been done on rainfall and temperature prediction (Wang and Warner 1988; Buizza et al. 1999; Chen and Hwang 2000; Silverman and Dracup 2000; Barnston et al. 2005; Lin et al. 2005; Abukhshim et al. 2006; Smith et al. 2007; Schepen et al. 2012; Badr et al. 2014; Song et al. 2014; Nicholson 2014). The most widely studied temporal scales are seasonal to interannual (Diallo et al. 2013; Landman et al. 2012; Schepen et al. 2012; Charles et al. 2013; Jiang et al. 2013; Lee et al. 2013; Singh et al. 2013; Kang et al. 2014; Nicholson 2014) and weather forecasts, through either quantitative precipitation forecasts (QPFs) or numerical weather prediction (NWP) (e.g., Lin et al. 2005), typically up to 7 days, and radar-based nowcasting (Ligda 1953; Golding 1998; Wilson et al. 1998; Foresti et al. 2015; Olsson et al. 2014; Sokol et al. 2013; Dai et al. 2014), that is, a short-term forecast, usually a few hours. The statistics of precipitation at seasonal to interannual time scales have been shown to have some usable predictability for regions that have strong teleconnections to the ENSO phenomena. Reasonable skill from QPF is demonstrated for lead times of 0–72 h in most places. However, there is a dearth of literature exploring predictability between the two time scales, and there is interest in subseasonal to seasonal predictability (Brunet et al. 2010; Vitart et al. 2012).

Recently, atmospheric rivers (Zhu and Newell 1998; Bao et al. 2006) have been widely studied as the major contributor to extreme precipitation and floods (Ralph et al. 2006; Leung and Qian 2009; Lavers et al. 2011, 2013; Ralph and Dettinger 2011). Extreme precipitation in the midlatitudes is often associated with anomalous atmospheric moisture transported from tropical and subtropical warmer oceanic areas. Lu et al. (2013) identified the major moisture sources for the 1995 January flood in western France as the Gulf of Mexico and tropical North Atlantic Ocean east to the Bahamas, and demonstrated the predictability of the extreme precipitation given midlatitude SLP fields. The atmospheric circulation patterns led to a coherent and persistent transport of moisture from these sources, with convergence of flow and precipitation over a large region covering western France for a period of nearly a month. Nakamura et al. (2013) provided a similar analysis for the Ohio River basin, where a persistent dipole in the SLP led to the wavelike transport of moisture from the Gulf of Mexico into the Ohio River basin every 4–7 days over the March–May season (Nakamura et al. 2013).

Several research groups are considering climate as a network of dynamical systems and are applying ideas from graph theory to a global dataset to study its collective behavior (Hsieh 2001; Tsonis and Roebber 2004; Tsonis et al. 2006, 2011; Yamasaki et al. 2008; Donges et al. 2009; Steinhaeuser et al. 2011, 2012; Berezin et al. 2012). The connections of these climate networks to ENSO, NAO, and so forth have been studied by these authors at interannual time scales. Berezin et al. (2012) considered daily surface temperature and the 850-hPa geopotential height fields in this context and identified a robust global network pattern that reflects physical coupling across the two fields that is stable over time. A significance of the findings from this literature is that there are certain locations whose climatic evolutions at different time scales tend to be more connected to each other, and potentially more predictable. Climate scientists have traditionally analyzed the global climate data fields using principal component analysis (PCA). At one level, the correlation network literature can be viewed as a similar dimension reduction method. However, it differs in the criteria used to derive the reduced representation and in the objective of finding a graphically connected system that represents the spatiotemporal associative memory and multiple variables' coexpression rather than explaining the maximal correlation in the full dataset. The particular method used here considers correlation associated with locations in space but also with time delays, and hence it differs in that regard from PCA, which typically examines only the spatial correlation structure.

We use the spatiotemporal correlation network algorithm developed and documented by (Kawale et al. 2013) to explore the spatiotemporal relationships among the climate variables of interest, and study the following questions:

1) Do correlation networks that can be identified from global SST and SLP fields at a pentad resolution inform global precipitation extremes over the subsequent month?
2) Are there regions in the world where the dipoles identified from the SST–SLP analysis provide robust predictability of precipitation extremes?
3) How does this predictability compare with what is achieved using traditional principal component analysis applied to global SLP or SST data?

This paper is organized as follows. Section 2 describes the data. The approach and methodology are presented in section 3. The results of the 1) exploration of the global dipole networks between SST and SLP, 2) examples of using dipole signals for regional monthly precipitation extremes predictability for different seasons, and 3) the model assessments are shown and discussed in section 4. The final section has a summary of the work.

2. Data

For our analysis we considered the NCEP–NCAR reanalysis dataset (Kalnay et al. 1996, available online at http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html) and the CPC Merged Analysis of Precipitation (CMAP) dataset (Xie and Arkin 1997, available online at http://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html) provided by the NOAA/OAR/ESRL Physical Sciences Division (PSD; http://www.esrl.noaa.gov/psd). The primary variables we used are SST anomalies and SLP anomalies from each pentad mean from the NCEP–NCAR reanalysis dataset and pentad mean precipitation from the CMAP dataset. The anomalies were calculated by subtracting the corresponding pentad climatology values for each grid. All variables have a grid resolution of 2.5° latitude × 2.5° longitude. The analysis and modeling were done using data from 1985 to 2011. Additionally, we assessed the skill of the model fit to the abovementioned data in a blind prediction mode using data from 2012 to 2014.

3. Approach

We start with a summary of the important steps of our approach and map them to the subsections that describe the crucial methods.

(i) Dipole identification (section 3a):
   1) Construct lagged negative correlation network, dipoles, between global SST and SLP.
   2) Filter network to retain strong dipole regions.

(ii) Principal component analysis (section 3b) of filtered dipole SST and dipole SLP:
   1) A PCA of the strong dipole SST and SLP fields separately.
   2) Comparison of the correlations (section 3c) with seasonal precipitation extremes of these dipole PCs with those from PCA with the full global fields.

(iii) Generalized linear regression for predicting occurrence of precipitation extremes:
   1) Logistic regression using standardized dipole SST principal components (PCs) 1–30 and dipole SLP PCs 1–30 as candidate predictors to predict the occurrence of precipitation extremes at different regions in their respective wet seasons (section 3d).
   2) For performance evaluation, the dependence between successive forecasts of the binary precipitation extreme variable is considered using a bootstrapping and leave-one-year-out cross-validation setup such that no serially correlated samples are selected in any cross-validation set (section 3e).
   3) Model performance is assessed in terms of the deviance statistic, error rate in the bootstrapping and leave-one-year-out cross-validation mode, and blind prediction using data from 2012 to 2014 (section 3f).

a. Dipole identification using spatiotemporal correlation network

We first use the spatiotemporal correlation network algorithm developed and documented by Kawale et al. (2013) to identify the time-lagged bivariate dipole networks between global SST anomalies and global SLP anomalies. We use data from 1985 to 2011 to train the dipole networks without dividing them into different seasons, because preliminary analyses showed no significant differences in the seasonality of the dipole networks between the two variables. Thus, the analysis is a priori limited to the annual dipoles. The algorithm is based upon the concept of reciprocity to generate cluster pairs of locations such that two locations within the same cluster have similar lagged correlations with the members of other clusters. The identification of these cluster pairs is done considering different time lags. The lagged correlation networks are built upon both positive and negative lagged correlations. The algorithm first constructs the complete directed lag graph based on the lagged correlations and then computes the clusters by performing a sequence of

thresholding, *k*-nearest neighbor (KNN) filtering, reciprocity, and identification of shared nearest neighbors. Finally, cluster pairs are identified as the ones that maximize a lagged correlation. Liess et al. (2014) demonstrated an application of the algorithm for identifying teleconnection patterns based on negative correlations around Australia. The constructed networks exhibit a high degree of local clustering but a small number of long-range connections. The details of each step of the algorithm are presented below.

1) STEP 1: CONSTRUCT COMPLETE DIRECTED LAG GRAPH

For each time lag l, a directed weighted graph $\mathbf{G}^l = (\mathbf{V}, \mathbf{E}^l)$ is constructed with nodes $\mathbf{V} = \{v_1, \ldots, v_n\}$ representing the spatial grids of interest; $\mathbf{E}^l$ is the matrix of directed edges in the graph; each entity E^l_{ij}, $1 \le i, j \le n$ is the edge strength linking location v_i to v_j and E^l_{ij} is directed; and $E^l_{ij} \neq E^l_{ji}$, since it specifies that the ith location's variable is leading that at the jth location by time l. The mathematical expression of E^l_{ij} is as follows:

$$E^l_{ij} = \frac{\sum_{u=1}^{m-l} (x_i^u - \overline{x}_i) \cdot (x_j^{u+l} - \overline{x}_j)}{[(x_i - \overline{x}_i)^{\mathrm{T}}(x_i - \overline{x}_i) \cdot (x_j - \overline{x}_j)^{\mathrm{T}}(x_j - \overline{x}_j)]^{1/2}}, \quad (1)$$

where x_i is the time series at location i, m is the length of the time series, l is the lag, and x_i^u is the value of x_i at time u.

The graph $\mathbf{G}^l = (\mathbf{V}, \mathbf{E}^l)$ is further separated into two groups—one positive, $\mathbf{G}^l = (\mathbf{V}, \mathbf{E}^{l+})$, and the other one negative, $\mathbf{G}^{l-} = (\mathbf{V}, \mathbf{E}^{l-})$—by separating $\mathbf{E}^l$ into two groups as follows:

$$E^{l+}_{ij} = \begin{cases} E^l_{ij}, & \text{if } E^l_{ij} > 0 \text{ and } E^l_{ij} > E^l_{ji} \\ 0, & \text{otherwise} \end{cases} \text{ and} \quad (2)$$

$$E^{l-}_{ij} = \begin{cases} E^l_{ij}, & \text{if } E^l_{ij} < 0 \text{ and } E^l_{ij} < E^l_{ji} \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

This separation process also eliminates the edges whose strength is not significant by setting their edge weight to be zero. The separation deals with the polarity of the lag correlation, which conveys different meanings: positive lag correlation indicates the traveling of a phenomenon from one location to another, whereas negative lag correlation indicates a receiver of information. Together the pair specifies a lagged dipole.

2) STEP 2: CONSTRUCT RECIPROCAL LAG GRAPH

Based on the directed positive and negative edge matrices ($\mathbf{E}^{l+}$ and $\mathbf{E}^{l-}$, respectively), a KNN list for any location node v_i is constructed by sorting all edges of v_i based on edge weights and then picking the top K nodes with edge weights from v_i. The KNN list for location node v_i is denoted as $\mathbf{S}^{lo+}$, $\mathbf{S}^{li+}$, $\mathbf{S}^{lo-}$, $\mathbf{S}^{li-}$. Terms $\{\mathbf{S}^{lo+}, \mathbf{S}^{li+}\}$ are from $\mathbf{E}^{l+}$; $\{\mathbf{S}^{lo-}, \mathbf{S}^{li-}\}$ are from $\mathbf{E}^{l-}$. For example, $\mathbf{S}^{lo+}$ is constructed by picking all outward edges from v_i and then sorting the edges from highest to lowest based on the edge weights in $\mathbf{E}^{l+}$ and then picking the top K edges. The superscripts "o" and "i" represent "outgoing" from and "incoming" to any location l, respectively. Kawale et al. (2011) demonstrated the effect of choosing different K values ($K = 25$ vs $K = 100$) on the final network. In this paper, we choose $K = 50$ as suggested by Kawale et al. (2011, 2013). The following steps of the algorithm ensure that the final dipole network is less sensitive to the choice of K.

The KNN list $\{\mathbf{S}^{lo+}, \mathbf{S}^{li+}, \mathbf{S}^{lo-}, \mathbf{S}^{li-}\}$ is the foundation for reciprocal lag graphs or matrices. The concept of directed reciprocity (Kawale et al. 2011, 2013) helps remove spurious correlations while considering the effect of spatial autocorrelation on the interconnections: graph nodes a and b are said to be connected by the directed (time lagged) reciprocal edge $a \rightarrow b$ if and only if a lies in b's *in*-KNN list and b appears in a's *out*-KNN list. Based on this directed reciprocity concept, the reciprocal graph $\mathbf{G}^l = (\mathbf{V}, \mathbf{E}^l)$ is constructed as follows:

$$E^r_{ij} = \begin{cases} 1, & \text{if } v_i \in S^{li+}_j \text{ and } v_j \in S^{lo+}_i \\ -1, & \text{if } v_i \in S^{li-}_j \text{ and } v_j \in S^{lo-}_i \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

where E^r_{ij} is the entity of $\mathbf{E}^r$. Note that only one of the three conditions can be satisfied for any location node.

3) STEP 3: CLUSTERING THE GRAPH NODES

The shared reciprocal nearest neighbor (SRNN) (Kawale et al. 2011) is used to cluster the nodes. It constructs SRNN graphs or matrices $\{\mathbf{G}^{so+}, \mathbf{G}^{si+}, \mathbf{G}^{so-}, \mathbf{G}^{si-}\}$ by computing the edge weights between two nodes based on how many edges they have in common. For example, $\mathbf{G}^{so+} = (\mathbf{V}, \mathbf{E}^{so+})$ is constructed by considering the positive (with edge weight 1) outgoing edges in $\mathbf{E}^r$. Mathematically,

$$\begin{aligned} E^{so+}_{ij} &= |\{v_k : \forall k, E^r_{ik} = 1 \text{ and } E^r_{jk} = 1\}|, \\ E^{si+}_{ij} &= |\{v_k : \forall k, E^r_{ki} = 1 \text{ and } E^r_{kj} = 1\}|, \\ E^{so-}_{ij} &= |\{v_k : \forall k, E^r_{ik} = -1 \text{ and } E^r_{jk} = -1\}|, \text{ and} \\ E^{si-}_{ij} &= |\{v_k : \forall k, E^r_{ki} = -1 \text{ and } E^r_{kj} = -1\}|. \end{aligned} \quad (5)$$

The clustering is based upon these four SRNN graphs $\{\mathbf{G}^{so+}, \mathbf{G}^{si+}, \mathbf{G}^{so-}, \mathbf{G}^{si-}\}$. The idea is to find cluster pairs A and B such that the locations in A have a lot of outgoing edges to locations in B in the graph $\mathbf{G}^{so}$ and locations in B should have a lot of incoming edges from A in the graph

$\mathbf{G}^{si}$ (Kawale et al. 2013). The SRNN ensures that two locations within one end of a dipole not only share high positive correlations due to their spatial autocorrelation but also share high negative correlations with the members of the other end of the dipole. This process is based upon the edge densities $\{\mathbf{D}^{so+}, \mathbf{D}^{si+}, \mathbf{D}^{so-}, \mathbf{D}^{si-}\}$ for the graphs $\mathbf{G}^{so}$ and $\mathbf{G}^{si}$, and each of the positive and negative correlations uses a local density attractor algorithm (Kawale et al. 2011). For example, the edge density for negative correlations is computed by taking the sum of the edge weights at each node in the graph as follows:

$$\forall_{i\in V} D_i^{so-} = \forall_{j\in V} \sum_{j=1}^{N} E_{ij}^{so-} \quad \text{and}$$

$$\forall_{i\in V} D_i^{si-} = \forall_{j\in V} \sum_{j=1}^{N} E_{ij}^{si-}. \tag{6}$$

The first step is to find clusters based on densities $\{\mathbf{D}^{so+}, \mathbf{D}^{si+}, \mathbf{D}^{so-}, \mathbf{D}^{si-}\}$ in each of the two graphs $\mathbf{G}^{so}$ and $\mathbf{G}^{si}$ using the local attractor algorithm, which iteratively attaches nodes in the graph to the local attractors, which are determined as points with the highest density in the neighborhood of locations; then the second step is to find the links across the clusters that have reciprocal edges. Note that the algorithm does allow a fixed edge threshold to be forced into the construction of edges. However, a very low edge threshold will not generate too many edges either, as the algorithm is built upon reciprocity and clustering that consider both inter- and intraconnections of a dipole. The spurious edges are eliminated through the iterations, and the final clusters that are considered to form a dipole are entitled with strong connections with the other end of a dipole. We use -0.01 to eliminate very small floating-point numbers. A higher threshold will yield fewer dipoles. A more detailed illustration of the steps from merging the shared nearest neighbors graphs to form the shared reciprocal nearest neighbors can be found in Kawale et al. (2011, 2013).

The algorithm can identify both positive and negative lagged correlations. We only focus on the negative lagged correlation, which specifies a lagged dipole (Kawale et al. 2013).

We constructed the time-lagged dipole networks between global SST and SLP, with SST leading SLP for up to 30 days using the pentad average data. The dipole networks identify strong linked regions such that the leading SST signals (areas in red in Fig. 1) are sent out to the receiving SLP areas (in blue), and the regions are linked by colored lines indicating their dipole strength, that is, the correlation between the two identified clusters.

The spatiotemporal networks constructed at different lags are further cleaned by retaining only the dipole regions whose dipole strength exceeds the 80th percentile of dipole strength at each time lag. This helps to identify the primary signals across the globe by eliminating weak to moderate dipoles and eliminating random connections. Since the cleaning is done at each time lag, the process also allows for the identification of the dipoles' persistency with time.

b. *Principal component analysis*

With the aim of understanding the leading modes of the strong dipole patterns and their relationship to global precipitation extremes, a PCA of the strong dipole [i.e., identified dipole regions between SST_{t-l} and SLP_t fields, with $l = 5, 10 \ldots, 30$ days (5 days corresponding to 1 pentad data point)], whose dipole strength exceeds the 80th percentile, was used. This criterion was chosen to select an adequate sample size while retaining potentially important features; 427 and 452 grid boxes ($2.5° \times 2.5°$) were selected for the SLP and SST regions, respectively. The PCA was done separately on the strong dipole SST regions and strong dipole SLP regions, to obtain two sets of principal components that extract the leading patterns of the interconnected 1) slowly changing boundary conditions (SST) and 2) responses of atmospheric circulation (SLP).

The pentad SST and SLP anomaly fields for the selected strong dipole areas contain a total of P spatial grids. Here, P is the number of grids retained after the selection process. Let $\mathbf{Z}$ be a $N \times P$ data matrix, where N is the total number of pentad data points from 1985 to 2011, so $N = 73 \times 27 = 1971$. For each single spatial field, each column is standardized (column mean subtracted and then dividing by the column standard deviation). Since $\mathbf{Z}$ has more columns than rows, a singular value decomposition is used to extract the principal components. The PCA of $\mathbf{Z}$ is then a factorization of the form

$$\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^{\mathrm{T}}, \tag{7}$$

where $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ and $\mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}$. The columns of $\mathbf{U}$ are orthonormal eigenvectors of $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$, the columns of $\boldsymbol{W}$ are orthonormal eigenvectors of dimensions of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$, and $\mathbf{\Sigma}$ is a diagonal matrix containing the square roots of eigenvalues from $\mathbf{U}$ or $\mathbf{W}$ in descending order, and the dimensions of $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{W}^{\mathrm{T}}$ are $N \times N$, $N \times P$, and $P \times P$, respectively. The analysis partitions a field into orthogonal (independent) modes. The eigenvalues provide a measure of the variance explained by each mode. The temporal and spatial variabilities of the field are isolated and are represented by the columns of $\mathbf{U}$ and $\mathbf{W}$, respectively. From this point, we will use dipole SST/SLP PCs to refer to the PCs extracted from pentad SST and SLP anomaly fields of the selected strong dipole regions that exceed the 80th percentile of the dipole strength. The

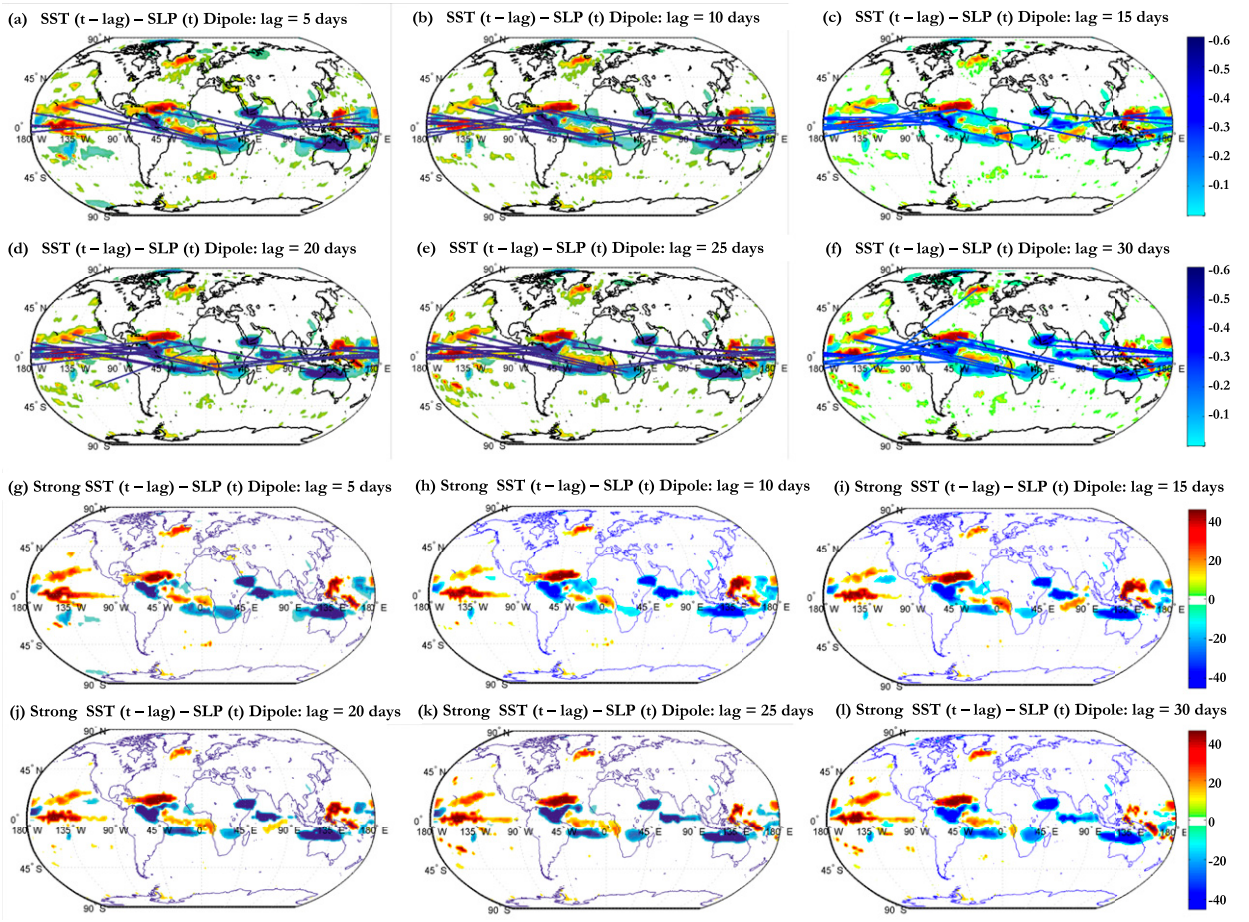

FIG. 1. Spatiotemporal dipoles between SST and SLP with different time lags, from 5 to 30 days. (a)–(f) The red regions represent outgoing SST signals, while the blue are receiving SLP regions. The edges link the centers of clusters identified as connected. The average correlation between the two clusters is shown in blue as the edge strength, and data are pentad normalized anomalies. (g)–(l) Selected strong dipole regions for different lags. The threshold is chosen as the 80% percentile of dipole strength. Persistence of strong dipole regions is observed and concentrated in the tropical region.

PCA is also done on global SST and SLP anomaly fields (80°S–80°N, all longitudes; excluding the North and South Poles), to extract the leading modes (referred to as global SST/SLP PCs) in order to provide a comparison to the strong dipole SST/SLP PCs in terms of their correlations with global precipitation extremes as follows.

c. *Correlation maps with precipitation extremes*

We constructed the time-lagged correlation maps between the climate variables' PCs (i.e., SST and SLP) with the binary global precipitation extremes defined as follows [Eq. (8)]. The first three PCs from the dipole SST and dipole SLP fields were compared with their counterparts from the global SST and SLP fields. The first three PCs were selected because they explained much of the total variance of their fields: 32% and 29% of dipole SLP and SST fields, and 21% and 20% of global SLP and SST fields. A binary precipitation extremes indicator (PE) was calculated for each 30-day period over the entire globe for each season, as follows:

$$\mathrm{PE}_{i,t}^{S} = \begin{cases} 0, & \text{if} \quad p_{i,t}^{S} < p_{i}^{S}* \\ 1, & \text{if} \quad p_{i,t}^{S} \geq p_{i}^{S}* \end{cases}, \tag{8}$$

where $\mathrm{PE}_{i,t}^{S}$ is the binary indicator for precipitation extreme over the subsequent 30 days, at location (grid) i, time step t, in season $S = 1, 2, 3$, and 4, corresponding to the seasons 1) December–February (DJF), 2) March–May (MAM), 3) June–August (JJA), and 4) September–Nov (SON); $p_{i,t}^{S}$ is the sum of the pentad mean precipitation over the subsequent 30 days from CMAP, at location (grid) i, time step t, in season S; and $p_{i}^{S}*$ is the 90% percentile of all subsequent 30-day precipitation totals at location i, in season S, over the entire 27 years of data (1985–2011). Note that there is an overlapping period of the pentad data between any two consecutive PE,

and the resulting serial correlation of PE is considered in the model fitting setup in section 3e by using bootstrapping in the cross-validation sampling process.

The point biserial correlation (Kornbrot 2014) maps are constructed 1) between the dipole SST/SLP PCs and the binary global PE and 2) between the global SST/SLP PCs and the binary global PE, separately, with the SST's and SLP's PCs 5 days (1 pentad data point) ahead of the 30-day PE. The point biserial correlation is the value of Pearson's product moment correlation when one of the variables is dichotomous, taking on only two possible values coded 0 and 1, and the other variable is metric (Kornbrot 2014). It is defined as follows:

$$\gamma(\mathrm{PC}_j, \mathrm{PE}_i) = \frac{\overline{\mathrm{PC}_j^{\mathrm{PE}_i=1}} - \overline{\mathrm{PC}_j^{\mathrm{PE}_i=0}}}{\sigma_{\mathrm{PC}_j}} \sqrt{\frac{N_{\mathrm{PE}_i=1} N_{\mathrm{PE}_i=0}}{N_{\mathrm{PE}_i}(N_{\mathrm{PE}_i}-1)}}, \tag{9}$$

where PE_i is the time series of the binary variable defined in Eq. (8) for 30-day precipitation extremes at location i; PC_j is the jth PC of either dipole SST/SLP fields or global SST/SLP fields, 5 days (1 pentad data point) ahead of PE_i; $\overline{\mathrm{PC}_j^{\mathrm{PE}_i=1}}$ and $\overline{\mathrm{PC}_j^{\mathrm{PE}_i=0}}$ are the means of the jth PCs coded 0 and 1 by PE_i, respectively; $N_{\mathrm{PE}_i=1}$ and $N_{\mathrm{PE}_i=0}$ are the number of observations coded by $\mathrm{PE}_i = 1$ and $\mathrm{PE}_i = 0$; N_{PE_i} is the total number of observations, $N_{\mathrm{PE}_i=1} + N_{\mathrm{PE}_i=0}$; and σ_{PC_j} is the standard deviation of the jth PC. The significance test for the point biserial correlation is the same as that for the Pearson correlation. More detailed information can be found in Kornbrot (2014). Note that the effective sample size—that is, the degree of freedom—for the significance test is adjusted with consideration for the sequential dependency of the PCs, with a reduction to $M[(1-\alpha_1)/(1+\alpha_1)]$ (Mitchell et al. 1966), where M is the total number of samples and α_1 is the lag-1 autocorrelation coefficient.

All correlation maps are constructed excluding the North and South Poles—that is, 80°S – 80°N, all longitudes. The purpose of constructing the two sets of correlation maps is to compare their spatial correlations (primary dipole SST/SLP PCs and global SST/SLP PCs, respectively) with the binary global precipitation extremes and see whether the dipole SST/SLP PCs correlate better or worse with global precipitation extremes compared to the global SST/SLP PCs, especially for certain regions of interest.

d. LPCR—Predictability of precipitation extremes

The first 30 PCs extracted from strong dipole SST and SLP fields—that is, dipole SST PCs 1–30 or dipole SLP PCs 1–30—are used as candidate predictors for global PE over the subsequent month [Eq. (8)], with a lead time of 5 days (1 pentad point), that is, $\mathrm{PC}_{t-1}^{\mathrm{SLP}}$ and $\mathrm{PC}_{t-1}^{\mathrm{SST}}$ for PE_t for the next 30 days. Many methods have been proposed to address the question of component selection (Preisendorfer and Mobley 1988; Jolliffe 2002; Jackson 2005). These methods may be categorized as either heuristic or statistical approaches (Jackson 1993). Our approach is heuristic, based on a combination of the graphical scree test and the assessment of the proportion of the total variance explained by the principal components retained in the model (e.g., Jackson 1993). The first 30 PCs of the strong dipole SST and SLP fields explain 73% and 81% of the total variance of the two fields, respectively. Taking seasonality of precipitation into consideration, the logistic principal component regression (LPCR) is applied at different geographical locations for different seasons (i.e., DJF, MAM, JJA, and SON) to demonstrate the predictability of global precipitation extreme occurrence over the subsequent month using the extracted dipole SST/SLP PCs.

For each season S, and each location i, we fit a logistic regression model (Bishop 2006), with standardized PC candidates (subtracted by their means and divided by their standard deviation). It is mathematically defined as follows:

$$\mathrm{logit}(y_{i,t}^S) = \ln \frac{y_{i,t}^S}{1-y_{i,t}^S} = \mathbf{X}_{t-1}^S \boldsymbol{\beta}_i^S \quad \text{and}$$
$$y_{i,t}^S = \Pr(\mathrm{PE}_{i,t}^S = 1), \tag{10}$$

where $y_{i,t}^S$ is the probability of the next 30 days' precipitation total exceeding the 90% percentile of the season's 30-day precipitation total at location i at time step t; $\mathbf{X}_{t-1}^S$ is a predictor matrix whose first column is the unit vector and subsequent vectors represent all the candidate predictors—that is, the standardized dipole SST PCs 1–30 and SLP PCs 1–30—for the corresponding S leading $y_{i,t}^S$ by 5 days. Term $\mathbf{X}_{t-1}^S$ is the same for all the locations at which forecasts are made. The vector $\boldsymbol{\beta}_i^S$ contains the intercept and regression coefficients for all predictors estimated using maximum likelihood. Note that we first fit the two sets of LPCR models with y_t^S covering the entire globe (80°S–80°N, all longitudes) for different seasons (i.e., DJF, MAM, JJA, and SON) separately with 1) $\mathbf{X}_{t-1}^S$ being the first 30 PCs of selected dipole SST fields and 2) $\mathbf{X}_{t-1}^S$ being the first 30 PCs of the selected dipole SLP fields. This step assists to select the regions of interest that have strong association with dipole SST/SLP PCs (significant LPCR coefficients) and prescreen the predictors for the selected regions for each season. Only statistically significant (p value < 0.05, at the 95% significance level) predictors for the target seasons and regions of interest are retained in the final model: y_t^S only covering the region of interest for each season and $\mathbf{X}_{t-1}^S$ only containing the preselected subset of the first 30 PCs of the selected dipole SST and SLP

fields. Both steps estimate the associated $\boldsymbol{\beta}_i^S$ using the maximum likelihood criterion.

e. Bootstrapping and cross validation

Since forecasts are made for every 30-day period using predictor data lagged by 5 days, forecast performance evaluation periods are overlapping and may overly emphasize the same information, since successive values of PE_t will be potentially correlated. The effect of serial correlation of PE is considered in the setup of fitting LPCR to include bootstrapping (Efron 1982) to generate nonserially correlated samples, together with leave-one-year-out cross validation. The bootstrapping and cross-validation process is described as follows:

Let year κ move from 1985 to 2011, and for each S of interest,

1) Leave 1 year out: Choose 26 years from 1985 to 2011 excluding year κ, which will be used to test the performance of the model built from these 26 years.
2) Draw nonoverlapping bootstrap samples for each year selected:
   (i) Randomly draw a starting position SP in index 1: 6 within each year and within each season. This is a pentad index.
   (ii) For that year and season select a paired sample of {**X**, PE} values for selected grids using the pentad indices { SP, SP + 6, SP + 12} to assure that there is no overlapping 30-day period in the model fitting or validation set.
3) Repeat the previous two steps 100 times to generate 100 bootstrap samples that are designed so that there is no serial correlation due to overlapping of the data used in model building or validation.

LPCR models are fit for each S with this bootstrapping and cross-validation setup, and the statistics associated with the model fitting are analyzed cross all 27×100 samples.

f. Performance assessment through deviance, error rate, and blind prediction

We use deviance and error rate to evaluate the goodness of fit of the fitted logistic models with the cross-validation and bootstrapping sampling technique described above. In addition, the entire approach is assessed with blind prediction using data from 2012 to 2014 (Michaelsen 1987; Elsner and Schmertmann 1994). Deviance is used widely as a measurement of the lack of fit to the data in a logistic regression model. Here, deviance is calculated within the bootstrapping and leave-one-year-out cross-validation process. We have 27 years of data from 1985 to 2011. Moving sequentially from 1985 to 2011, 100 models are fit using nonserially correlated samples without the data of that year and then used to estimate the observations for that year given that year's predictors. The average deviance of S's models against the null model, Δ^S, is then computed using these estimated values. The deviance is averaged across all the selected spatial grids and models fitted using bootstrapped samples with leave-one-year-out cross validation. The key equations are provided as follows:

$$\Delta^S = \frac{-\sum_{i=1}^{N_S}\sum_{\theta=1}^{N_\theta} 2\times \ln\left[\dfrac{L(\hat{\boldsymbol{\beta}}_i^{S,\theta})}{L_{\mathrm{null}}^{S,\theta,i}}\right]}{N_S \times N_\theta} = \frac{2\times \sum_{i=1}^{N_S}\sum_{\theta=1}^{N_\theta} [\ell_{\mathrm{null}}^{S,\theta,i} - \ell(\hat{\boldsymbol{\beta}}_i^{S,\theta})]}{N_S \times N_\theta},$$

$$L(\hat{\boldsymbol{\beta}}_i^{S,\theta}) = \prod_{t=1}^{N_T^{S,\theta}} (\hat{y}_{i,t}^{S,\theta})^{\mathrm{obs}_{i,t}^{S,\theta}} (1-\hat{y}_{i,t}^{S,\theta})^{(1-\mathrm{obs}_{i,t}^{S,\theta})},$$

$$L_{\mathrm{null}}^{S,\theta,i} = \prod_{t=1}^{N_T^{S,\theta}} (0.1)^{\mathrm{obs}_{i,t}^{S,\theta}} (0.9)^{(1-\mathrm{obs}_{i,t}^{S,\theta})},$$

$$\ell(\hat{\boldsymbol{\beta}}_i^{S,\theta}) = \sum_{t=1}^{N_T^{S,\theta}} [\mathrm{obs}_{i,t}^{S,\theta} \times \ln(\hat{y}_{i,t}^{S,\theta}) + (1-\mathrm{obs}_{i,t}^{S,\theta}) \times \ln(1-\hat{y}_{i,t}^{S,\theta})], \quad \text{and}$$

$$\ell_{\mathrm{null}}^{S,\theta,i} = \sum_{t=1}^{N_T^{S,\theta}} [\mathrm{obs}_{i,t}^{S,\theta} \times \ln(0.1) + (1-\mathrm{obs}_{i,t}^{S,\theta}) \times \ln(0.9)], \tag{11}$$

TABLE 1. Summary of the selected location for LPCR models for different seasons and the chosen informative predictors with their estimated coefficients averaged over the selected grids (irregular areas from the location below due to selection process) and over the bootstrapped cross validation process.

Season	Location	Predictors and average coefficients	$\chi^2_{k^S}$ test p values
DJF	North Africa 20°N – 45°N, 5°W–20°E	SLP PC1 (−1.14); SLP PC2 (+0.92)	$<10^{-5}$
MAM	Central America 0°–20°N, 90°–40°W	SLP PC1 (−1.33); SLP PC2 (+1.48); SLP PC3 (−1.42); SST PC4 (+1.13)	$<10^{-5}$
JJA	Central Africa 15°S–10°N, 10°–50°E	SLP PC1 (+1.33); SLP PC2 (−1.45); SLP PC3 (+1.16); SST PC4 (−2.28)	$<10^{-5}$
SON	East Brazil 15°–5°S, 60°–45°W	SLP PC1(−1.37); SLP PC2 (+1.34); SLP PC17 (−1.31); SST PC1 (+1.94); SST PC3 (+1.59)	$<10^{-5}$

where $L(\hat{\boldsymbol{\beta}}_i^{S,\theta})$ is the likelihood of the fitted model $\theta = 1, \ldots, N_\theta$ for location $i = 1, \ldots, N_S$, in S; $L_{\text{null}}^{S,\theta}$ is the likelihood of the null model; $\ell(\hat{\boldsymbol{\beta}}_i^{S,\theta})$ and $\ell_{\text{null}}^{S,\theta}$ are their corresponding log likelihood. Term $N_\theta = 2700$ is the total number of models for each season, N_S is the total spatial grids selected for S, $\hat{y}_{i,t}^{S,\theta}$ is the estimated probability of extreme precipitation over the subsequent 30 days at location i in S for sample $t = 1, \ldots, N_T^{S,\theta}$, $N_T^{S,\theta}$ is the sample size of the testing set (the observations of the year that were left out of the model fitting) for S and model θ, $\text{obs}_{i,t}^{S,\theta}$ is the corresponding observation of precipitation extreme in the testing set—that is, $\text{obs}_{i,t}^{S,\theta} = 1$ if the subsequent 30 days' precipitation total exceeds the seasonal 90% percentile threshold for a 30-day rainfall amount. The average difference of deviance against the null model (no predictor) is assessed on a chi-square distribution $\chi^2_{k^S}$ with k^S degree of freedom (Hosmer and Lemeshow 2004; Seager et al. 2007). Here k^S is the number of predictors (Table 1) for S plus 1.

We also assessed the error rates in the leave-one-year-out cross-validation process to assist in the evaluation of the predictability of precipitation extremes with the LPCR for each season. The error rate is calculated as the proportion of predictions in which the model assigns $\Pr(\text{PE}_{i,t}^S = 1) > 0.5$ to 0 observations ($\text{PE}_{i,t}^S = 0$) or $\Pr(\text{PE}_{i,t}^S = 1) < 0.5$ to 1 observations ($\text{PE}_{i,t}^S = 1$). Error rates assist the measurement of goodness of fit with a comparison to the baseline in which the null model predicts 1 with $\Pr(\text{PE}_{i,t}^S = 1)_{\text{null}} = 0.1$ and 0 with $\Pr(\text{PE}_{i,t}^S = 0)_{\text{null}} = 0.9$, as the extreme precipitation is defined as events exceeding the seasonal 90% percentile; the corresponding baseline error rate here is $1 - 0.1 \times 0.1 - 0.9 \times 0.9 = 18\%$.

The model fit is also tested with the most recent data (2012–14) with a blind prediction process. The error rates of the blind prediction are tabulated in Tables 2e–f) and compared to those for the training period (Tables 2a–d).

4. Results

a. *Time-lagged dipole network between SST and SLP*

The global dipole networks between the SST and SLP fields with different time lags (SST leading SLP) are shown in Figs. 1a–f). The red regions in Figs. 1a–f represent outgoing SST signals, while the blue areas are receiving SLP regions. The edges link the centers of clusters identified as connected. The average correlation between the connected clusters is shown in blue as the edge strength, that is, dipole strength. The key findings are as follows:

1) The dipole networks are identified are persistent through 30 days, indicating persistent informative signals—that is, a strong spatiotemporal relationship between linked dipole SST clusters and dipole SLP clusters. As expected from the construction, this is a sparse network.
2) The persistent outgoing/leading SST signals are concentrated in the tropical regions, consistent with the knowledge that most of the climate signals are from the tropical ocean, and are possibly the dominant drivers of climate variability at this time scale, throughout the year. This suggests that coupled ocean–atmosphere variability and the time-lagged responses of atmosphere to the variance of the ocean play a dominant role in the tropics on the 30-day time scale we presented here.
3) Persistent outgoing SST dipole areas are in (i) the tropical Atlantic ocean, consistent with the region Lu et al. (2013) identified as the major tropical moisture sources for western France precipitation extremes and Nakamura et al. (2013) identified as important for Ohio River floods; (ii) the equatorial Pacific Ocean at the Niño-3.4 region, which is well identified as classifying ENSO episodes; and (iii) the west equatorial Pacific at the Mariana Trench, east to the Philippines and Indonesia.

TABLE 2. Summary of prediction accuracy probabilities for different seasons: (a) DJF, (b) MAM, (c) JJA, and (d) SON. (e)–(h) Summary of the blind prediction's performance (error rates and its 95% uncertainty intervals) using data from 2012 to 2014.

(a) DJF		Observations 0	Observations 1
Predictions	1	0.007[a][0, 0.003]	0.022[b]
	0	0.893[b]	0.078[a][.067, 0.08]

(b) MAM		Observations 0	Observations 1
Predictions	1	0.02[a][.003, 0.1]	0.036[b]
	0	0.88[b]	0.064[a][.048, 0.07]

(c) JJA		Observations 0	Observations 1
Predictions	1	0.017[a][.002, 0.07]	0.03[b]
	0	0.88[b]	0.07[a][.053, 0.077]

(d) SON		Observations 0	Observations 1
Predictions	1	0.002[a][.003, 0.1]	0.029[b]
	0	0.90[b]	0.071[a][.054, 0.076]

Baseline		Observations 0	Observations 1
Predictions	1	0.09[a]	0.01[b]
	0	0.81[b]	0.09[a]

(e) DJF	Observations 0	Observations 1
Error rates	0.006 [0, 0.041]	0.1 [.096, 0.1]

(f) MAM	Observations 0	Observations 1
Error rates	0.04 [.003, 0.17]	0.1 [.09, 0.16]

(g) JJA	Observations 0	Observations 1
Error rates	0.167 [.042, 0.4]	0.069 [.050, 0.08]

(h) SON	Observations 0	Observations 1
Error rates	0.027 [.002, 0.14]	0.080 [.070, 0.082]

[a] Mean error rate for each scenario: the sum of the entities in each season is the error rate, and the 95% uncertainty interval is in the bracket. Entities larger than the baseline table's entities indicate worse than the null model.

[b] Accuracy rate: the sum of the entities in each season is the accurate prediction rate. Entities larger than the baseline table's entities indicate better than the null model.

4) The receiving SLP regions are also quite persistent for different lags. They include northern South America, the tropical Indian Ocean, East Africa, and northern Australia and its nearby Indian Ocean region.

We eliminated weak to moderate signals that are less than the 80th percentile of the dipole strength at each time lag. The resulting strong dipole network is shown in Figs. 1g–l. The strong dipole regions (blue and red areas in Figs. 1g–l) remain almost the same from the 5- to the 30-day lag, consistent with what we observed from Figs. 1a–f.

b. Correlation maps with precipitation extremes

PCA as described in section 3b is used to extract the leading patterns of the interconnected 1) slowly changing boundary conditions (SST) and 2) responses of atmospheric circulation (SLP) from the selected strong dipole network to generate dipole SST PCs and dipole SLP PCs. The same analysis is done on the global SST and SLP fields to generate global SST PCs and global SLP PCs.

We then constructed time-lagged correlation maps between the primary climate variables' PCs (i.e., SST and SLP) with the binary global PE, defined above in Eq. (8): 1) dipole SLP PCs 1–3 versus PE (Figs. 2a–c); 2) global SLP PCs 1–3 versus PE (Figs. 2d–f); 3) dipole SST PCs 1–3 versus PE (Figs. 2g–i); and 4) global SST PCs 1–3 versus PE (Figs. 2j–l). All the correlation maps are constructed with the SST and SLP PCs 5 days (1 pentad data point) ahead of PE. From the correlation maps with the first three PCs of each of the field in Fig. 2, we can see that the strong dipole PCs appear to be more informative in terms of this metric and we subsequently use them for prediction with the logistic regression. Though we only provide the comparison of the first three PCs for illustration, the first 30 PCs are used as candidate predictors in the LPCR model described in section 3d, and a corresponding comparison of forecast skill using PCs of both global SST/SLP fields and dipole SST/SLP are also pursued to examine the merits of using dipole predictors for prediction.

Comparing Figs. 2a–c to Figs. 2d–f, the patterns are similar between the two sets for strong correlations in the maps with dipole fields. Both the dipole SLP PCs and global SLP PCs show high correlations with central Africa, Central America, Brazil, India, and southeastern China. The correlation strengths associated with these

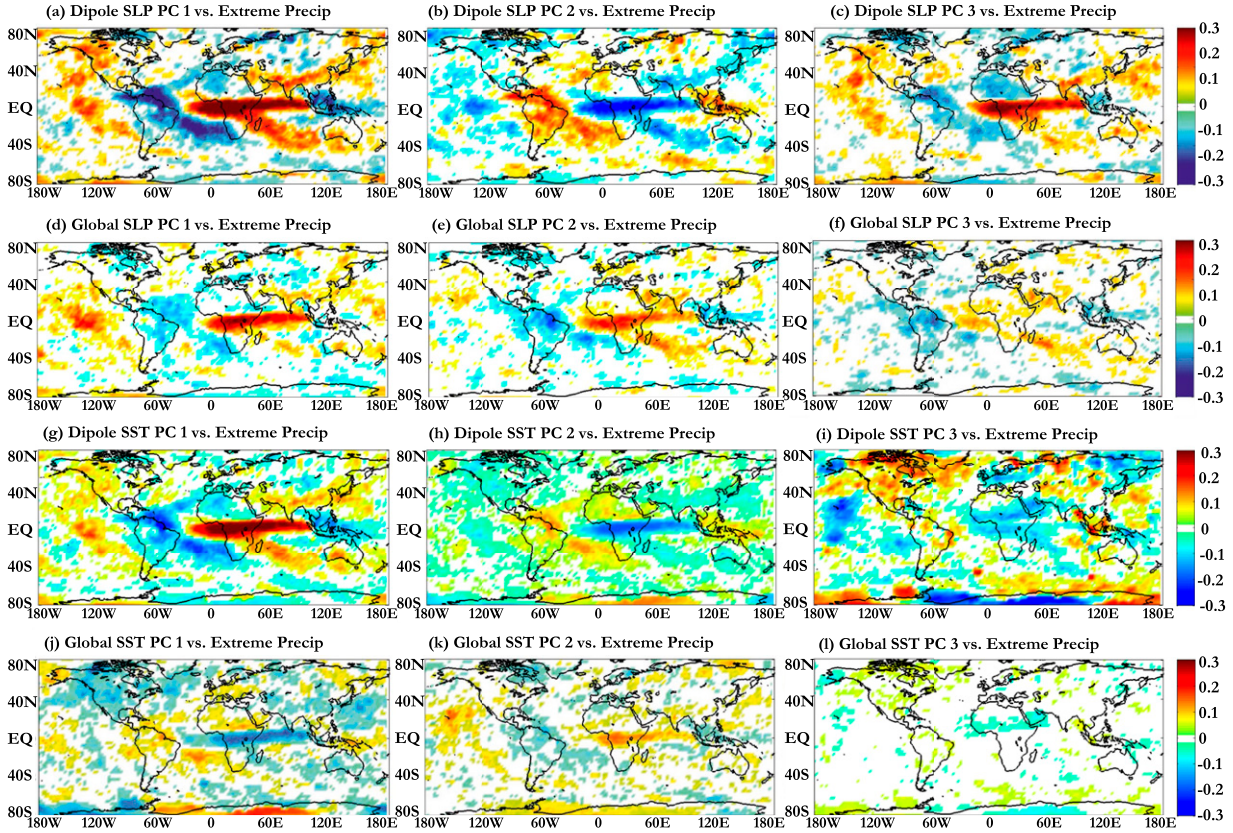

FIG. 2. (a)–(c) Correlation maps between 5-day-ahead dipole SLP regions' PCs 1–3 and global PE over the subsequent month; the three PCs explain 16%, 9%, and 7% of the total variance, respectively. (d)–(f) Correlation maps between 5-day-ahead global SLP regions' PCs 1–3 and global PE over the subsequent month; the three PCs explain 9%, 7%, and 5% of the total variance, respectively. (g)–(i) Correlation maps between 5-day-ahead dipole SST regions' PCs 1–3 and global PE over the subsequent month; the three PCs explain 12%, 10%, and 7% of the total variance, respectively. (j)–(l) Correlation maps between 5-day-ahead global SST regions' PCs 1–3 and global PE over the subsequent month; the three PCs explain 8%, 7%, and 5% of the total variance, respectively. For the 95% significance level (p value $<$ 0.05), the effect of the highest serial correlations of SST (0.83) and SLP (0.5) on the effective degree of freedom of the significance test is considered; the effective degree of freedom for SST and SLP are 189 and 660, respectively. The significance level of SST and SLP are 0.119 and 0.064, respectively.

identified regions are higher in the dipole correlation maps, indicating a potentially better prediction skill than for global SLP PCs. The North Africa, Indonesia, and Thailand regions are significantly correlated to the dipole SLP PCs but not to the global SLP PCs. The strengths of the correlations are strong for the first three dipole SLP PCs, while they degrade rapidly with the global SLP PCs. Similar results are observed in Figs. 2g–l for the SST PCs. The dipole SST PC correlation maps (Figs. 2g–i) share some regions with the global SST PCs maps (Figs. 2j–l), and they have overall higher correlations with precipitation extremes. While both Fig. 2g and Fig. 2j indicate correlations with extreme precipitation in central Africa, Central America, North America, Indonesia and Thailand, and China, the dipole SST PC1 shows much stronger correlation with these regions, especially central Africa, Central America, and Brazil. These are subsequently chosen as examples for prediction with logistic regression for different seasons—specifically, we consider the following examples: central Africa (JJA), Central America (MAM), and Brazil (SON).

c. *Seasonal predictability of precipitation extremes*

We fit the LPCR models for different seasons with the first 30 PCs of the selected dipole SST and dipole SLP fields as candidate predictors. The predictors included in the final models are summarized in Table 1. The areas of extreme precipitation to predict as examples are chosen based on an initial exploratory analysis (described in section 3d) of the predictability of precipitation extremes over the subsequent month using the candidate PC predictors. For each grid point (2.5° × 2.5° grid, 10 512 grid boxes in total), in the exploratory analysis, an

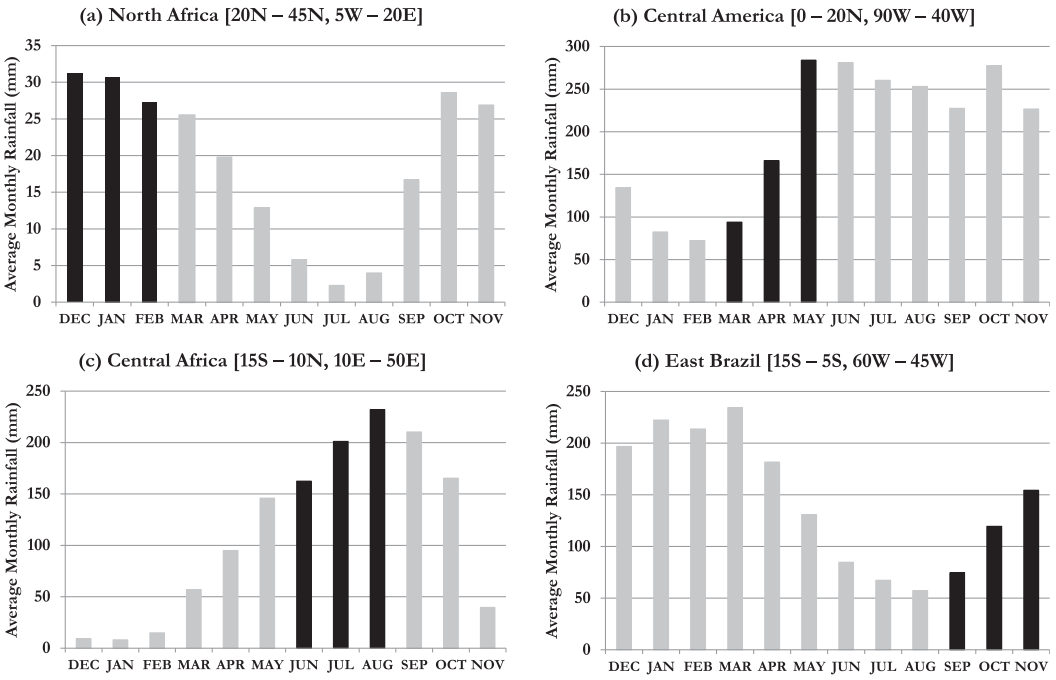

FIG. 3. Average monthly rainfall (1900–2009) of the selected regions for different seasons: (a) DJF, North Africa (20°–45°N, 5°W–20°E); (b) MAM, Central America (0°–20°N, 90°–40°W); (c) JJA, central Africa (15°S–10°N, 10°–50°E); and (d) SON, east Brazil (15°–5°S, 60°–45°W); the information is from the World Bank Group website (http://sdwebx.worldbank.org/climateportal/). The selected seasons' rainfall amounts for the four regions are plotted in black.

LPCR model is fit to each candidate predictor set. For each of these locations, we identify the wet season. For each wet season, we chose the areas that have significant coefficients (at the 95% level) for the leading predictors, and then fit an LPCR model using the predictors identified (Table 1) from the global exploratory analysis. The chosen areas might be irregular due to the selection process; that is, the chosen grids might be subset areas from the candidate rectangular regions in Table 1. The results for different seasons for the example locations are shown next.

1) DJF

The selected region for this season is the North Africa–Mediterranean area (20°–45°N, 5°W– 20°E). The total number of selected grid boxes is 53, a subset area from the abovementioned box. DJF is the wet season for the area according to the average monthly rainfall (Fig. 3a) based on data from 1900 to 2009. The statistically significant predictors for the area are the standardized dipole SLP PC1 and PC2. Some dipole SST PCs also show their strong association with the area's precipitation extremes. However, the likelihood-ratio test showed that the model fit using dipole SLP PC1 and PC2 is the best. The dipole SLP PC1 and PC2 explain 16% and 9% of the total variance of the SLP dipole field, respectively. The estimated coefficients associated with the two PCs have opposite signs, PC1 (−1.14) and PC2 (+0.92), indicating their changes have the opposite effect on the odds ratio of the probability of having precipitation extremes over the subsequent month. The fitted model passed the χ^2_{ks} test with p value $< 10^{-5}$ (Table 1) against the null model in the bootstrapped cross-validation process. The average error rate of the prediction in this season is 8.4%, less than half of the baseline error rate (18%) of the null model. Table 2a tabulated the prediction accuracy/errors of the LPCR model with bootstrapped cross validation, with the sum of the error entities as the error rate of the season—that is, prediction = 1 with observation = 0 and prediction = 0 with observation = 1, the sum of accuracy entities as the prediction rate of precipitation extremes [defined in Eq. (8)] over the subsequent month in the season. The accurate predictions are both larger than the baseline probability of the null model, that is, the fitted LPCR achieves 0.89 (0.022) against a baseline of 0.81 (0.01). The wrong predictions are both smaller than the baseline probability of the null model; that is, the fitted LPCR achieves 0.007 (0.078) against 0.09 (0.09).

2) MAM

A Central American region (30 grid boxes out of the box 0°–20°N, 90°–40°W) is selected to assess the predictability of using dipole signals for precipitation extremes over the subsequent 30 days in MAM, which corresponds to the beginning and peak of the wet season for the region (Fig. 3b). Four predictors are included in the final LPCR model: dipole SLP PCs 1–3 and dipole SST PC4. The coefficients of dipole SLP PCs 1–3 have different signs—PC1 (−1.33), PC2 (+1.48), and PC3 (−1.42)—as summarized in Table 1. The first three PCs of the dipole SLP have strong year-round correlations (Figs. 2a–c) with the precipitation extremes in Central America. The signs of the coefficients for PC1, PC2, and PC3 are all consistent with their year-around correlations (Figs. 2a–c). The fitted model passed the test with p value $< 10^{-5}$ (Table 1) against the null model in the cross-validation process. The average error rate of the prediction in this season is 8.3%, again much better than the null model's error rate. Table 2b tabulated the prediction accuracy/errors of the LPCR model for precipitation extremes over the subsequent month in MAM. The accurate predictions are both larger than the baseline probability of the null model; that is, the fitted LPCR achieves 0.88 (0.036) against a baseline of 0.81 (0.01). The wrong predictions are both smaller than the baseline probability of the null model; that is, the fitted LPCR achieves 0.02 (0.064) against 0.09 (0.09).

3) JJA

This is the wet season for central Africa (15°S–10°N, 10°–50°E). The total number of selected grid boxes is 52 from the abovementioned box. Four predictors are included in the final LPCR model: dipole SLP PCs 1–3 and dipole SST PC4 (Table 1). Table 1 summarizes the estimated coefficients for the four predictors of the fitted LPCR models, averaged over the selected grids and over the leave-one-year-out cross validation. The estimated coefficients are all statistically significant (p value < 0.05). The coefficients of dipole SLP PCs 1–3 have alternating signs—PC1 (+1.33), PC2 (−1.45), and PC3 (+1.16)—which is consistent with their year-round spatial correlation, shown in Figs. 2a–c. The fitted model passed the test with p value $< 10^{-5}$ (Table 1) against the null model in the cross-validation process. The average error rate of the prediction in this season is 8.7% (baseline error rate is 18%). Table 2c tabulated the prediction accuracy/errors of the LPCR model for precipitation extremes over the subsequent month for this season. The probability of the fitted model telling extreme when the observation is nonextreme is 0.017, which is better than the null model's 0.09, while the probability of the fitted model telling nonextreme when the observation is actually extreme is 0.07, better than the baseline of the null model, 0.09. The probability of the fitted model telling extreme—consistent with the observation, that is, 0.03—is twice better than the null model's baseline, 0.01. Finally, the model is better than the null model in terms of predicting nonextremes as having a probability of 0.88 and a baseline of 0.81.

4) SON

SON is the beginning of the wet season in east Brazil (15°–5°S, 60°–45°W). The total number of selected grid boxes is 106; their precipitation extremes are selected as predictands for the model. Five predictors are included in the final LPCR model: dipole SLP PCs 1, 2, and 17 and dipole SST PCs 1 and 3 (Table 1). Table 1 summarizes the estimated coefficients for the predictors in the fitted LPCR models, averaged over the selected grids and over the bootstrapped leave-one-year-out cross validation. The fitted model passed the test with a p value $< 10^{-5}$ (Table 1) against the null model in the cross-validation process. The coefficients of dipole SLP PC1 and PC2 have opposite signs, PC1 (−1.37) and PC2 (+1.34), which is the same as their year-round spatial correlation, shown in Figs. 2a,b. The opposite signs indicate their changes have the opposite effect on the odds ratio of the probability of having precipitation extremes over the subsequent month. The coefficients of dipole SST PC1 and PC3 also have the same signs, PC1 (+1.94) and PC3 (+1.59), which is different from their year-round spatial correlation, shown in Figs. 2g,i. The average error rate of the prediction in this season is 7.3%, less than half of the baseline error rate, 18%. Table 2d tabulates the prediction accuracy/errors of the LPCR model for precipitation extremes over the subsequent month for this season. The predictions are superior to the baseline probability of the null model; that is, the fitted LPCR achieves 0.9 (0.029) against a baseline of 0.81 (0.01). The probability of the fitted model telling extreme when the observation is nonextreme is 0.002, better than the null model's 0.09, while the probability of the fitted model telling nonextreme when the observation is actually extreme is 0.071, less than the baseline of the null model, 0.09.

d. Blind prediction

In addition to the cross-validated and bootstrap verifications, we felt that it was important to see whether the relationships were robust in a completely out of sample fitting and prediction setting. Given the short dataset, a limited experiment was performed. We used each of the models fitted above to the 1985–2011

SST, SLP, and precipitation data and checked the forecast skill for the same seasons and locations as above for data from 2012 to 2014, which was not used in model fitting. The pentad dipole SST/SLP fields for 2012–14 are projected to the orthogonal PC bases derived from 1985 to 2011. The selected predictors for each region and season presented are used in the models presented in section 4c and Table 1 to predict the 30-day extreme precipitation. By comparing with the observations from 2012 to 2014, the error rates are summarized in Tables 2e–h. We note that the patterns in forecast skill identified originally are similar to those noted under this blind prediction setting.

5. Summary and discussion

This paper presents a novel application of the correlation network extraction algorithm to the identification of stable, correlated SST and SLP patterns using pentad data and considering persistent correlation up to at least 30 days. The motivation was to see whether the identification of such a pattern could provide useful forecasts of the probability of extreme rainfall at any region in the world over the ensuing 30-day period. Such a forecast would be potentially very useful for flood preparation at a lead time that is well beyond the lead time of meteorological forecasts, and it corresponds to a gap in predictability between QPF and seasonal to interannual climate prediction.

We focused on the identification of dipoles between SST and SLP with the idea that the warm SSTs that indicate strong, persistent convection would then link to SLPs where there is a strong, persistent low pressure, which increases the chance of extreme precipitation. Tropical SST and SLP pairs were identified where there were strong regional dipoles, as defined by a contiguous spatial cluster of SST being negatively correlated with a contiguous spatial cluster of SLP. An arbitrary threshold corresponding to the 80th percentile of global dipole strength across clusters was used to prune the network, and then the principal components of the SST and of the SLP fields associated with the subset of retained dipole grids were used as potential predictors of 30-day precipitation that exceeds the 90th percentile of the climatological seasonal 30-day precipitation totals. A number of regions across the world were identified where there may be potential skill in forecasting 30-day extreme precipitation, defined as precipitation in excess of the 90th percentile of that 30-day period. Of these, one example was selected for each season, and its forecasting skill was assessed under leave-one-year-out cross validation. The misclassification rate in these cross-validated forecasts ranged from 7.3% to 8.7% (compared to the baseline error rate, 18%, of the null model) in the example, and each of the models was found to be different from the null model at a p value less than 10^{-5} based on the chi-square test applied to the model deviance under cross validation.

A comparison of the correlations of the strong dipole PCs and of the global field PCs with extreme precipitation is provided, and it shows that the correlations of the dipole PCs are uniformly stronger. Consequently, it appears that an approach toward dimension reduction using the correlation networks rather than a brute force PCA on the global gridded fields of SST and SLP could potentially be more effective in building empirical models for extended range forecasting of extreme precipitation. A corresponding comparison of forecast skill using the first 30 PCs of global SST and SLP fields as candidate predictors was also pursued; the global SST/SLP PCs are uninformative (p value > 0.05) in the fitting of the logistic regression model. Of course, issues of nonstationarity and nonlinearity were not addressed in this work, and a rather specific application of correlation networks was pursued. We recognize that for a nonlinear or nonstationary system, a finite sample may not be able to capture the full state space of the underlying system. Thus, 1) our model may not be able to effectively forecast in regions of the state space that have not been visited in the record used and 2) if we subsample the past record to construct the predictors, it is also possible that the prediction skill in the blocks of data used for validation will suffer since the construction of the correlation network and the PCA used to develop the predictors will not have sampled this part of the state space. We believe that these issues apply to any statistical model that relies on a dataset with a limited exploration of the underlying state space. Given these caveats, we believe that the cross validation of the forecast part of the model is justified. We note that the record used is quite short in terms of number of years, even though the sample size in terms of the number of pentads is reasonably large. These issues, as well as a test of correlation networks and their predictability in SST, SLP, and precipitation fields simulated by model integrations from phase 5 of CMIP (CMIP5) for historical and climate change scenarios, still need to be performed.

An alternate formulation of the approach could be to consider a specific region where the prediction of precipitation extremes is of interest and to explore a correlation network that is relevant directly to it. Our interest in this paper was to screen globally for SST/SLP dipole patterns at the pentad to monthly time scales using the basic correlation network algorithm, and to then see if the patterns identified have any utility for the prediction of precipitation extremes anywhere in the

world. This is a first step in an exploratory process of model building and for making the case for potential predictability of precipitation extremes over the ensuing 30-day period.

Acknowledgments. This work is part of the Global Flood Initiative of the Columbia Water Center. It has been supported by the National Research Foundation of Korea by a grant funded by the Korean government (NRF-2010-220-D00083); by NOAA Grants NA10OAR4310159, NA10OAR4320137, and NA10OAR4310137 (Global Decadal Hydroclimate Predictability, Variability and Change); and by NSF Grant IIS-1029711. U. Lall was also supported by an IPA from the U.S. Army Corps of Engineers.

REFERENCES

Abukhshim, N. A., P. T. Mativenga, and M. A. Sheikh, 2006: Heat generation and temperature prediction in metal cutting: A review and implications for high speed machining. *Int. J. Mach. Tools Manuf.*, **46**, 782–800, doi:10.1016/j.ijmachtools.2005.07.024.

Badr, H. S., B. F. Zaitchik, and S. D. Guikema, 2014: Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel. *J. Appl. Meteor. Climatol.*, **53**, 614–636, doi:10.1175/JAMC-D-13-0181.1.

Bao, J.-W., S. A. Michelson, P. J. Neiman, F. M. Ralph, and J. M. Wilczak, 2006: Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. *Mon. Wea. Rev.*, **134**, 1063–1080, doi:10.1175/MWR3123.1.

Barnston, A. G., A. Kumar, L. Goddard, and M. P. Hoerling, 2005: Improving seasonal prediction practices through attribution of climate variability. *Bull. Amer. Meteor. Soc.*, **86**, 59–72, doi:10.1175/BAMS-86-1-59.

Berezin, Y., A. Gozolchiani, O. Guez, and S. Havlin, 2012: Stability of climate networks with time. *Sci. Rep.*, **2**, 666, doi:10.1038/srep00666.

Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer-Verlag, 738 pp.

Brunet, G., and Coauthors, 2010: Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bull. Amer. Meteor. Soc.*, **91**, 1397–1406, doi:10.1175/2010BAMS3013.1.

Buizza, R., A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **14**, 168–189, doi:10.1175/1520-0434(1999)014<0168:PPOPUT>2.0.CO;2.

Cañón, J., J. González, and J. Valdés, 2007: Precipitation in the Colorado River Basin and its low frequency associations with PDO and ENSO signals. *J. Hydrol.*, **333**, 252–264, doi:10.1016/j.jhydrol.2006.08.015.

Charles, A., B. Timbal, E. Fernandez, and H. Hendon, 2013: Analog downscaling of seasonal rainfall forecasts in the Murray Darling Basin. *Mon. Wea. Rev.*, **141**, 1099–1117, doi:10.1175/MWR-D-12-00098.1.

Charney, J. G., R. G. Fleagle, V. E. Lally, H. Riehl, and D. Q. Wark, 1966: The feasibility of a global observation and analysis experiment. *Bull. Amer. Meteor. Soc.*, **47**, 200–220.

Chen, S.-M., and J.-R. Hwang, 2000: Temperature prediction using fuzzy time series. *IEEE Trans. Syst. Man Cybern.*, **30B**, 263–275, doi:10.1109/3477.836375.

Dai, Q., M. A. Rico-Ramirez, D. Han, T. Islam, and S. Liguori 2014: Probabilistic radar rainfall nowcasts using empirical and theoretical uncertainty models. *Hydrol. Processes*, **29**, 66–79, doi:10.1002/hyp.10133.

Diallo, I., M. B. Sylla, M. Camara, and A. T. Gaye, 2013: Interannual variability of rainfall over the Sahel based on multiple regional climate models simulations. *Theor. Appl. Climatol.*, **113**, 351–362, doi:10.1007/s00704-012-0791-y.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths, 2009: Complex networks in climate dynamics. *Eur. Phys. J. Spec. Top.*, **174**, 157–179, doi:10.1140/epjst/e2009-01098-2.

Efron, B., 1982: *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 38, SIAM, 92 pp., http://dx.doi.org/10.1137/1.9781611970319.

Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624, doi:10.1175/1520-0434(1994)009<0619:AFSTCV>2.0.CO;2.

Foresti, L., L. Panziera, P. V. Mandapaka, U. Germann, and A. Seed, 2015: Retrieval of analogue radar images for ensemble nowcasting of orographic rainfall. *Meteor. Appl.*, **22**, 141–155, doi:10.1002/met.1416.

Golding, B. W., 1998: Nimrod: A system for generating automated very short range forecasts. *Meteor. Appl.*, **5**, 1–16, doi:10.1017/S1350482798000577.

Hosmer, D. W., Jr., and S. Lemeshow, 2004: *Applied Logistic Regression*. John Wiley & Sons, 392 pp.

Hsieh, W. W., 2001: Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *J. Climate*, **14**, 2528–2539, doi:10.1175/1520-0442(2001)014<2528:NCCAOT>2.0.CO;2.

Huntington, T. G., G. A. Hodgkins, G. D. Keim, and R. W. Dudley, 2004: Changes in the proportion of precipitation occurring as snow in New England (1949–2000). *J. Climate*, **17**, 2626–2636, doi:10.1175/1520-0442(2004)017<2626:CITPOP>2.0.CO;2.

Jackson, D. A., 1993: Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, **74**, 2204–2214, doi:10.2307/1939574.

Jackson, J. E., 2005: *A User's Guide to Principal Components*. John Wiley & Sons, 592 pp.

Jeong, J.-H., B.-M. Kim, C.-H. Ho, and Y.-H. Noh, 2008: Systematic variation in wintertime precipitation in East Asia by MJO-induced extratropical vertical motion. *J. Climate*, **21**, 788–801, doi:10.1175/2007JCLI1801.1.

Jiang, X., S. Yang, Y. Li, A. Kumar, X. Liu, Z. Zuo, and B. Jha, 2013: Seasonal-to-interannual prediction of the Asian summer monsoon in the NCEP Climate Forecast System version 2. *J. Climate*, **26**, 3708–3727, doi:10.1175/JCLI-D-12-00437.1.

Jolliffe, I. T., 2002: *Principal Component Analysis*. 2nd ed. Springer Series in Statistics, Springer-Verlag, 488 pp., doi:10.1007/b98835.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kang, S., J. Hur, and J.-B. Ahn, 2014: Statistical downscaling methods based on APCC multi-model ensemble for seasonal prediction over South Korea. *Int. J. Climatol.*, **34**, 3801–3810, doi:10.1002/joc.3952.

Kawale, J., and Coauthors, 2011: Data guided discovery of dynamic climate dipoles. *Proc. 2011 Conf. on Intelligent Data Understanding (CIDU 2011)*, Mountain View, CA, NASA, 30–44.

——, and Coauthors, 2013: A graph-based approach to find teleconnections in climate data. *Stat. Anal. Data Min.*, **6**, 158–179, doi:10.1002/sam.11181.

Kim, T.-W., J. B. Valdés, B. Nijssen, and D. Roncayolo, 2006: Quantification of linkages between large-scale climatic patterns and precipitation in the Colorado River Basin. *J. Hydrol.*, **321**, 173–186, doi:10.1016/j.jhydrol.2005.07.043.

Kornbrot, D., 2014: Point biserial correlation. *Wiley StatsRef Statistics Reference Online*, doi:10.1002/9781118445112.stat06227.

Landman, W. A., D. DeWitt, D.-E. Lee, A. Beraki, and D. Lötter, 2012: Seasonal rainfall prediction skill over South Africa: One- versus two-tiered forecasting systems. *Wea. Forecasting*, **27**, 489–501, doi:10.1175/WAF-D-11-00078.1.

Lavers, D. A., R. P. Allan, E. F. Wood, G. Villarini, D. J. Brayshaw, and A. J. Wade, 2011: Winter floods in Britain are connected to atmospheric rivers. *Geophys. Res. Lett.*, **38**, L23803, doi:10.1029/2011GL049783.

——, ——, G. Villarini, B. Lloyd-Hughes, D. J. Brayshaw, and A. J. Wade, 2013: Future changes in atmospheric rivers and their implications for winter flooding in Britain. *Environ. Res. Lett.*, **8**, 034010, doi:10.1088/1748-9326/8/3/034010.

Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528, doi:10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2.

Lee, D. Y., J.-B. Ahn, and K. Ashok, 2013: Improvement of multimodel ensemble seasonal prediction skills over East Asian summer monsoon region using a climate filter concept. *J. Appl. Meteor. Climatol.*, **52**, 1127–1138, doi:10.1175/JAMC-D-12-0123.1.

Leung, L. R., and Y. Qian, 2009: Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophys. Res. Lett.*, **36**, L03820, doi:10.1029/2008GL036445.

Liess, S., and Coauthors, 2014: Different modes of variability over the Tasman Sea: Implications for regional climate. *J. Climate*, **27**, 8466–8486, doi:10.1175/JCLI-D-13-00713.1.

Ligda, M. G. H., 1953: The horizontal motion of small precipitation areas as observed by radar. MIT Dept. of Meteorology Tech. Rep. 21, 60 pp.

Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.*, **32**, L14801, doi:10.1029/2005GL023451.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 18 pp.

Lu, M., U. Lall, A. Schwartz, and H. Kwon, 2013: Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in France in 1995. *Water Resour. Res.*, **49**, 6381–6392, doi:10.1002/wrcr.20512.

Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600, doi:10.1175/1520-0450(1987)026<1589:CVISCF>2.0.CO;2.

Mitchell, J. M. J., B. Dzerdzeevskii, H. Flohn, W. L. Hofmeyr, H. H. Lamb, K. N. Rao, and C. C. Wallén, 1966: Climatic change. Rep. for the Commission for Climatology, AGM Tech. Note 79, WMO-195, 79 pp.

Nakamura, J., U. Lall, Y. Kushnir, A. W. Robertson, and R. Seager, 2013: Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. *J. Hydrometeor.*, **14**, 485–504, doi:10.1175/JHM-D-12-059.1.

Nicholson, S. E., 2014: The predictability of rainfall over the Greater Horn of Africa. Part I: Prediction of seasonal rainfall. *J. Hydrometeor.*, **15**, 1011–1027, doi:10.1175/JHM-D-13-062.1.

Olsson, J., L. Simonsson, and M. Ridal 2014: Rainfall nowcasting: Predictability of short-term extremes in Sweden. *Urban Water J.*, **11**, 605–615, doi:10.1080/1573062X.2013.847465.

Preisendorfer, R. W., and C. D. Mobley, 1988: *Principal Component Analysis in Meteorology and Oceanography*. Developments in Atmospheric Sciences, Vol. 17, Elsevier, 444 pp.

Ralph, F. M., and M. D. Dettinger, 2011: Storms, floods, and the science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union*, **92**, 265–266, doi:10.1029/2011EO320001.

——, P. J. Neiman, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, doi:10.1029/2006GL026689.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

——, and ——, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626, doi:10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

Schepen, A., Q. J. Wang, and D. Robertson, 2012: Evidence for using lagged climate indices to forecast Australian seasonal rainfall. *J. Climate*, **25**, 1230–1246, doi:10.1175/JCLI-D-11-00156.1.

Seager, R., and Coauthors, 2007: Model projections of an imminent transition to a more arid climate in southwestern North America. *Science*, **316**, 1181–1184, doi:10.1126/science.1139601.

Silverman, D., and J. A. Dracup, 2000: Artificial neural networks and long-range precipitation prediction in California. *J. Appl. Meteor.*, **39**, 57–66, doi:10.1175/1520-0450(2000)039<0057:ANNALR>2.0.CO;2.

Singh, A., N. Acharya, U. C. Mohanty, and G. Mishra, 2013: Performance of Multi Model Canonical Correlation Analysis (MMCCA) for prediction of Indian summer monsoon rainfall using GCMs output. *C. R. Geosci.*, **345**, 62–72, doi:10.1016/j.crte.2012.11.003.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796–799, doi:10.1126/science.1139540.

Sokol, Z., D. Kitzmiller, P. Pesice, and J. Mejsnar, 2013: Comparison of precipitation nowcasting by extrapolation and statistical-advection methods. *Atmos. Res.*, **123**, 17–30, doi:10.1016/j.atmosres.2012.10.013.

Song, Y., Y. Li, B. Bates, and C. K. Wikle, 2014: A Bayesian hierarchical downscaling model for south-west Western Australia rainfall. *J. Roy. Stat. Soc.*, **63C**, 715–736, doi:10.1111/rssc.12055.

Steinhaeuser, K., N. V. Chawla, and A. R. Ganguly, 2011: Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science. *Stat. Anal. Data Min.*, **4**, 497–511, doi:10.1002/sam.10100.

——, A. R. Ganguly, and N. V. Chawla, 2012: Multivariate and multiscale dependence in the global climate system revealed

through complex networks. *Climate Dyn.*, **39**, 889–895, doi:10.1007/s00382-011-1135-9.

Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118, doi:10.1126/science.1109496.

Trigo, R. M., D. Pozo-Vázquez, T. J. Osborn, Y. Castro-Díez, S. Gámiz-Fortis, and M. J. Esteban-Parra, 2004: North Atlantic oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula. *Int. J. Climatol.*, **24**, 925–944, doi:10.1002/joc.1048.

Tsonis, A. A., and P. J. Roebber 2004: The architecture of the climate network. *Physica A*, **333**, 497–504, doi:10.1016/j.physa.2003.10.045.

——, K. L. Swanson, and P. J. Roebber, 2006: What do networks have to do with climate? *Bull. Amer. Meteor. Soc.*, **87**, 585–595, doi:10.1175/BAMS-87-5-585.

——, G. Wang, K. L. Swanson, F. A. Rodrigues, and L. da Fontura Costa, 2011: Community structure and dynamics in climate networks. *Climate Dyn.*, **37**, 933–940, doi:10.1007/s00382-010-0874-3.

Vitart, F., A. W. Robertson, and D. L. T. Anderson, 2012: Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *WMO Bull.*, **61** (2), 23–28.

Walker, G. T., 1923: Correlation in seasonal variations of weather, VIII: A preliminary study of world weather. *Mem. Indian Meteor. Dept.*, **24**, 75–131.

——, 1924: Correlation in seasonal variations of weather, IX: A further study of world weather. *Mem. Indian Meteor. Dept.*, **24**, 275–332.

——, 1928: World weather III. *Mem. Roy. Meteor. Soc.*, **2**, 97–104.

——, and E. W. Bliss, 1930: World weather IV. *Mem. Roy. Meteor. Soc.*, **3**, 81–95.

——, and ——, 1932: World weather V. *Mem. Roy. Meteor. Soc.*, **4**, 53–84.

——, and ——, 1937: World weather VI. *Mem. Roy. Meteor. Soc.*, **4**, 119–139.

Wang, W., and T. T. Warner, 1988: Use of four-dimensional data assimilation by Newtonian relaxation and latent-heat forcing to improve a mesoscale-model precipitation forecast: A case study. *Mon. Wea. Rev.*, **116**, 2593–2613, doi:10.1175/1520-0493(1988)116<2593:UOFDDA>2.0.CO;2.

Wilson, J. W., N. A. Crook, C. K. Mueller, J. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079–2099, doi:10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2.

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Yamasaki, K., A. Gozolchiani, and S. Havlin, 2008: Climate networks around the globe are significantly affected by El Niño. *Phys. Rev. Lett.*, **100**, 228501, doi:10.1103/PhysRevLett.100.228501.

Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, doi:10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.