EPA Public Access

Author manuscript

J Hydrol (Amst). Author manuscript; available in PMC 2019 September 18.

About author manuscripts

Submit a manuscript

Published in final edited form as:

J Hydrol (Amst). 2017 ; 548: 305–321.

Multi-scale quantitative precipitation forecasting using nonlinear and nonstationary teleconnection signals and artificial neural network models

Ni-Bin Chang^a, Y. Jeffrey Yang^b, Sanaz Imen^a, Lee Mullon^a

^aDepartment of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, USA

^bU.S. EPA, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH, USA

Abstract

Global sea surface temperature (SST) anomalies are observed to have a significant effect on terrestrial precipitation patterns throughout the United States. SST variations have been correlated with terrestrial precipitation via ocean–atmospheric interactions known as climate teleconnections. This study demonstrates how the scale effect could affect the forecasting accuracy with or without the inclusion of those newly discovered unknown teleconnection signals between Adirondack precipitation and SST anomaly in the Atlantic and Pacific oceans. Unique SST regions of both known and unknown telecommunication signals were extracted from the wavelet analysis and used as input variables in an artificial neural network (ANN) forecasting model. Monthly and seasonal scales were considered with respect to a host of long-term (30-year) nonlinear and nonstationary teleconnection signals detected locally at the study site of Adirondack. Similar intra-annual time-lag effects of SST on precipitation variability are salient at both time scales. Sensitivity analysis of four scenarios reveals that more improvements of the forecasting accuracy of the ANN model can be observed by including both known and unknown teleconnection patterns at both time scales, although such improvements are not salient. Research findings also highlight the importance of choosing the forecasting model at the seasonal scale to predict more accurate peak values and global trends of terrestrial precipitation in response to teleconnection signals. The scale shift from monthly to seasonal may improve results by 17% and 17 mm/day in terms of R squared and root of mean square error values, respectively, if both known and unknown SST regions are considered for forecasting.

Keywords

Climate change; Precipitation; Sea surface temperature; Wavelet; Teleconnection patterns; Artificial neural networks; Forecasting

1. Introduction

Hydrologic time series prediction has been a successful modeling tool to address the underlying physical mechanisms by a holistic approach (Coulibaly and Baldwin, 2005). A problem arises in this the type of physical system modeling due to its highly dynamic and

even chaotic nature, exhibiting nonlinearity and non-stationarity. This variability means that the characteristics of the system change over time, possibly due to external forcing and internal feedback mechanisms simultaneously, which are difficult to represent by traditional numerical models with physics-based governing equations. Inputs and outputs of this type of physical system are not proportional and are therefore subject to frequent abrupt change, even in a chaotic mode as opposed to slow and gradual, during which multiple equilibrium states are the norm (Rial et al., 2004). This tendency contrasts with traditional assumptions of stationarity and linearity in many environmental systems analyses.

Climate is generally deemed the best example of nonlinear dynamics driven by interactions of its subsystems with multiple scales (e.g., the ocean, the terrestrial systems, and the atmosphere) (Gonzalez et al., 2013). Hence, the main challenge in studying the climate system is to detect the key linkages and their interactions/feedbacks that modulate the dynamic behaviour of the system (Gonzalez et al., 2013). As an example of the climate systems' functional response, we focused on the precipitation in the northeastern United States, a region with well documented, strong correlations with known climate indices or teleconnection patterns regarded as a set of index regions in the oceans, such as the Pacific North American (PNA), El-Nino Southern Oscillation (ENSO), and the North Atlantic Oscillation (NAO) Patterns. The warm El Nino phase of the ENSO has been linked with cold and stormy winter weather in the northeastern United States, whereas the cold La Nino cycle brings warmer and milder winter weather into the same region (Greene, 2012). The NAO, traditionally measured by sea level pressure differences between the Azores and Iceland, is the dominant pattern of Atlantic climate variability and has been shown to have broad influence over the Atlantic storm track (Marshall et al., 2001). During positive NAO phases, Iceland's low pressure systems and Azores' high pressures lock the jet stream in the northeastern zonal path across North America and cause greater maximum temperatures and fewer extreme cold days in the winter in the northeastern United States. (Wettstein and Mearns, 2002). In addition, during negative NAO phases, above-average snowfall is produced in New England as a result of eastward displacement of the typical East Coast storm track (Bradbury et al., 2003). PNA is also a dominant extratropical teleconnection pattern in the Northern Hemisphere, characterized by above average sea level pressure (SLP) or geospatial heights over the Hawaiian region and mountainous regions of western North America, as well as below average heights south of Alaska and over the southeastern United States. The positive cycle of PNA has been linked to increased precipitation over the northeastern United States by up to 15% over the last century (Beckage et al., 2008, Tang and Beckage, 2010).

Focusing on teleconnection signals generated by these known patterns in climate science, however, limits the predictive potential for hydrometeorological forecasting, mainly depending on those unstable relationships between precipitation variability and some of these known patterns (Gan et al., 2007; Jiang, 2013). The strength of these climate or teleconnection patterns, which are normally composed of some key index regions in oceans, could change in time and space, making their applications unreliable for terrestrial precipitation forecasting. Hence, one of the main efforts of this study was to explore the potential or unknown teleconnection signals affecting terrestrial precipitation over a specific geographic location, signals mainly selected as sectors of sea surface temperature (SST) not

previously reported. In this context, these selected sectors of SST are termed herein as “unknown SST regions” or “unknown index regions”, relative to the teleconnection power tied to these known patterns (such as NAO, ENSO, PNA, and others). These same areas are also termed “known SST regions” in the Pacific and Atlantic oceans and have a significant effect on precipitation variability over the terrestrial study site. Regardless of whether or not they are known or unknown, they are herein called “SST index regions” or simply “SST regions”, all of which collectively contribute to a unique “teleconnection pattern”.

Recently, artificial neural network (ANN) models have been found to be useful predictor variables for hydrologic forecasting. The ENSO pattern, for example, is an important coupled oceanic–atmospheric phenomenon with global significance to the Earth’s climate and strong manifestations in temperature and precipitation worldwide (Solomon et al., 2007). Abbot and Marohasy (2012) found a strong performance enhancement to Australian precipitation forecasting when combining multiple climate-related teleconnection patterns or index regions, including the Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO), and the Nino 3.4 index, and this scientific discovery gave rise to evidence-based support for our study.

ANNs are inherently reliant on the basic time series predictor data of the model. Early ANN-based hydrologic studies reported limited forecasting success, which has been attributed to weaknesses in the time series inputs. To solve this issue, Wang and Ding (2003) proposed pre-treating the time series data by decomposing wavelets into their fundamental signals, creating a new hybrid ANN model called the Wavelet Neural Network (WNN). Wavelet decomposition provides a mathematical process for distilling a signal into multiple levels of details while also extracting local information of the time series (Satyaji Rao and Krishna, 2009), providing a more robust representation of the time series data. WNNs also solved a common problem in ANN forecasting. Multivariate hydrologic time series data are routinely combined from many sources of varying frequencies into a single model, such as evapotranspiration and streamflow data (Ancil and Tape, 2004), a process that can have a conflicting effect on the ANN, causing the model to diverge from the solution. By extracting the underlying frequency behavior of these datasets, the WNN models can more successfully interpret the interrelationships.

Mwale et al. (2004) used a WNN model to forecast seasonal precipitation across a gridded array of rainfall data at a continental scale. Using this a novel approach, Mwale et al. (2004) applied wavelet principal component analysis (WPCA), a technique that analyzes individual scale-averaged wavelet power (SAWP) datasets, to identify homogeneous zones of rainfall variability and to improve precipitation predictability in East Africa. Interestingly, Mwale et al. (2004) did not rely on the known SST regions such as ENSO and NAO in their predictive modeling, and instead included independent teleconnection patterns or index regions in the relevant oceans to obtain predictor data.

Long-term teleconnection analysis is hampered by using linear correlation directly between SST and terrestrial response when evaluating some nonlinear systems (Franzke, 2009) and when depending on some teleconnection signals with low signal-to-noise ratios. This study extends an earlier short-term study (Mullon et al., 2013) by exploring the time scale effect

on the precipitation within Adirondack State Park based on a host of long-term (30-year) teleconnection signals between SST in the North Atlantic and Pacific oceans. This study does not, however, aim to show how much improvement can be gained by including some unknown teleconnection signals not previously considered because most known teleconnection patterns have been identified and proved influential for terrestrial precipitation variability. Instead, the aim of this research was to demonstrate how the scale effect from monthly to seasonal could affect the forecasting accuracy with or without the inclusion of those newly discovered unknown teleconnection signals between Adirondack precipitation and SST anomaly in the Atlantic and Pacific oceans. Based on a wavelet-based ANN forecasting model, teleconnection signals associated with specific geographic locations (e.g., the Adirondack site herein) can be comprehensively compared without relying solely on the known SST regions. The unknown teleconnection patterns were therefore derived independently and based exclusively on the SST and terrestrial precipitation databases from the National Aeronautics and Space Administration (NASA) and National Oceanographic and Atmospheric Administration (NOAA), respectively. Nevertheless, the inputs of local meteorological information associated with mesoscale and synoptic-scale weather phenomena were not included in this forecasting approach because they are associated with a more deterministic weather prediction scheme for a short-term (i.e., 5–7 day) meteorological analysis.

The following science questions are explored herein: (1) Do unknown teleconnection patterns exist between the Adirondack region precipitation and SST in the North Atlantic and Pacific oceans? (2) What is the effect of considering the time lag in temporal aggregation on detecting SST index regions that significantly affect precipitation variability over the Adirondack region? (3) How is precipitation forecasting affected by shifting the time scale from seasonal to monthly? (4) How would the performance of the developed forecasting model be affected if the unknown teleconnection patterns were not included as predictors in the model? We hypothesized that both influential known and unknown teleconnection signals exist for the study region, and including the unknown teleconnection patterns could largely improve the performance of the developed forecasting model. Because terrestrial precipitation variability and SST forcing exhibit significant seasonality, we hypothesized that the seasonal time scale greatly improves the forecasting accuracy. We also hypothesized that a high commonality exists between the SST regions, affecting the seasonal-mean precipitation and precipitation at a monthly scale.

2. Study area

An important aspect of local-scale teleconnection research is to identify a study region with limited signal distortion from anthropogenic activities or natural disasters. Often, anthropogenic activities such as land use and land cover change can mask underlying teleconnection signals, causing undue noise in the time series data. In this study, Adirondack State Park in upstate New York was identified as an ideal candidate site in the northeastern United States because of limited anthropogenic activities, disease impact, or natural disasters within the study time frame, as well as the strong literature evidence of teleconnection power over this area.

Adirondack State Park is a ~22,000 km² (6 million acres) region of public and private lands in northeastern New York, consisting primarily of undeveloped forested lands (Fig. 1). The area was originally dedicated in 1892 to protect the region from mass deforestation common during the 19th century, and much of the park is now afforded constitutional protections that prevent the removal of timber as well as other degrading land management and development practices. Areas in both the North Atlantic and Pacific oceans were selected to examine SST anomalies to terrestrial precipitation over the Adirondack site, along with the locations of the known teleconnection patterns as ENSO, NAO, and PNA associated with the northeastern United States. (Fig. 1).

3. Data and methods

3.1. Data

The precipitation dataset used for this study was obtained from the Climate Prediction Center's Unified Gauge-Based Analysis of Daily Precipitation over CONUS (CPC data). The CPC data have a continuous spatial coverage of 0.25×0.25 degrees at monthly mean temporal resolution from 1948 to the present. The data were derived from >13,000 daily rain gauges reporting across the conterminous United States (NOAA, 2014), and quality control procedures included ground-based radar precipitation for corrections based on the GOES 8 and GOES 10 IR satellite data. Thirty-nine gridded stations fit within the Adirondack State Park boundaries.

SST data were obtained from the Group for High Resolution Sea Surface Temperature (GHRSST) global Level 4 SST analysis, which produces a daily 0.25×0.25 degree grid of SST and uses optimum interpolation (OI) data from the 4 km AVHRR Pathfinder Version 5 time series. The data were bias-adjusted using *in situ* ship and buoy observations (NASA, 2014).

Both precipitation and SST datasets covering the study period from 1983 to 2010 were converted into monthly and seasonal anomaly time series. SST data were converted to a spatial resolution of 5°. A total of 116 and 255 grids cover the North Atlantic and Pacific basins, respectively (Fig. 1). The collected SST and precipitation data were used in the forecasting model after the processing steps described in the following sections.

3.2. Methods

An analytical framework includes a suite of essential steps to identify the teleconnection patterns over both Atlantic and Pacific oceans specific to the study area, as well as the wavelet-based ANN modeling analysis used to calibrate the model and forecast precipitation (Fig. 2). It should be noted that the wavelet-based ANN model in the current study is a hybrid model in which wavelet and principal component analysis are woven together, which is different with the WNN model in the literature. The first step was to obtain the precipitation and SST data for the selected time-scale, including monthly or seasonal in this study and then calculate the standardized anomaly of both precipitation and SST data. This step was followed by a wavelet transformation over the time-frequency domain, in which the original time series values of SST and precipitation were decomposed via various wavelets

into their fundamental signals and used to extract the scaled average wavelet power (SAWP) within selected frequency bands, defined by the associated 2-dimensional time frequency wavelet power spectra contour graph. Principal component analysis was then performed on the precipitation data array to create a single time series, representing the majority of the precipitation variability over the study area (denoted as WPC-SAWP hereafter). The new single precipitation time series was then mapped, using pixel-wise correlation with the SST SAWP data to create the teleconnection pattern maps. Statistically significant SST regions were then extracted from the standardized anomaly SST data and used as inputs into the forecasting model. Historic precipitation values (raw, unfiltered) were also used as target data in the ANN model for forecasting analyses. The forecasting model (i.e., a wavelet-based ANN model) was trained and validated based on the 29-year dataset of raw SST and precipitation from 1983 to 2009. In addition, forecasting skill was assessed based on unseen data from 2010. Final statistical assessment was carried out for the scale-dependent comparisons of both monthly and seasonal effects on forecasting the terrestrial precipitation in the study region.

3.2.1. Wavelet analysis—Wavelet transformation is the process of decomposing time series data into an associated time-frequency domain. A type of spectral analysis, namely wavelet decomposition, allows the determination of dominant localized variations of power (i.e., where the variance of the time series is largest for a given frequency) (Keener et al., 2010). By decomposing a time series into the time-frequency domain, the dominant modes of variability in frequency domain and how those identified modes vary in time can both be determined (Torrence and Compo, 1998). For this study, wavelet analysis was used to isolate dominant modes of North Atlantic and Pacific SST variability and to correlate these modes with the first principal component of precipitation variability over Adirondack State Park, allowing identification of both known and unknown teleconnection patterns. Wavelet analysis in this study was performed using the MATLAB software script developed by Torrence and Compo (1998). A Morlet wavelet function was used to decompose anomalous time series data using the continuous wavelet transform (CWT) method; the Morlet wavelet adopted in this study is defined as a sine wave modulated by a Gaussian curve. For a detailed description of the wavelet methodology, see Torrence and Compo (1998), and for the equations used in this study, see Equations 2–5 in Mullon et al. (2013).

3.2.2. Wavelet analysis for teleconnection signals—First, monthly standardized anomaly data of precipitation were produced by calculating the monthly averages of each month over the specified time range and subtracting that value from each gridded time series value, representing the 30-year time series for each gridded location. Each monthly anomaly time series was then divided by the climatological standard deviation to generate monthly standardized anomaly data. For the seasonal standardized anomaly data of precipitation, the seasonal-mean values of precipitation were calculated by averages of precipitation of March–April–May (MAM) for spring, June–July–August (JJA) for summer, September–October–November (SON) for fall, and December–January–February (DJF) for winter. The standardized anomaly of each seasonal-mean precipitation was then calculated as described for the case of monthly scale; however, the seasonal average and standard deviation were used instead of monthly average and standard deviation. To establish the possible

spatiotemporal associations between SSTs forcing and terrestrial responses in precipitation, lag correlation analysis was performed considering SST time series with different time lags. For instance, to find the lagged correlation between precipitation of MAM, the associated SST time series data were averaged for a 3-month period in sequence starting with MAM (0 lag), FMA (1 month lag), JFM (2 months lag), ..., MAM (12 months lag). Note that the standardized anomaly time series of SST were generated and applied in the same way as its precipitation counterpart in this lag correlation analysis.

To improve understanding of the following wavelet analysis, a typical example of spectral decomposition of a single SST grid using wavelet (Fig. 3, Fig. 4) for both scenarios of monthly and seasonal scale, respectively, can be used to show the monthly and seasonal anomaly time series values and the associated local wavelet power spectra. The dark red areas illustrate regions, in both time and frequency, where high wavelet power exists, and the blue areas illustrate regions of little power. The area within the black bold lines are regions of statistical significance at the 95% confidence level. The inverted arch shape expressed by the pink dash line (Fig. 3, Fig. 4) represents the cone of influence in the wavelet analysis, determined by the length of the time series. For a wavelet analysis to determine the power in each varying period, the non-dimensional wavelet can expand to encompass larger periods and contract to encompass smaller ones flexibly. Thus, larger periods suffer from “edge effects”, where the larger wavelets have a much smaller window over which they can completely propagate through the time series; therefore, zeroes are “padded” to either end for the time series to compensate. For this reason, regions of the graph outside of this pink dash line are not fully reliable because the wavelet is influenced by artificial zero values. Because of edge effects, longer periods have a narrow region of the graph where the data are considered valid.

In this study, the wavelet power of the selected grid of SST was concentrated between the 0.25- and 1-year period bands for monthly scale, which show a dominant intra-annual signal (Fig. 3b). In addition, the wavelet power of the same grid of SST is concentrated between the 2- and 4-year period bands for seasonal scale (Fig. 4b). The global wavelet spectrum summarized on the right-hand-side of the wavelet power sub-diagram (Figs. 3c and 4c) provides an unbiased estimation of the true power spectrum of the time series and therefore provides a robust characterization of the time series variability (Santos et al., 2003). Frequency between the selected bands of the time series was confirmed by an integration of power over time (Figs. 3c and 4c), showing a significant peak above the 95% percent confidence level for the global wavelet spectrum. In both scenarios, it is indicative that several dominant frequencies within the selected period band exists that influence the SST and precipitation variability, especially in the study with the seasonal scale.

The SAWP (Figs. 3d and 4d) is defined as a time series of the average variance in the selected band used to examine modulation of that time series by another. Fig. 3d derives from the average of Fig. 3b over all scales between the 0.25- and 1-year band period in wavelets, and Fig. 4d from the average of Fig. 4b over all scales between the 2- to 4-year band period. To efficiently capture the dominant frequencies, for the monthly scale, SAWP was calculated with the scales set at 0.25- and 1-year for all gridded data. In addition, for the seasonal scale, SAWP was calculated within the inter-annual frequency band (2-4 year)

period for all gridded data. The gridded time series data were therefore converted from the original monthly and seasonal standardized anomaly time series into a new wavelet variability time series between the 0.25- and 1-year period band and the 2- and 4-year period band, respectively. These data are termed as scale-average time series of the precipitation SAWP and SST SAWP in this context to perform pixel-wise correlation analysis and create the subsequent teleconnection maps.

3.2.3. Multivariate principal component wavelet processing—To establish specific oceanic teleconnection regions between the Adirondack precipitation and SST, pixel-wise correlation analysis was employed to identify regional teleconnection “hotspots”. Comparing how each of the 39 individual precipitation grids within Adirondack correlates with each individual SST grid within the oceanic boundaries is impractical; therefore, in this study, principal component analysis was used to consolidate the multivariate precipitation SAWP data into a single time series.

Principal component analysis uses empirical orthogonal functions to transform multivariate data into spatially uncorrelated modes of variability (Venegas et al., 1997). The resulting principal components (PCs) are time series data of the same length as the original and represent the relative variability within the multivariate data. For large amounts of data with limited variability, few principal components are needed to represent the majority of the variability within the original data.

The SAWP precipitation time series data were consolidated into a single PC time series. Calculating PCs on SAWP wavelet time series data has previously been denoted by Mwale et al. (2004) as wavelet principal components (WPCs) and is denoted as such herein. The resulting WPCs are considered frequency-compacted energy variability and indicate the magnitude of events of time series averaged from the individual scales (Mwale et al., 2004). WPCs for the monthly and seasonal precipitation dataset were calculated, and the first WPC (i.e., WPC1) was selected to represent the entire dataset. To determine the spatial variability of WPC1 relative to the entire SAWP precipitation data, spatial correlation was conducted between WPC1 and the SAWP of the individual grids of the study site (Fig. 5, Fig. 6). The WPC1 is highly representative of the underlying dataset, with some individual grids explaining more than 90% of the variability. The overall variability explained across all grids was 67% at the monthly scale. For the seasonal scale, the overall variability explained by WPC1 was 63% in fall, 83% in spring, 50% in summer, and 67% in winter, suggesting that the precipitation within the Adirondack region is relatively homogeneous, and that the WPC1 is a good representation of the entire dataset.

3.2.4. Mapping the teleconnection patterns or index regions—Pixel-wise correlation maps were generated based on the statistical linkages between the WPC1 precipitation and the SAWP SST dataset at various lags from 1 to 12 months. This mapping technique is similar to that in Mullon et al. (2013) but uses the variability trends of SAWP in lieu of simple anomalous data to capture the nonlinearity of potential SST index regions. To detect SST index regions in the Pacific and Atlantic oceans having a significant effect on precipitation variability over the Adirondack site, these oceanic regions with long-term

(more than 3 months) and consistently high correlation were extracted from the correlation analysis.

3.2.5. Forecasting model—As mentioned earlier (Section 3.2.3), precipitation data are homogeneous over the spatial coverage area of 39 grids of the Adirondack site. Hence, the creation of separate precipitation forecast models for each grid was unnecessary in this study, and the average raw precipitation value was calculated across the entire precipitation grids to represent a regional analysis of the selected site. This average historical precipitation time series was thus used as the target in the ANN forecasting model. In addition, the standardized anomaly SST values were extracted over the relevant grids from the oceanic location of each detected SST region and averaged to consolidate into a single time series. The time series of the detected SST regions were used as inputs to the ANN model.

The goal of ANN modeling analysis is to learn the underlying teleconnection patterns and cross-linkage relationships from a subset of input data used to train the model for monthly or seasonal forecasting, depending on the learning datasets. To test the forecasting capabilities or skills of the trained ANN model, an unseen input (input not used in the training step) was used for prediction purposes. In nonlinear systems, it is important to train an ANN model based on a representative set of data to avoid overfitting. In a limited nonlinear time series such as precipitation, low frequency recurrent signals might not be “seen” in the training set if the time series is not long enough. It is therefore important to use the longest training set available without compromising the test set. For this purpose, training and validation was conducted based on the datasets from 1983 to 2009. Note that 60% of the data were used for training and the rest were applied for validation to meet the fundamental philosophy of cross validation. Data from 2010 were then used as inputs into the trained forecasting model as the unseen dataset to evaluate the model’s capability to forecast future precipitation at the designated time scale.

The ANN model, obtained from the MATLAB neural network fitting toolbox, is a feed-forward neural network trained by the Levenberge-Marquardt back propagation algorithm consisting of three layers: input, hidden, and output layer. This model can predict the future value of a time series from the past values of inputs (i.e., precipitation time series). The same model architecture was used to formulate different scenarios in this study. A sensitivity analysis was performed by comparing four extended scenarios to assess the impact of unknown teleconnection patterns and the scale shift from monthly to seasonal on the performance of the developed forecasting models. In this context, two scenarios of four include only the detected known teleconnection patterns for monthly and seasonal scale; the other two scenarios contain both known and unknown teleconnection patterns for monthly and seasonal scale. Note that the neural network capacity test associated with hidden layers was applied to check the potential overfitting of the proposed forecasting models (Narasimha et al., 2008; Sheela and Deepa, 2013; Wang et al., 2014).

To evaluate the performance of the developed forecasting model, R -squared values (i.e., coefficient of determination, R^2) and root mean square error (RMSE) were calculated based on Equations 1a and 1b. The statistical indices are given in the following equations, where y_{ip} is the i th predicted value, y_{io} is the i th observed value, N is the number of sample, y^{-P} is

the mean of the predicted value, $\bar{y}^o$ is the mean of the observed value. Model performance ranks well if RMSE is close to 0 or RSQ (or R^2) is close to 1.

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^p - y_i^o)^2} \quad (1a)$$

- Square of the Pearson product moment correlation coefficient (RSQ)

$$RSQ = \frac{\sum_{i=1}^N (y_i^p - \bar{y}^p)(y_i^o - \bar{y}^o)}{\sqrt{\sum_{i=1}^N (y_i^p - \bar{y}^p)^2} \sqrt{\sum_{i=1}^N (y_i^o - \bar{y}^o)^2}} \quad (1b)$$

- Coefficient of Determination (R^2)

$$R^2 = RSQ^2 \quad (1c)$$

In this test, the training samples were measured against a theoretical lower and upper bound of the capacity, based on the categorical number of inputs, the categorical number of outputs, and the number of neurons required for the ANN model:

$$L \leq P \leq L(1 + N/M) \quad (2)$$

where P is the total number of samples, L is the required number of neurons, N is the categorical number of inputs, and M is the categorical number of outputs in an ANN model.

3.2.6. Statistical assessment of scale effect on forecasting analyses—Scale effect over the temporal domain with spatial implications from one month to a seasonal time horizon could significantly affect the forecasting accuracy. In addition to the R -squared values and RMSE, other statistical descriptors may also be employed to discern the inherent difference of the time scale between the forecasted and observed time series of terrestrial precipitation. They include but are not limited to first-order descriptors, such as mean, variance, skewness, and kurtosis; second-order statistical descriptors, such as entropy, energy, and correlation; and high-order statistical descriptors, such as multi-scale covariance and Fourier analysis. In an attempt to improve understanding of the reliability of those selected forecasting models, this study applied 6 statistical descriptors, including mean (measures the location of a region), variance (measures region roughness), skewness (measures relative intensity to average), kurtosis (measures uniformity), entropy (measures the level of disordered information), and energy (measures how much intensity variation is there in a region). This endeavor leads to the integrated assessment of precipitation histograms of three time-series datasets including the ground observations, monthly

predictions, and seasonal predictions from 1983 to 2009 to determine how reliable the model training and validation would be without considering the R -squared values and RMSE.

On a comparative basis, statistical distance is defined herein as the numerical difference between the predicted descriptor values associated with either monthly or seasonal time scales and the corresponding values associated with ground observations. We hypothesized that the statistical distance between seasonal forecasts and ground observations would be smaller than that between monthly forecasts and ground observations based on first- and second-order statistical descriptors in this study. The definition of these four first-order statistics (mean, variance, skewness, kurtosis) can be found in statistical textbooks, and the definition of the two second-order statistical descriptors (entropy and energy) are expressed by Equation 3 below. Details of how to compute and compare these statistical descriptors from spatial and temporal variations can be found in the literature (Aggarwal and Agrawal, 2012).

- Energy (E)

$$E = -k \sum_i P_i^2 \quad (3a)$$

- Entropy (S)

$$S = -k \sum_i P_i \log P_i \quad (3b)$$

where P_i is the probability that the value appears during the system's fluctuations; the quantity k is a physical constant known as Boltzmann's constant, which, like entropy, has units of heat capacity; and the logarithm is dimensionless. In our study, k can be neglected on a comparative basis when dealing with statistical distance.

4. Results and discussion

4.1. Scale effect on the detection of SST regions

Each climate variable (i.e., SST) has temporal and spatial dimensions. The SST data were organized to be $t \times y$, including “ t ” SST regions representing known and unknown teleconnection patterns at given spatial locations and their minor time units of interest in this study (i.e., month or season), and “ y ” year representing the major time unit (i.e., year). To detect key climate drivers, represented here as SST regions, the lagged pixel-wise correlation was calculated between the SAWP of SSTs and the WPC-SAWP of precipitation for the selected time scale (i.e., either monthly or seasonal). Note that selection of the SAWP data for this study prevents us from using a wide range of unknown frequencies embedded in the raw dataset of SST and precipitation time series (Mwale et al., 2004). With this setting, the produced teleconnection pattern map for the monthly time scale can provide a concise view of the climate drivers globally applicable for the study site (Fig. 7). In the Atlantic basin, the

strongest correlation zones are aligned with the known NAO regions typically associated with high pressure systems centered near the Azores and lower pressure systems northward, a phenomenon caused by the westerly atmospheric jet stream and the oceanic Gulf Stream. A signature of the NAO is an atmospheric pressure difference that causes a layered variability affect, where high and low pressures exist along a fine gradient from north to south. The strong negative correlation from 45° to 60°N and a positive correlation from 15° to 45°N are consistent with this NAO pattern. Another known teleconnection pattern that seems to be dominant in the Atlantic basin is the Tropical North Atlantic (TNA) pattern.

In the Pacific region, the strongest correlation occurs south of Alaska as well as along the eastern and western equatorial region, areas associated with the PNA and ENSO regions, respectively. Other known teleconnection patterns, including the Nino 1 + 2 and the Western Pacific (WP) Pattern also appear. The strong adherence of Adirondack precipitation to the known teleconnection patterns confirms the historical evidence of the known teleconnections with the northeastern United States, discussed in the introduction. In addition to the known teleconnection patterns, high correlation regions also exist that are not part of known teleconnection areas. A strong negative correlation off the western coast of Baja California, as well as regions in the western North Atlantic, are present and are generally consistent with teleconnection patterns observed by Mullon et al. (2013) for northeastern teleconnections. This finding emphasizes the importance of considering the unknown SST regions as well as the known SST regions.

The question remains of how the temporal aggregation with varying time scales can affect the detection of SST regions. The detected SST regions for the seasonal time scale (Fig. 8, Fig. 9) when compared to the results for the monthly time scale (Fig. 7) show the commonality of detected SST regions at different time scales. In addition, the detection of SST regions at the seasonal scale highlights the existence of two important unknown SST regions on the western coast of Baja California and regions in the western North Atlantic.

NAO and PNA affect the precipitation variability of Adirondack site in all seasons (Fig. 8, Fig. 9). In addition, except fall and spring, the Western Pacific pattern shows a significant impact on precipitation. Nino 3 and the unknown SST region at the western North Atlantic show a significant effect on precipitation variability within the study area in all seasons except summer and fall, respectively. The other detected unknown SST region impacts precipitation in summer and fall. Nino 4 only affects precipitation in two seasons, winter and spring. Hence, detecting SST regions at seasonal scale helps assess the relative contributions of the detected SST regions on precipitation variability in Adirondack site.

4.2. Scale effect on Time-lag detection of SST regions

In addition to detecting the significance of those index regions, it is important to determine their corresponding time lags. The correlation magnitude shows the intensity of SST forcing effect on terrestrial precipitation, whereas the corresponding time lags represent the possible time period for the precipitation to respond to SST forcing (Wu et al., 2008). Hence, prominent time lags corresponding to the selected SST regions were determined for both monthly and seasonal scales (Table 1, Table 2). At the monthly scale, except for SST region 2, most of the dominant lags for all SST regions are limited to an intra-annual time domain.

Likewise, at the seasonal scale, all of the dominant lags for all SST regions remain limited to an intra-annual time domain. This finding implies that the short-term memory effect is limited to a few months, whereas the long-term memory effect may be up to the time scales to be selected. The longer the time scale, such as seasonal scale, the longer the memory effect, such as 2–4 years. Overall, the corresponding time lag of the same SST region may vary when shifting the time scale from monthly to seasonal scale (Table 1, Table 2).

4.3. Comparative performance of the ANN models for forecasting analyses

For each selected time scale, a forecasting model was developed to fulfill the settings of the proposed scenarios. Each input dataset was fed into the ANN model to train networks for the time period of 1983–2009. The average historic precipitation across all 39 grids was also used as the target of these models to predict. For the proposed forecasting model at the monthly scale, the number of samples was 352, the categorical number of inputs was 11, and the categorical number of outputs was 1; therefore, number of required neurons in the only hidden layer was equal to 40 at most, satisfying the criterion of Equation 2 for both scenarios. In addition, the proposed forecasting model for the seasonal scale had 28 samples, 7–11 categorical number of inputs based on the selected season, and 1 categorical output. Therefore, 10 neurons at most were required in the only one hidden layer based on the same reasoning.

After the proposed ANN model was tested fully in the training and validation stage, it was applied to forecast precipitation in the year 2010. The ANN output was compared with observed values of precipitation for 1983–2009 by including both known and unknown SST regions for both time scales (Fig. 10). Considering the two time series confirms that both forecasted time series with different scales are neither nonlinear nor stationary time series because neither has a constant mean, a constant variance, or an autocovariance that does not depend on time (i.e., they are not second-order stationary or stationary order 2). Because the modeling analysis randomly divided inputs into two datasets for training and testing (validation), the exact outcome of training and validation, respectively, could not be shown. To gain a qualitative perspective (Fig. 10), the proposed forecasting model was able to estimate the precipitation trend for both time scales; however, the precipitation forecast skill in detecting the peak values of precipitation was significantly improved by shifting the time scale from monthly to seasonal.

Finally, the forecasting ability of the well-trained and well-tested ANN model was assessed by predicting precipitation based on unseen values of SST for 2010. The performance of the model in predicting precipitation using unseen data was also improved at the seasonal scale (Fig. 11). At the monthly scale, however, the forecasting model tended to overestimate the values of precipitation, except during spring.

To answer the last science question, how the performance of the developed forecasting model will suffer from excluding the unknown teleconnection patterns, the performance of the ANN model was assessed for four extended scenarios (Table 3). For the monthly scale, the inclusion of those unknown SST regions did improve the R -squared value and decrease the RMSE value between the observed and ANN output by 2% and 0.05 mm/day, respectively. For the seasonal scale, 4% and 0.02 mm/day improved results were observed

for R -squared and RMSE, respectively. The scale shift from monthly to seasonal may improve results by 17% and 17 mm/day in terms of R -squared and RMSE, respectively, if both known and unknown SST regions are considered for forecasting.

To assess the performance of the proposed ANN model in determining the spatiotemporal variability of precipitation within the Adirondack region, the ANN model was applied to each of the individual precipitation gridded time series, thus forecasting 39 separate precipitation values. The spatial correlation (R -squared) between the ANN outputs and observed values during 1983–2009 and the forecast error (mm/day) contours in 2010 were shown for monthly (Fig. 12) and for seasonal (Fig. 13, Fig. 14) time scales.

Strong correlation trends were observed throughout most of the Adirondack region (Fig. 12), with correlation values varying from 0.41 to 0.76. The forecast error image (Fig. 12, right), however, shows that some high correlation areas significantly underestimated the precipitation, suggesting that some high correlation regions are somewhat misleading and not producing as a forecast as strong as predicted. The same results were obtained at the seasonal scale (Fig. 13, Fig. 14). Results of the ANN model showed higher correlation with observed values when compared to the monthly time scale. In addition, winter had the highest forecast error relative to other seasons, emphasizing the importance of considering other predictor variables, such as temperature, to improve the forecast skill. The areas with higher average precipitation errors could be due to the inherent uncertainties in the datasets, the local precipitation variability over the grids at the Adirondack site, and the topographical conditions, such as the mountainous northeast, which may produce anomalous precipitation that complicates model forecasting. Note that the relatively small size of the Adirondack (~6 million acres) relative to broad teleconnection regions of the ocean and the relatively high spatial resolution of the precipitation data likely limits the teleconnection influence and thus the ANN model's predictive capacity. However, Adirondack State Park was selected for this study because of the relatively low anthropogenic disturbances within the study time frame.

As previously mentioned (Section of 3.2.6), a statistical assessment on the well-trained models with two different scales can be formalized by using 6 descriptors to explore the global changes of predictive capacity. A histogram analyses of terrestrial precipitation over three bases including predicted values, observed values, and ground truth observations (Fig. 15) provides a firm basis for a holistic statistical assessment with respect to the selected 6 descriptors. The frequency distribution is visually similar (Fig. 15(a) and (c)), implying that precipitation prediction based on the seasonal scale can capture the global pattern compared to that based on a monthly scale. To reinforce the comparison, the summary of monthly time scales, seasonal time scales, and ground observations (Table 4) confirms the hypothesis that the distance between seasonal predictions and ground observations were smaller than those between monthly predictions and ground observations based on both first-order and second-order statistical descriptors, except for variance. Note that statistical distance in this context is defined as the difference of descriptor values between the ground observations at a specific time scale (Fig. 10). These statistical descriptors provide a holistic viewpoint of prediction capacity (Table 4), in addition to the two R -squared and RMSE statistical indicators (Table 3).

5. Conclusions

With the aid of wavelet theory, objectively identified nonlinear and nonstationary teleconnection signals associated with Adirondack State Park in the northeastern United States were used to forecast precipitation considering two different time scales (i.e., monthly and seasonal). Wavelet transformation with spectral decomposition reveals that the dominant wavelet power period for both precipitation and SST domains falls into the seasonal to annual frequency band for the monthly scale and 2–4 year band for the seasonal scale. Using the dominant wavelet energy band, the SAWP was calculated to extract the wavelet variability within the selected frequency band. Using principal component analysis, the precipitation SAWP dataset was consolidated into its first principal component (WPC of precipitation). Teleconnection mapping was performed using lagged pixel-wise correlation between the WPC-SAWP precipitation and the SAWP SST domain to complete this comparative study.

The identified teleconnection patterns in the 12-month lag map are consistent with the teleconnections defined in the literature known to influence the northeastern United States as well as potential unknown teleconnections associated with the North Atlantic and Pacific oceans. Results also confirmed the common SST regions for both seasonal and monthly scales. In addition, the seasonal scale helps determine the contribution of more identified SST regions. For instance, results showed the importance of NAO and PNA and their impact on precipitation variability on Adirondack site in all seasons. Interestingly, the same unknown SST regions were detected at both time scales. Both time scales show the intra-annual time-lag effect of SST on precipitation variability in this study region. The corresponding time lag of the same SST region varies by shifting the time scale from monthly to seasonal, however, and may improve results by 17% and 17 mm/day in R -squared values and RMSE, respectively, if both known and unknown SST regions are considered for forecasting. This result can also be evidenced by the much smaller statistical distance between the predicted time series for the seasonal scale and the ground truth observations of terrestrial precipitation at the study site.

The averages of each detected SST region were used as inputs into the ANN model. Forecasting analysis revealed that the forecasting model performed better at the seasonal scale, confirmed by assessing the performance in predicting precipitation using the unseen data from 2010. In addition, comparing the results of the forecasting model for the proposed scenarios showed the minor effects of considering unknown teleconnection patterns as well as the known teleconnection patterns on the performance of forecasting models.

Overall, the methodology proposed herein used a stochastic technique with inductive reasoning for long-range forecasting of precipitation based on historic precipitation observations and teleconnection patterns derived from SST regions. We anticipate that a longer time series record as well as the incorporation of other predictor variables, such as temperatures, SLPs, mesoscale and synoptic scale weather phenomena, and wind directions, will improve the long-range forecasting skills recommended for future research.

Acknowledgement

The authors acknowledge the data and reports cited and used in this analysis. Support from the USEPA National Risk Management Research Laboratory is gratefully acknowledged. The research described herein has been subjected to the Agency's peer and administrative review and has been approved for external publication. Any opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Agency, therefore, no official endorsement should be inferred. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

- Aggarwal N, Agrawal RK, 2012 First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images. *J. Signal Inf. Process* 3, 146–153.
- Abbot J, Marohasy J, 2012 Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Adv. Atmos. Sci* 29 (4), 717–730.
- Ancil F, Tape DG, 2004 An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. *J. Environ. Eng. Sci* 3, S121–S128.
- Beckage B, Osborne B, Gavin DG, Pucko C, Siccama T, Perkins T, 2008 A rapid upward shift of a forest ecotone during 40 years of warming in the Green Mountains of Vermont. *Proc. Natl. Acad. Sci. U. S. A* 105, 4197–4202. 10.1073/pnas.0708921105. [PubMed: 18334647]
- Bradbury JA, Keim BD, Wake CP, 2003 The influence of regional storm tracking and teleconnections on winter precipitation in the northeastern United States. *Ann. Assoc. Am. Geogr* 93, 544–556.
- Coulibaly P, Baldwin CK, 2005 Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *J. Hydrol* 307, 164–174.
- Franzke C, 2009 Multi-Scale analysis of teleconnection indices: climate noise and nonlinear trend analysis. *Nonlinear Process. Geophys* 16, 65–76.
- Gan TY, Gobena A, Wang Q, 2007 Precipitation of western Canada –wavelet, scaling and multifractal analysis and teleconnection to large scale climate anomalies. *J. Geophys. Res. Atm.* 112 10.1029/2006JD007157.
- Gonzalez DL, Pendse SV, Padmanabhan K, Angus MP, Tetteh IK, Srinivas S, Villanes A, Semazzi F, Kumar V, Samatova NF, 2013 Coupled heterogeneous association rule mining (CHARM): Application toward inference of modulatory climate relationships. In: *IEEE 13th International Conference on Data Mining (ICDM)*, 1055–1060, Dallas, TX 10.1109/ICDM.2013.142.
- Greene CH, 2012 The winters of our discontent. *Sci. Am* 307, 50–55. 10.1038/scientificamerican1212-50.
- Jiang RG, Gan TY, Xie J, Ni W, 2013 Spatiotemporal variability of Alberta's seasonal precipitation, their teleconnection with large-scale climate anomalies and sea surface temperature. *Int. J. Climatol* 34, 2899–2917. 10.1002/joc.3883.RMS.
- Keener VW, Feyersisen GW, Lall U, Jones JW, Bosch DD, Lowrance R, 2010 El-Nino/Southern Oscillation (ENSO) influences on monthly NO₃ load and concentration, stream flow and precipitation in the Little River Watershed. *J. Hydrol* 381, 352–363.
- Marshall J, Kushnir Y, Battisti D, Chang P, Czaja A, Dickson R, Hurrell J, McCartney M, Saravanan R, Visbeck M, 2001 North Atlantic climate variability: phenomena, impacts and mechanisms. *Int. J. Climatol* 21, 1863–1898.
- Mullon L, Chang NB, Yang YJ, Weiss J, 2013 Integrated remote sensing and wavelet analyses for screening short-term teleconnection patterns in northeast America. *J. Hydrol* 499, 247–264.
- Mwale D, Gan TY, Shen SSP, 2004 A new analysis of variability and predictability of seasonal rainfall of central southern Africa for 1950–94. *Int. J. Climatol* 24, 1509–1530.
- National Aeronautics and Space Administration (NASA), 2014 http://podaac.jpl.nasa.gov/dataset/NCDC-L4LRblend-GLOB-AVHRR_OI (Accessed by Jan. 2014).
- National Oceanic and Atmospheric Administration (NOAA), 2014 <http://www.esrl.noaa.gov/psd/data/gridded/data.unified.html> (Accessed by Jan. 2014).
- Narasimha P, Manry M, Maldonado F, 2008 Upper bound on pattern storage in feedforward networks. *Neurocomputing* 71 (16–18), 3612–3616.

- Rial JA, Pielke RA Sr., Beniston M, Claussen M, Canadell J, Cox P, Held H, et al., 2004 Nonlinearities, feedbacks and critical thresholds within the earth's climate system. *J. Clim. Change* 65, 11–38.
- Santos CAG, Galvão CO, Trigo RM, Servat E, 2003 Rainfall data analysis using wavelet transform. *Hydrol. Mediterranean Semiarid Reg* 278, 195–201.
- Satyaji Rao YR, Krishna B, 2009 Modelling hydrological time series data using wavelet neural network analysis, *New Approaches to Hydrological Prediction in Data-sparse Regions* In: Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention. IAHS Publications, Hyderabad, India, pp. 101–110.
- Sheela K, Deepa S, 2013 Review on methods to fix number of hidden neurons in neural networks. *Math. Probl. Eng* 2013, 1–11.
- Solomon S, Qun D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, 2007 Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Tang G, Beckage B, 2010 Projecting the distribution of forests in New England in response to climate change. *Divers. Distrib* 16, 144–158. 10.1111/j.1472-4642.2009.00628.x.
- Torrence C, Compo GP, 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc* 79 (1), 61–78.
- Venegas SA, Mysak LA, Straub DN, 1997 Atmosphere-ocean coupled variability in the South Atlantic. *J. Clim* 10, 2904–2920.
- Wang W, Ding J, 2003 Wavelet network model and its application to the prediction of hydrology. *Nat. Sci* 1, 67–71.
- Wang J, Wang C, Chen LP, 2014 The bounded capacity of fuzzy neural networks (FNNs) via a new fully connected neural fuzzy inference system (F-CONFIS) with Its applications. *IEEE Trans. Fuzzy Syst* 22 (6), 1373–1386.
- Wettstein JJ, Mearns LO, 2002 The influence of the North Atlantic-Arctic Oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada. *J. Clim* 15, 3586–3600.
- Wu R, Kirtman BP, Pegion K, 2008 Local rainfall-SST relationship on sub- seasonal time scales in satellite observations and CFS COLA Technical Report 266, 17.

Fig. 1. Gridded boundary regions for Adirondack State Park and the North Atlantic and Pacific basins selected to examine SST anomalies to terrestrial precipitation over the Adirondack site, shown in yellow and gray, respectively. The locations of the known teleconnection patterns associated with the northeastern United States are outlined in blue (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Fig. 2.
The analytical framework of this study.

Fig. 3.

Wavelet decomposition example of a single Precipitation grid for monthly time scale. (a) The top image shows the monthly anomaly data with the X and Y axes representing time and SST in degrees, respectively. (b) The middle image shows the expanded two-dimensional time frequency wavelet power spectra contour graph, with the X and Y axes representing time and period in years, respectively. (c) The right graph shows the global wavelet power, with X and Y axes representing wavelet power in degrees C and period in years, respectively. (d) The bottom image represents the SAWP time series, with X and Y axes indicating average variance over years.

Fig. 4.

Wavelet decomposition example of a single Precipitation grid for seasonal time scale. (a) The top image shows the seasonal anomaly data with the X and Y axes representing time and SST in degrees, respectively. (b) The middle image shows the expanded two-dimensional time frequency wavelet power spectra contour graph, with the X and Y axes representing time and period in years, respectively. (c) The right graph shows the global wavelet power, with X and Y axes representing wavelet power in degrees C and period in years, respectively. (d) The bottom image represents the SAWP time series, with X and Y axes indicating average variance over years.

Fig. 5. Spatial correlation between WPC1 of monthly precipitation and SAWP of individual precipitation grids within the study site.

Fig. 6.
Spatial correlation between WPC1 of seasonal precipitation and SAWP of individual precipitation grids within the study site (a) DJF, (b) MAM, (c) JJA, (d) SON.

Fig. 7.

Results of the 12-month lagged pixel-wise correlations between Adirondack precipitation (WPC1) SAWP and SST SAWP time series at monthly scale. The numbers marked in this figure stand for teleconnection patterns in Table 1. The colored contours represent correlation values with yellow and red indicating statistically significant positive correlation at the 95% and 99% confidence interval, respectively, as well as light blue and dark blue indicating negative correlation at the 95% and 99% confidence intervals, respectively.

Fig. 8. Results of the 12-month lagged pixel-wise correlations between Adirondack precipitation (WPC1) SAWP and SST SAWP time series at seasonal scale: (a) DJF, (b) MAM. The numbers marked in this figure stand for teleconnection patterns in Table 2.

Fig. 9. Results of the 12-month lagged pixel-wise correlations between Adirondack precipitation (WPC1) SAWP and SST SAWP time series at seasonal scale: (a) JJA, (b) SON. The numbers marked in this figure stand for teleconnection patterns in Table 2.

Fig. 10.

Comparison between the ANN output with observed values of precipitation for the time period of 1983–2009 by including both known and unknown SST regions: (a) monthly time scale, (b) seasonal time scale.

Fig. 11.
Comparison between the ANN output with observed values of precipitation for the year 2010 by including both known and unknown SST regions.

Fig. 12.

The image on the left displays the spatial correlation (R -squared) between the ANN output and observed values during 1983–2009. The image on the right displays the forecast error (mm/day) contours in the year 2010.

Fig. 13. Spatial correlation (R -squared) between the ANN output and observed values during 1983–2009 in: (a) DJF, (b) MAM, (c) JJA, (d) SON.

Fig. 14.
Forecast error (mm/day) for the year 2010 in (a) DJF, (b) MAM, (c) JJA, (d) SON.

Fig. 15. Histogram analyses of terrestrial precipitation over varying cases (unit of precipitation interval: mm/day).

Table 1.

The list of selected SST regions and their corresponding lags- monthly scale.

SST Region#	Teleconnection Pattern	Abbrev.	Selected Lag (month)				
			1	3	6	12	24
1	Pacific North American Pattern	PNA	+	⊕	+	+	
2	Pacific North American Pattern	PNA				+	⊕
3	unknown	–	+	+	⊕		
4	Pacific North American Pattern	PNA	+	⊕	+	+	
5	Nino 3.4	Nino 3.4			+	⊕	
6	North Atlantic Oscillation	NAO	+	+	+	⊕	
7	North Atlantic Oscillation	NAO		+	+	⊕	
8	Unknown	–		+	+	⊕	
9	Tropical North Atlantic	TNA		+	⊕	+	
10	West Pacific	WP	+	+	⊕	+	
11	Nino 4	Nino 4			+	⊕	

The influential lag time for each SST region is shown with +.

The topmost influential lag for each SST region is marked with ⊕.

Table 2.

The list of selected SST regions and their corresponding lags- seasonal scale.

Season	SST Region	Teleconnection Pattern	Abbrev.	Selected Lag (month)												
				0	1	2	3	4	5	6	7	8	9	10	11	12
Winter	1	Pacific North American Pattern	PNA	+	+	⊕	+	+	+	+	+	+	+	+	+	+
	2	Pacific North American Pattern	PNA	⊕	+	+	+	+	+	+	+		+	+	+	+
	3	Nino 3	Nino 3	⊕	+	+	+	+	+			+	+	+	+	+
	4	Pacific North American Pattern	PNA	+	+	+	+	+	+	+	⊕	+	+	+	+	+
	5	North Atlantic Oscillation	NAO	+		+	+	+	+	+	⊕		+	+	+	+
	6	North Atlantic Oscillation	NAO	+	+	+	+	+	+	+	+	⊕	+	+	+	+
	7	Tropical North Atlantic	TNA	+	+	+	⊕	+	+		+	+	+	+	+	+
	8	Unknown	—	+	+	+			+	+	+	⊕	+	+	+	+
	9	Nino 4	Nino 4	+	+	+	⊕	+	+	+	+	+	+	+	+	+
	10	West Pacific	WP	+	+	+	+	⊕	+	+	+	+	+	+	+	+
	11	West Pacific	WP	+	+		+	+	+	⊕	+	+	+	+	+	+
Spring	1	Pacific North American Pattern	PNA	+	⊕	+	+	+	+	+	+	+	+	+	+	+
	2	Nino 3	Nino 3	+	+	+	+	⊕	+	+	+	+			+	+
	3	Pacific North American Pattern	PNA	+	+	+	+	+	⊕	+	+		+	+	+	+
	4	Pacific North American Pattern	PNA	+		+	+	⊕	+	+	+	+	+	+	+	+
	5	North Atlantic Oscillation	NAO	+	+	+	+	+	+	+	+	+	+	⊕	+	+
	6	North Atlantic Oscillation	NAO	⊕	+	+	+	+	+	+	+	+	+	+	+	+
	7	Nino 4	Nino 4	+	+	+	+	+	+	⊕	+	+	+	+	+	+
Summer	1	Pacific North American Pattern	PNA	+	+	⊕	+	+	+	+	+	+	+	+	+	+
	2	Pacific North American Pattern	PNA	+	+	+	+	⊕	+	+	+	+	+	+	+	+
	3	Unknown	—	+	+	+	+	+	+	+	+	+	⊕	+	+	+
	4	Pacific North American Pattern	PNA	+	+	⊕	+	+	+	+	+	+	+	+		+
	5	Unknown	—	+	+	+	+	+	+	+	+	⊕	+	+	+	+
	6	North Atlantic Oscillation	NAO	+	+	+	⊕	+	+	+	+	+	+	+	+	+
	7	Tropical North Atlantic	TNA	+	+	⊕	+	+			+	+	+	+	+	+
	8	West Pacific	WP	+	+	+	+	+	+	+	+	⊕	+	+	+	+
Fall	1	Nino 3 & Nino 3.4	Nino 3 Nino3.4	+	+		+	+	+	⊕	+	+	+	+	+	+
	2	Unknown	—	+	+	+	+	+	+	+	+	+	⊕	+	+	+
	3	Pacific North American Pattern	PNA	+	+	+	+	+	⊕	+	+	+	+	+	+	+
	4	North Atlantic Oscillation	NAO	+		+	+	+	⊕		+	+	+	+	+	
	5	Unknown	—	+	+	⊕	+	+	+	+	+	+	+	+	+	+

Season	SST Region	Teleconnection Pattern	Abbrev.	Selected Lag (month)												
				0	1	2	3	4	5	6	7	8	9	10	11	12
	6	Tropical North Atlantic	TNA	+	+		+	+		+	+	⊕	+	+	+	+
	7	Tropical North Atlantic	TNA	⊕	+								+	+	+	+

The influential lag time for each SST region is shown with +.

The topmost influential lag for each SST region is marked with ⊕.

Table 3.

R-squared and root-of-mean square error (RSME) between the observed value and forecasted ANN output for the proposed scenarios.

Scenario	R-squared	RMSE (mm/day)
Monthly-Known and Unknown SST Regions	0.48	0.63
Monthly – Known SST Regions	0.46	0.68
Seasonal- Known and Unknown SST Regions	0.65	0.46
Seasonal-Known SST Regions	0.61	0.48

Table 4.

Statistical descriptor assessment of forecasting analyses with varying time scales based on the selected first order and second order statistical descriptors.

	Monthly Time Scale		Seasonal Time Scale		Ground Observations
	Value	Statistical Distance *	Value	Statistical Distance *	Value
Data range (mm/day)	3.26	0.41	3.38	0.29	3.67
Mean (mm/day)	3.39	-0.33	3.02	0.04	3.06
Variance ([mm/day] ²)	0.54	0.04	0.52	0.06	0.58
Skewness ([mm/day] ³)	0.46	0.05	0.46	0.05	0.51
Kurtosis ([mm/day] ⁴)	0.55	-0.37	0.19	-0.01	0.18
Entropy (% ²)	1.76	-0.15	1.6	0.01	1.61
Energy (% ²)	0.11	-0.02	0.09	0	0.09

* Statistical distance is defined as the difference between the predicted descriptor values associated with either monthly or seasonal time scale and the corresponding ones associated with ground observations.