# A Decision-tree Approach to Seasonal Prediction of Extreme Precipitation in Eastern China

Wenguang Wei, Zhongwei Yan*

*Key Laboratory of Regional Climate–Environment for Temperate East Asia, Institute of Atmospheric Physics; University of Chinese Academy of Sciences; Beijing 100029 China*

Phil D. Jones

*Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK*

Short Title: Decision-tree Approach to Seasonal Prediction of Extreme Precipitation

*Corresponding author: Zhongwei Yan

E–mail: yzw@tea.ac.cn

Postal Address: P. O. Box 9804, RCE-TEA, Institute of Atmospheric Physics, Building 40, HuaYanLi, ChaoYang District, Beijing 100029 China

# Abstract

Seasonal prediction of extreme precipitation has long been a challenge especially for the East Asian Summer Monsoon region, where extreme rains are often disastrous for the human society and economy. This paper introduces a decision-tree (DT) method for predicting extreme precipitation in the rainy season over South China in April-June (SC-AMJ) and the North China Plain in July-August (NCP-JA). A number of preceding climate indices are adopted as predictors. In both cases, the DT models involving ENSO and NAO indices exhibit the best performance with significant skills among those with other combinations of predictors and are superior to their linear counterpart, the binary logistic regression model. The physical mechanisms for the DT results are demonstrated by composite analyses of the same DT path samples. For SC-AMJ, an extreme season can be determined mainly via two paths: the first follows a persistent negative NAO phase in February-March; the second goes with decaying El Niño. For NCP-JA, an extreme season can also be traced via two paths: the first is featured by 'non El Niño' and an extremely negative NAO phase in the preceding winter; the second follows a shift from El Niño in the preceding winter to La Niña in the early summer. Most of the mechanisms underlying the decision rules have been documented in previous studies, while some need further studies. The present results suggest that the decision-tree approach takes advantage of discovering and incorporating various nonlinear relationships in the climate system, hence is of great potential for improving the prediction of seasonal extreme precipitation for given regions with increasing sample observations.

KEY WORDS decision tree, seasonal prediction, extreme precipitation, eastern China

## 1. Introduction

Seasonal extreme precipitation events have disastrous influences especially in the densely populated East Asian regions during the rainy monsoon season. The disasters related to extreme precipitation (e.g. flooding, urban waterlogging and landslides) happen almost every year. As a prominent example, devastating floods due to excessive extreme rains over the whole season hit most of eastern China in the summer of 1998, causing an economic loss of hundreds of billions of dollars and a death toll of thousands (National Climate Center, 1998). A recent example was in May 2016, when successive extreme rains hit South China leading to waterlogging, landslides, debris flow and other subsequent disasters across the region (Li et al., 2018). Prediction of whether there will be such extreme rainfall events in a specific region in upcoming months or season is undoubtedly helpful for reducing the risk of disastrous extreme events.

However, few operating agencies over the world make seasonal prediction of regional extreme precipitation events. One of the most common targets of the seasonal climate prediction is the seasonal total precipitation (usually in form of the percentage precipitation anomaly for a given region). Clearly, an anomaly of seasonal total precipitation does not necessarily indicate the signal of seasonal extreme precipitation events. A typical case was in 2016 in South China, where the seasonal total precipitation did not show a significant anomaly but severe floods happened due to excessive extreme rains (Wang et al., 2017). It is implied that the physical mechanism for anomalous total precipitation should be different from that for extreme rains. Therefore, it is beneficial to explore the predictability and develop direct predictive methods for the seasonal extreme precipitation events for affected regions.

Previous studies have suggested that the seasonal extreme precipitation accumulation during the rainy season in eastern China should be of considerable potential predictability (Wei et al., 2017). However, the signal at any individual station is weak due to strong local weather noise. Using a summarizing index of extreme precipitation for a reasonably large region and a typical temporal aggregation period is a natural way to enhance the signal linking to large-scale predictors (Li and Wang, 2017).

One of the most common means for seasonal prediction is the use of a coupled general circulation model (CGCM) by operational agencies. However, the seasonal prediction of precipitation over the East Asian Summer Monsoon (EASM) region remains a long-standing challenge for dynamical models. Recent studies showed that the prediction of the seasonal total precipitation by physical models such as CGCMs has remained at a limited level of skill (Wang et al., 2009; Wang et al., 2015), not to mention that of the extreme precipitation.

A number of empirical methods have been proposed to predict seasonal precipitation in the EASM region (Fan et al., 2008; Wu et al., 2009; Yim et al., 2014). Various precursors were discovered and some of the associated physical mechanisms have been well documented. For example, many studies have noted that the decaying phase of El Niño influences the climate of East Asian by inducing a persistent anomalous anticyclone over the western North Pacific (Wang et al., 2000; Wang et al., 2003; Wu et al., 2010). Some studies suggested that the tri-pole pattern of sea surface temperature anomaly (SSTA) associated with a negative phase of the North Atlantic Oscillation (NAO) could persist in different seasons and have impacts on the climate in East Asia by triggering a wave train in the mid-high latitudes (Watanabe, 2004; Wu et al., 2009). Gong and Ho (2003) found that the boreal spring Arctic Oscillation (AO) had a negative correlation with the following summer rainfall in the mid-lower reaches of the Yangtze River; while Nan and Li (2003) showed significant positive correlations between the boreal spring Southern Hemisphere annular mode (SAM) and the following summer rainfall in the same region. The influence from decadal and multi-decadal factors such as Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO) is also reported and documented by a number of studies (Zhu and Yang, 2003; Zhang et al., 2013; Zhu et al., 2016; Si and Ding, 2016; Pei et al., 2017, Yang et al., 2017). Different predictive models were then developed. Most of the published models are linear and for prediction of the seasonal total precipitation. Li and Wang (2017) followed similar procedures to establish multiple linear regression models for prediction of the number of extreme rainy days in regions of China.

Since the climate system is nonlinear, any linear model is an approximation to the underlying physical process and usually only suitable for a limited time period. Outside the given time window, the model's prediction skill decreases rapidly. This is a common problem in the field of statistical climate prediction, which sometimes is attributed to inter-decadal climate shifts or nonstationary relationships between different components of the climate system. Another problem arises from linear models usually requiring linearly 'independent' predictors. This is not easily satisfied since the components in the climate system are often related to each other to varying degrees. Consequently, a linear model can only incorporate very few nearly independent predictors but omits many potentially important factors simply due to their linear correlation with the selected predictor. However, the effect of a predictor cannot be simply represented by another correlated predictor in a nonlinear system.

In this study, we introduce a decision-tree (DT) approach to prediction of the seasonal extreme precipitation events in given regions in China, and compare its performance with that of binary logistic regression model, a class of generalized linear model. DT is a classic data mining method but has not yet been well applied in climate prediction. The method is not constrained by independence between predictors and hence allows the discovery and involvement of all possible relationships between the input factors and the target variable as long as there are sufficient training samples. This is suitable for prediction of a nonlinear system such as the climate.

The data and the target variables of prediction are described in Section 2. The methods are introduced in Section 3. The resultant models and their skills are demonstrated in Section 4, followed by the physical interpretation of the DT models in Section 5. A summary of the study with discussion is in Section 6.

## 2. Data and target variables

### 2.1 Data

Daily precipitation records from 824 stations over China were obtained from the National Meteorological Information Center, China Meteorological Administration. We selected a subset of 675 stations without missing records during the period between 1 January 1960 and 31 December 2013. Eastern China is densely covered by this subset of stations.

The monthly-mean sea level pressure (SLP), 850 hPa horizontal winds, and 500 hPa geopotential heights (GPH), gridded at a horizontal resolution of 2.5°× 2.5°, were taken from the National Center for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis datasets (Kalnay et al., 1996). The monthly-mean sea surface temperature (SST) records from the COBE-SSTs dataset were also used (Ishii et at., 2005).

A number of climate indices were applied as the potential predictors. The Niño-3.4, AO, NAO, AMO and detrended AMO indices are available from the NOAA database. The PDO index series is from Nathan Mantua at UW/JISAO (http://research.jisao.washington.edu/pdo/). An East Asian winter monsoon (EAWM) index is available following Wang and Chen (2014). The southern annular mode index is available following Nan and Li (2003). All the climate indices are monthly, based on which the seasonal mean indices are calculated when necessary.

### 2.2 Definition of an Extreme Precipitation Event

In eastern China, most of precipitation in a year occurs during the EASM season. To focus on this rainy season's extreme precipitation, we adopt an accumulated index similar to that of Li and Wang (2017), i.e., the number of extreme precipitation days

(EPD) during the rainy season for a given region. The procedures to decide whether a wet day is an EPD are as follows: (1) use all the available wet days' rainfall amounts to obtain a cumulative distribution function (cdf) for a station; (2) determine the empirical $90^{th}$ percentile of the cdf as the threshold to identify an EPD for this station. Following these procedures, all EPDs can be identified for each station. Thus, we can obtain the accumulated number of EPDs (AEPD) within a time period (e.g. a month, season and year), for each station. Averaging all stations' AEPDs within a region results in a regional mean AEPDs (MAEPD) for the region.

To distinguish between extreme event and non-extreme event for a region, a threshold (e.g. one standard deviation above the mean value of the MAEPD) was adopted to partition the yearly samples into two categories: one for those above the threshold, representing a "real" extreme event (labeled as "above") and one for those below the threshold (labeled as "below"). Varying the threshold (e.g. from 0.5 to 1.1 standard deviations above the mean of the MAEPD), we can obtain different partition results representing extreme and non-extreme events to the different extreme levels.

### 2.3 Target of Prediction

The climatological distribution of monthly AEPD is given in Figure 1. For eastern China, the seasonal cycle is prominent, with most EPDs in the warm season (from April to August). Few EPDs occur in the winter (from December to February, not shown in the figure). As the summer monsoons advance northward during the warm season, the peak of EPDs demonstrates a propagation from south to north. In April-Jun, there are more EPDs in southern China; in July-August, the center of extreme precipitation shifts to the North China Plain. Two target regions are therefore outlined as (1) Southern China (20°N - 32°N, 110°E - 122°E) for April–June (SC-AMJ, hereafter) and (2) North China Plain (32°N - 42°N, 110°E - 135°E) for July–August (NCP-JA, hereafter).

The time series of the seasonal precipitation indices for the two target regions are shown in Figures 2. The numbers of total precipitation days in both regions exhibits a decreasing trend during the past decades. This was mainly due to decreases of light rains across the country in association with global warming as explored by previous studies (Yan and Yang 2000; Qian et al., 2007). However, MAEPD demonstrates quite stationary interannual variability, which implies that the mechanisms for extreme precipitation days and total precipitation days could be different. In the present predictive modeling analysis, we use the MAEPD partition results in these two regions as the target variables.

# 3. Methods

## 3.1 Decision Tree Model

When the extreme precipitation frequencies are divided into two categories with a given threshold (above or below the threshold), a prediction model for such categorical data is essentially a classifier. Such a classifier holds a set of rules related to the predictors. Suppose $X = [X_1, \ldots, X_p]$ is the predictor vector. Each of its components $X_i, i = 1, \ldots, p$ represents a predictor, either a discrete or continuous variable. A realization of the predictor vector is expressed as $[x_1, \ldots, x_p]$. The response variable or predicted target is denoted as $Y$, whose values are taken as a two-element set, say

$\{1,0\}$. A realization of the response variable is expressed as $y$. The rule in a classifier is a mapping or function $Y = f(X)$. Based on a specification $[X_1, \ldots, X_p]$ of the predictors, the classifier is to determine the response value $y$ of the predictand. Typically, the rule is built by analyzing or learning from a training set of samples. An independent set of samples is needed for validation of the performance of the built model. The generation of the classification rules is critical for building a category-predictive model.

The DT model is one type of classifiers. As indicated in its name, DT has a tree-like structure, where each internal node denotes a test on a predictor, each branch is the outcome of the test, and the leaf node holds a class label (Han et al., 2011). The rule induction of DT is based on the information entropy (IE) proposed in the pioneering work by Claude Shannon in his information theory (Shannon, 1948). Assume the response variable Y has m possible outcomes and each outcome holds a probability of $p_i, i = 1, \ldots, m$ ($m = 2$ in the present case). The Shannon's information entropy, as defined in the formula (1), can serve as an index to measure the impurity of the variable.

$$Info(Y) = -\sum_{i=1}^{m} p_i \, log_2(p_i) \quad (1)$$

A large value of $Info(Y)$ implies a high level of impurity. It is easy to show that more categories in Y or a more even distribution of the categories in Y should result in a larger value of $Info(Y)$, or in other words, a higher level of impurity. This is in accordance with common physical intuition. An alternative index to measure the impurity of a variable is the Gini index, defined as

$$Info(Y) = 1 - \sum_{i=1}^{m} {p_i}^2 \quad (2)$$

It has similar characteristics as Shannon's IE. In this study, we will use both indices to generate the DT for extreme precipitation event prediction and compare their results.

Select a predictor $X_i$. A binary split on $X_i$ partitions the training set $S$ into $S_1$ and $S_2$. Another index is defined to measure the impurity of the variable after the partition:

$$Info_{X_i}(Y) = \frac{|S_1|}{|S|} Info(S_1) + \frac{|S_2|}{|S|} Info(S_2) \quad (3)$$

where $|\cdot|$ denotes the number of sample in a set. This index is the weighted average of IEs for the subsets after the partition. The more impurity, the larger the value of $Info_{X_i}(Y)$, and vice versa. For the seasonal prediction here, we prefer a binary split on $X_i$ generating two branches from a node rather than a multiway split leading to more than two branches. This is partly because multi-splits fragment the data too quickly, leaving insufficient data at the next level down. Besides, multiway splits can be achieved by a series of binary splits (Hastie et al., 2008).

The reduction in impurity that would be incurred by a split on $X_i$ is

$$\Delta Info_{X_i}(Y) = Info(Y) - Info_{X_i}(Y) \quad (4)$$

The predictor that maximizes the reduction in impurity is selected as the splitting predictor. The predictor and either its splitting subset (for a discrete-valued predictor) or split-point (for a continuous-valued predictor) together form the splitting criterion. Iterating the above processes results in a decision tree. Theoretically, the training set can be finally split into a number of pure subsets, the leaf nodes, as long as there are enough predictors. However, it is easy to overfit the data when the sample size of a subset is too small. In this situation, continuing to partition the training data will only result in lengthy but meaningless branches. Thus, we need some criteria to decide when to stop partitioning and let the current set form a leaf node. We adopt a stopping criterion that there must be at least 5 samples in a leaf node, considering the relatively small sample size in the present study.

### 3.2 Binary Logistic Regression Model

For a comparison, the binary logistic regression model is also applied, which is a common method to estimate the probability that one case (e.g. extreme event) is present for a binary predictand, given the values of predictors. In fact, it is a type of generalized linear models and has the following form as

$$\log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \quad (5)$$

where $\pi$ is the probability of one of the two cases, $\mathrm{X} = [X_1, \cdots, X_p]$ is the predictor vector, and $\beta = [\beta_0, \cdots, \beta_p]$ is the regression coefficient vector. Although it is not a strictly linear model, we can still notice that it assumes a linear relationship between the natural logarithm of the odds (log odds) and the predictors, which makes it suffer from similar drawbacks with ordinary linear models.

### 3.3 Methods for Validation

Since the predictand is a binary categorical variable and both models are making probability prediction, the receiver operating characteristic (ROC) curve is an appropriate tool to validate the model and compare between different models. A ROC curve is constructed based on the probability prediction results of testing samples. It reflects the changing relationship between hit rate and false alarm rate when the probability threshold changes between 0 and 1, separating the probability prediction results into positive and negative events. Hit rate is the proportion of correct forecast positive events in all observed positive events, while the false alarm rate is the proportion of false forecast positive events in all negative events. False alarm rates and hit rates are shown on the horizontal and vertical axes, respectively. A perfect model should produce a ROC curve composed of the left and upper boundary lines, while a random model will produce the diagonal line as its ROC curve. A skillful model should produce a ROC curve located in the left-upper corner of the rectangle box. The closer to the left-upper corner the curve, the more skillful the model. Thus, the area under the curve (AUC) is a good measure of the model's skill. Quantitatively, AUC represents the probability for a model to distinguish between two given (positive and negative) samples. For the present study, AUC is applied as a primary index for model validation. The Wilcoxon-Mann-Whitney test (Wilks, 2011) is applied to estimate, in terms of AUC, whether the DT model performs statistically better than a random prediction. Based on AUC, the Brier Skill Score is also calculated for the model, using climatological probabilities as the reference forecasts.

Accuracy (ACC) is another commonly used index to validate a prediction model. ACC is simply defined as the ratio of all correct forecast events to the total number of samples. ACC may fail when applied to unbalanced sample sets, because a bad model may produce a high accuracy by simply predicting the dominant class but omitting the minor class. This is just the case for the extreme precipitation prediction, because the defined extreme rainfall seasons might be rare. Therefore, ACC, hit rate and false alarm rate should be combined to comprehensively determine a model's performance. In practical, an optimal cut-off point can be obtained given the costs under the four types of forecast, namely hit, false alarm, miss and correct rejection (Metz, 1978). Since these costs are usually application-oriented and unknown, we simply adopt the point with the largest ACC in the ROC curve as the optimal point and use the ACC, hit rate and false alarm rate corresponding to this point to evaluate the built model.

# 4. Results

## 4.1 Building the Predictive Models

A series of monthly predictors representing large-scale oceanic and atmospheric conditions between the preceding December and the first month of the target season (April for SC-AMJ and June for NCP-JA) are selected for building the predictive models. Therefore, the models make at least 0-lead predictions. The seasonal predictors, i.e., the 3-month-running averages of the corresponding monthly predictors, are also used. In summary, the climate indices used in the present study as potential predictors include those preceding monthly and seasonal indices of NINO3.4 (NINO34), EAWM, AO, NAO, PDO, AMO, and SAM. As mentioned *supra*, the relevant climate relationships have been well documented between these potential precursors and precipitation in eastern China. However, few studies synthesized their combined effects into a nonlinear predictive model for seasonal extreme precipitation events. The DT method provides a way to cope with this issue. More factors have been considered when building the models, including the regional mean anomalies of SLP, 500 hPa geopotential height, and SST in the regions of significant leading correlation with the

MAEPD time series. The method for selecting these factors (Table S1) can be found in the supplemental material.

To fit a model of true skill, the sample set should be partitioned into two subsets with one for model training and the other for model testing. In this study, we randomly select around 75% samples to train the model and the rest to test the fitted model. Moreover, the binary partition of the sample set should keep the ratio of the "above" class number to the "below" class number identical for the subsets. To evaluate the sample partition uncertainty, we repeat the above with random partitions and model building processes multiple times and use the mean ROC curve of the models to represent the performance of one experiment. Here, an experiment consists of a threshold for defining an extreme event and a combination of predictors. It is found that the mean ROC curve tends to be stable after 12 times of random partition. Thus, we build 12 models for each experiment.

Theoretically, the DT method is able to use the combination of all predictors as input and find the optimal paths to form a tree to classify between "above" and "below" classes for the training samples. However, since the sample size is relatively small for the present study, a simultaneous input of the predictors may result in an overfitted model, which usually performs badly on the test sample set. To avoid this problem, we carry out a series of experiments with all possible combinations of different types of predictors. For example, with p types of predictors, we firstly carry out p experiments, of which each considers only one type of predictor (e.g. NINO34). Then, we have $C_p^2$ experiments by including two types of predictors, $C_p^3$ experiments by including three …until including all types of predictors. This is the method of exhaustion. As mentioned above, an experiment also involves a threshold for defining an extreme event. In this study, we adopt a series of thresholds for each combination of predictors, such as 0, 0.1, 0.2, …,1.5 standard deviations above the mean climatology. Comparing the mean AUCs of the models between different combinations of predictors, the combination with the largest mean AUC value is selected as the best combination and the types of predictors used in this combination are considered to be the most important factors for the prediction target. With this best combination of predictors, a further comparison of the mean AUCs of the models corresponding to different thresholds of extremes leads to the threshold for defining an extreme event that has the best predictability. Finally, the most balanced DT model was chosen from the 12 models corresponding to this best threshold for physical interpretation. The same procedures are applied to build the binary logistic regression models. A flowchart illustrating the whole procedure for SC_AMJ is shown in the supplementary (Figure S1).

### 4.2 Selected Predictors

It is found that, for both regions, the maximum mean AUC values are taken when two types of predictors are used: NINO34 and NAO, no matter building a DT model or binary logistic regression model. Thus, NINO34 and NAO are deemed as two robust factors for the prediction of extreme precipitation event for both cases. For a DT model, the experiments using the Gini index have higher skills than those using Shannon's IE. In the following, therefore, we only show the modeling results based on the Gini index for a DT model.

### 4.3 Best Thresholds to Define Extreme Events

Within the models using the combination of ENSO and NAO as predictors, the predictability of events in different extreme levels is revealed by comparing the performance between models trained by different samples resulting from varying thresholds. Results from the DT models show that, for both regions, the mean AUCs demonstrate a first increasing then decreasing trend, peaking at around one standard deviation above the mean (red lines in Figure 3). Considering the decreasing trend is probably caused by deficiency of "above" samples to train a meaningful model when an extremely large threshold is adopted, we suggest that the reasonably extreme events are better predicted. Similar conclusions can also be made from the results of binary logistic regression model (blue lines in Figure 3). Thus, one standard deviation above the mean is a more robust and appropriate threshold to define an extreme precipitation season, regarding the modeling skill. For the following analysis, we have chosen the 12 models trained from the samples categorized by this threshold.

### 4.4 Comparison between DT models and Binary Logistic Regression Models

For SC-AMJ, the two mean ROC curves are shown in Figure 4a: one from DT model and the other from the binary logistic regression model. The logistic model shows a slightly higher value of AUC. However, its ROC curve shows a slower rising rate than that of DT model when the false alarm rate is low. This means that, to reach the same hit rate, the logistic model will make more false alarms, which will deteriorate its performance. For NCP-JA, the mean AUC of DT model is larger than that of logistic model and the rising rate of the ROC curve of DT model is also quicker than that of logistic model when we keep the false alarm rate at a relatively low level (Figure 4b).Thus, for the two regions, the performance of the DT method is superior to that of the binary logistic regression model. Moreover, the DT model provides by its decision rules a natural and intuitive way to interpret the nonlinear interaction between different predictors to generate an extreme precipitation event. This is different from traditional linear models, which always produce a prediction result based on superposition of the linearly independent predictors. DT is a knowledge-discovery process, automatically producing the nonlinear relationship when the predictive model is built. The discovered relationships in the decision rules of a DT model can be further analyzed to understand the underlying physics.

### 4.5 Balanced Models and Validations

To extract robust decision rules, we compare the decision rules between 12 models. It is found that all models demonstrate similar rules, even though there are minor differences due to the uncertainty from random partitioning between the training and testing sets. Such uncertainty arises from the fact that random partitions may lead to biased formations of the training and testing sets. For example, ideally, there should be nearly equal ratios of samples with different mechanisms in both the training and testing sets, but in practice, with limited samples, a larger ratio of samples with certain mechanisms may fall into the training set, compared to the testing set. In the ideal case, all mechanisms are properly induced by the training process, leading to relatively high prediction skill on the testing set. Otherwise, the mechanisms induced in the trained model do not match those in the testing set, hence leading to poor skills. For this reason, we choose the most balanced tree with a relatively high AUC value for the extraction of decision rules and physical interpretation since such a model most likely involves all mechanisms properly for the generation of extreme precipitation events.

For SC-AMJ, the selected model is marked as Model 0, with an AUC value of 0.9 and a BSS value of 38% (Table 1), which is strongly suggested as skillful by the Wilcoxon-Mann-Whitney test ($p=0.015$). The corresponding decision tree is shown in Figure 5a. For this model, the numbers of training samples and testing samples are 39 and 15 respectively. The "above" label samples in the training set are the years of 1962, 1975, 1977, 1983, 1998 and 2006 while the remaining 3 "above" samples, 1973, 1995 and 2010, fall into the testing set. In the ROC curve of this model, the hit rate corresponding to the maximum accuracy (87%) point is 100% and the false alarm rate is 15% (Table 1). This means that such a model is able to discover all above-threshold extreme precipitation events at the cost of a small false alarm rate. We can also find that this model contains two leaf nodes with relatively large portions of the "above" sample. The paths leading to these nodes involve possible physical processes generating the extreme precipitation events. The first path (Path1_SC) is related to negative NAO phases in February and March (NAO_MAR $\leq$ -0.56→ NAO_FEB $\leq$ -0.47) while the second path (Path2_SC) does not necessarily need a negative phase of NAO in February but requires an El Niño state in preceding winter (NAO_MAR > -0.56→NINO34_DEC>1.04). The "above" sample of 2010 in the testing set falls into the leaf node of Path1_SC while the other two (1973 and 1985) end in the leaf node of Path2_SC.

For NCP-JA, the selected model is marked as Model 8, with an AUC value of 0.97and a BSS value of 51% (Table1), which is also significantly skillful over a random prediction following the Wilcoxon-Mann-Whitney test (p=0.003). The decision tree is shown in Figure 5b. The training set for this model includes 40 samples with 9 above-labeled years, 1962, 1964, 1969, 1973, 1985, 1988, 1996, 1998 and 2007; the remaining 14 samples with 3 above-labeled years, 1963, 1995 and 2010, form the testing set. The maximum accuracy point in the ROC curve holds a value of 0.93 and the corresponding hit rate and false alarm rate are 67% and 0% respectively (Table 1). There are also two leaf nodes with a relatively high ratio of above-labeled sample. The first (Path1_NCP) indicates a weak positive Niño state and an extremely negative phase of NAO in the preceding winter (NINO34_JAN ≤ 0.81 → NAO_DEC ≤ -1.28). The second (Path2_NCP) involves a transition from a Niño state in preceding winter to a weak cold phase in early summer (NINO34_JAN>0.81→ NINO34_JUN ≤ -0.13). The "above" sample of 1963 in the testing set falls into the leaf node of Path1_NCP while another one (2010) ends in the leaf node of Path2_NCP.

To make a physical understanding of the mechanisms generating the regional extreme precipitation, we pool all "above" samples from both training and testing sets in a leaf node for a composite analysis. For SC-AMJ, the "above" samples of 1975, 1977, 2006 and 2010 fall into the leaf node of Path1_SC, while those of 1973, 1983, 1995 and 1998 fall into the leaf node of Path2_SC. For NCP-JA, the leaf node of Path1_NCP contains the "above" samples of 1962, 1963 and 1996, while that of Path2_NCP contains the "above" sample years of 1964, 1969, 1973, 1988, 1998, 2007 and 2010.

## 5. Physical Interpretation

Warm season precipitation over East Asia is always associated with the strength and position of the western North Pacific subtropical high (WNPSH). To produce superfluous rainfall over this region in two months or a season, a steady position of the WNPSH and mostly steady cold air mass activities from the inland north are important conditions. Under such conditions, a fierce and persistent interaction between the humid warm southerlies and cold northerlies meet along the northwestern flank of the WNPSH, leading to persistent extreme rains in the region. This fact is exactly reflected in the decision rules of the present models.

For SC-AMJ, the composite SLP and 850 hPa wind fields of Path1_SC show that a weak anomalous anticyclone is located over the Philippine Sea, favorable for transporting moisture into South China by the significant southwesterlies along its northwest flank (Figure 6a). In this case, the WNPSH extends more westward than its climatological position (Figure 7a). Over the mid-high latitudes, a wave train extends from the North Atlantic to the North Pacific, with two significant anomalous highs over the Ural Mountains and a large area from the Okhotsk Sea to the Aleutian Islands, respectively, and an anomalous low in the Eurasian Continent in between (Figure 6c). Previous studies showed that such a mid-high latitude circulation pattern favors inland cold air masses intruding to southeastern China (Zhao et al., 1998). Thus, a combination of these low latitude and mid-high latitude circulation patterns result in more-than-usual persistent subtropical fronts over South China, leading to an extremely rainy season. To maintain such persistent circulation patterns, the ocean condition should play an important role. As the simultaneous SSTA distribution shows, weak cold anomalies occur in the central and eastern equatorial Pacific and expand northwestward to the southeastern Philippine Sea, but from the South China Sea to the eastern Philippine Sea SST anomalies are warm (Figure 6e). The cooling in the southeastern Philippine Sea enhances the anticyclone over the area and drives it to extend westward. Meanwhile, the North Atlantic Ocean demonstrates a tripole SSTA pattern with a strong positive center to the north of 50°N, a weak positive center to the south of 30°N, and a weak negative center in between (Figure 6e). The tripole pattern triggers the wave train over the mid-high latitudes, as demonstrated in previous studies (Watanabe 2004; Sung et al., 2006; Wu et al., 2009). This pattern is usually accompanied by a negative NAO phase as a result of air-sea interaction (Pan et al., 2005). An analysis of the evolution of the SSTA from January to June reveals that under the rules of Path1_SC (Figure 8), the tripole pattern exists as early as in the preceding winter and persists into early summer (Ogi et al., 2003 and 2004). According to previous studies, the mechanisms for this tripole pattern to persist change with seasons. In winter, the negative NAO and the tripole SSTA pattern are coupled by a positive feedback (Pan, 2005); while in spring, a negative NAO induces the tripole SSTA pattern then the pattern maintains itself into early summer through the ocean memory (Wu et al., 2009). Anyway, a preceding persistent negative NAO phase favors an increase of extreme precipitation over South China in the AMJ season. In Path1_SC, there is little SSTA developing or decaying in the tropical Pacific (Figure 8). It is suggested that the mid-high latitude circulation pattern induced by the tripole SSTA pattern in the North Atlantic favors cold air mass activities into eastern Asia in the preceding months, thus preventing the WNPSH from moving northward and keeping it to the southeast of southern China during the AMJ season.

The composite results of Path2_SC show a significant anomalous anticyclone over western North Pacific (Figure 6b). It is a much stronger anomalous anticyclone than in the case of Path1_SC, extending from the South China Sea to south of Japan. The significant southwesterlies along its northwestern flank transport moisture into South China. The WNPSH extends extremely westward into the South China Sea (Figure 7b). In the mid-high latitudes, there is a weak anomalous high over the Ural Mountains and a saddle over the Okhotsk Sea (Figure 6d). The composite SSTA shows a Niño state in the eastern tropical Pacific (Figure 6f). In fact, such a circulation pattern results from decaying El Niño (Wang et al., 2000). The evolution of the SSTA indicates that the preceding winter is featured by a strong El Niño, decaying but not totally disappearing until the early summer (Figure 9). There is no consensus on the mechanism for maintaining the western North Pacific anomalous anticyclone. Some studies suggested that the air-sea interaction between the anomalous anticyclone and the SSTA pattern during the decaying phase of El Niño could favor its persistence (Wang et al., 2000; Wang et al., 2003). Others suggested that the warming in the Indian Ocean during the decaying phase of El Niño should play a more important role (Xie et al., 2009; Wu et al., 2010). However, not all decaying El Niño events result in extreme precipitation over

SC-AMJ. The composite analysis shows that those years following Path2_SC without extreme precipitation over SC-AMJ are corresponding to the decaying of a central Pacific El Niño (Figures not shown). A central Pacific El Niño shifts the tropical heating center into the area near the international dateline, resulting in two descending centers to its west and east, respectively. The one in the west strengthens and shifts the WNPSH westward, exerting more control over South China (Yuan et al., 2012). Thus, a decaying central Pacific El Niño is not favorable for extreme precipitation over South China. Since there are limited samples for El Niño events, the decision tree model is unable to identify such a rule. Nevertheless, the decaying of El Niño remains as a good indicator for predicting extreme precipitation events over SC-AMJ.

For NCP-JA, the first path is also featured by preceding negative NAO states, but also on the condition that the preceding NINO34 index is negative. The evolution of the SSTA in the tropical Pacific verified this point (Figure 11). The simultaneous tripole SSTA pattern in the North Atlantic remains but tends to be vague in the composite map for July-August while the north Pacific shows a strong warm center (Figure 10e). Under such conditions, there remains the wave train of two anomalous highs and one anomalous low over the Eurasian Continent. The two anomalous highs are weak but the low over Mongolia is quite strong (Figure 10c). This circulation pattern favors cold air mass activities invading into northern China. In the mid-lower latitudes, the seasonal advance of the WNPSH favors the formation of fronts over NCP-JA. An anomalous anticyclone extends from southern China to Japan (Figure 10a), favorable for transporting strong moisture along its northwest flank into North China. Another route of moist transportation originated from the Indian Ocean, traveling through southwestern China then into North China (Figure 10a). The fronts formed by humid warm and cold air interaction produce extremely excessive precipitation in the region. It is noteworthy that, in this case, the WNPSH is much weaker than usual (Figure 7c).

For Path2_NCP, the decision rule involves a shift from positive SSTA anomalies in the eastern tropical Pacific in the preceding winter to negative anomalies in the early summer. A significant anomalous anticyclone is located to the south of Japan and extends westward to cover southern China (Figure 10b). A significant positive anomalous high corresponds to a large-scale blocking situation over the northern Pacific. Over Mongolia, there is a weak anomalous low (Figure 10d). These favor formation of fronts over NCP following similar reasons to those in Path1_NCP. The simultaneous SST anomalies in the eastern tropical Pacific show La Niña status (Figure 10f). Tracing the development of La Niña, we find that it follows the decay of El Niño from the preceding winter to the early summer (Figure 12). Such a fact was also noted by Li and Wang (2017), who applied a regression analysis regarding the extreme rainfall day index over North China (north of 30°N). Two connected anomalous anticyclones over western North Pacific are a typical result from a decaying El Niño (Wang at al., 2000). But different from that, the anomalous anticyclone over the western North Pacific more northwestward (Figure 7d), possibly due to developing of La Niña. During the development of La Niña, cooling in the eastern tropical Pacific and warming in the southeast of the Philippine Sea strengthen the Walker cell over the Pacific and force the WNPSH to extend northwestward. Consequently, the anomalous anticyclone over the western North Pacific occurs between the north of Philippine Sea and the south of Japan and stretches westward over southern China, leading to a rain belt shifting from the mid-lower reaches of the Yangtze River to NCP.

## 6. Summary and Discussions

By this study, we developed decision tree models to predict the seasonal extreme precipitation for two regions in eastern China. The DT models output a probability prediction of a “yes” or “no” extreme precipitation season. A series of preceding monthly and seasonal climate indices were used as the predictors. The experiments with different combinations of predictors suggested that the models involving ENSO and NAO indices as the predictors should be the best for the regional cases. The DT models demonstrated the main rules to generate extreme precipitation over the regions, with underlying physical processes understood via composite analyses of the same-route sample observations.

For SC-AMJ, there were two main paths leading to extreme precipitation. Path 1 involved a persistent negative NAO phase in February-March, coupled with a tripole SSTA pattern in the North Atlantic. The air-sea interaction and the memory of the ocean maintain the tripole SSTA pattern, which triggers a wave train over the mid-high latitude Eurasian continent. Such an anomalous circulation pattern favors cold air mass intruding into eastern China with persistent front formation over South China, hence causing extreme rains. Path 2 was featured by the El Niño state in the preceding winter, followed by a decaying phase of El Niño, leading to more-than-usual extreme precipitation over South China.

For NCP-JA, there are also two main paths leading to more-than-usual extreme precipitation. The first involves an extremely negative NAO phase in the preceding winter coupled with the tripole SSTA pattern persisting from the preceding winter to the early summer. These trigger a wave train including an extremely strong anomalous low over Mongolia and an anomalous high over the North Pacific. This circulation pattern favors cold air activities into northern China and front formation over NCP. The second path involves a shift from El Niño in the preceding winter to weak La Niña in the early summer. A decaying El Niño helps to maintain an anomalous anticyclone during the spring and early summer in the northwestern Pacific, which favors more-than-usual extreme precipitation over NCP. Although the monthly or seasonal climate indices are selected as predictors, the interpretation of the physical mechanisms for seasonal extreme precipitation is different from that for seasonal total precipitation revealed by previous studies. Here, we emphasize that a seasonal extreme precipitation event is the result of a combination of different preceding climate states that should be above or below some 'extreme' levels (e.g., NINO34_JAN >0.8 & NINO34_JUN < -0.13 indicate extreme precipitation event in NCP_JA). If the preceding climate states are outside these ranges, no extreme precipitation event will be triggered.

The present results also suggested that the seasonal extreme precipitation over eastern China should be closely related with typical SSTA patterns in the Pacific and the North Atlantic. It is reasonable to have ENSO indices as decisive predictors in the present model, as ENSO is the most important source of interannual variability of global climate. The DT model also captured the influence of SSTA in the North Atlantic on the atmospheric circulation over the far downstream regions. Moreover, we noted that the DT model incorporating only SAM indices also had some skill for prediction of extreme precipitation over SC-AMJ. This is in accordance with the study of Nan and Li (2003), but the mechanism needs further study.

For comparison, we tried to use the climatological mean of the extreme precipitation index as a threshold to define an extreme precipitation season and then applied the same procedures to build the DT models. However, the resultant models showed little skill. Considering that such an undertaking makes little difference from partitioning the total precipitation into more- and less-than-usual classes, we suggest that the seasonal extreme precipitation should be more predictable than the seasonal total precipitation is for the study regions. This point was also implied in some previous studies (e.g., Wang and Yan, 2011).

Caveats exist due to the limited observations in the present study. With limited samples, any statistical modeling, including the DT, is easily influenced by sampling uncertainty and should be understood with caution. The analyses of underlying physical processes did help validating the modeling. Insufficient samples also restrict the DT method to discover more accurate or complete decision path for the generation of an extreme event. One example has been shown above that not all decaying Pacific El Nino events result in extreme precipitation events over SC_AMJ. Another example is the incomplete description of the decadal or multidecadal change due to lack of samples. Wu and Wang (2002) had documented a decadal change of the relationship between the seasonal precipitation anomaly over North China and the mean SST anomaly over Nino3.4 region, where the correlation was positive during 1962-77 but shifted to negative during 1978-93. They further pointed out that the decadal change was possibly due to two anomalous heating sources: one from the Philippine Sea and the other from Indian. The present DT model for NCP_JA only integrated the positive relationship between the extreme precipitation event over this region and the anomalous Nino3.4 index, even though the mean SST anomaly in Philippine Sea was also used as a potential predictor. The reason is also probably due to a lack of samples which prevents the DT model from discovering the modulating effects from other factors. Another issue arises from the use of accumulated extreme precipitation indices which probably mixes up extreme events induced by different weather or circulation systems (e.g., frontal system and landfall tropical cyclone). A possible solution is to model the extreme events from different sources separately.

Nevertheless, the DT method used here demonstrated great potential of skillful seasonal prediction of the regional extreme precipitation, with quite consistent performance even with limited samples. It is hopeful to incorporate more physical factors / mechanisms in the DT models with increasing observations, so as to improve the predictive performance with time.

**Acknowledgements:** This study was supported by the MOST project (2016YFA0600400) and the CAS projects (XDA20020201, 134111KYSB20160010 & 134111KYSB20160028).

## References

Fan, K., Wang, H., & Choi, Y. J. (2008). A physically-based statistical forecast model for the middle-lower reaches of the Yangtze River Valley summer rainfall. Chinese Science Bulletin, 53(4), 602-609.https://doi.org/10.1007/s11434-008-0083-1

Gong, D. Y., & Ho, C. H. (2003). Arctic oscillation signals in the East Asian summer monsoon. Journal of Geophysical Research: Atmospheres, 108(D2). https://doi.org/10.1029/2002JD002193

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques (3rd ed.). Elsevier.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statistical Learning; Data Mining, Inference and Prediction (2nd ed.). New York: Springer series in statistics.

Ishii, M., Shouji, A., Sugimoto, S., & Matsumoto, T. (2005). Objective analyses of sea‐surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. International Journal of Climatology, 25(7), 865-879. https://doi.org/10.1002/joc.1169

Jia, X., Chen, L., Gong, Z., Wang, Z., Yuan, Y., & Sun, C. (2011). Anomalies of ocean and atmospheric circulation in 2010 and their impacts on climate in China. Meteorological Monthly (in Chinese), 37(4), 446-453.doi: 10.7519/j.issn.1000-0526.2011.4.008

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... & Zhu, Y. (1996). The NCEP/NCAR 40-year reanalysis project. Bulletin of the American meteorological Society, 77(3), 437-471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Li, C., Tian, Q., Yu, R., Zhou, B., Xia, J., Burke, C., ... & Ciavarella, A. (2018). Attribution of extreme precipitation in the lower reaches of the Yangtze River during May 2016. Environmental Research Letters, 13(1), 014015.https://doi.org/10.1088/1748-9326/aa9691

Li, J., & Wang, B. (2017). Predictability of summer extreme precipitation days over eastern China. Climate Dynamics, 1-12. https://doi.org/10.1007/s00382-017-3848-x

Metz, C. E. (1978). Basic principles of ROC analysis. In Seminars in nuclear medicine (Vol. 8, No. 4, pp. 283-298). Elsevier. https://doi.org/10.1016/S0001-2998(78)80014-2

Nan, S., & Li, J. (2003). The relationship between the summer precipitation in the Yangtze River valley and the boreal spring Southern Hemisphere annular mode.

Geophysical Research Letters, 30(24). https://doi.org/10.1029/2003GL018381

National Climate Center of China. (1998). China's 1998 Severe Flood and Climate Extremes. Beijing: China Meteorological Press,152pp (in Chinese).

Ogi, M., Tachibana, Y., & Yamazaki, K. (2003). Impact of the wintertime North Atlantic Oscillation (NAO) on the summertime atmospheric circulation. Geophysical Research Letters, 30(13). https://doi.org/10.1029/2003GL017280

Ogi, M., Tachibana, Y., & Yamazaki, K. (2004). The connectivity of the winter North Atlantic Oscillation (NAO) and the summer Okhotsk High. Journal of the Meteorological Society of Japan. Ser. II, 82(3), 905-913. https://doi.org/10.2151/jmsj.2004.905

Orsolini, Y. J., Zhang, L., Peters, D. H., Fraedrich, K., Zhu, X., Schneidereit, A., & van den Hurk, B. (2015). Extreme precipitation events over north China in August 2010 and their link to eastward‐propagating wave‐trains across Eurasia: observations and monthly forecasting. Quarterly Journal of the Royal Meteorological Society, 141(693), 3097-3105.https://doi.org/10.1002/qj.2594

Pan, L. (2005). Observed positive feedback between the NAO and the North Atlantic SSTA tripole. Geophysical research letters, 32(6)., https://doi.org/10.1029/2005GL022427

Pei, L., Xia, J., Yan, Z., & Yang, H. (2017). Assessment of the Pacific decadal oscillation's contribution to the occurrence of local torrential rainfall in north China. Climatic Change, 144(3), 391-403., https://doi.org/10.1007/s10584-016-1610-8

Qian, W., Fu, J., & Yan, Z. (2007). Decrease of light rain events in summer associated with a warming environment in China during 1961–2005. Geophysical Research Letters, 34(11). https://doi.org/10.1029/2007GL029631

Shannon, C. E. (1948). A mathematical theory of communication. Bell system technical journal, 27(3), 379-423.https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Si, D., & Ding, Y. (2016). Oceanic forcings of the interdecadal variability in East Asian summer rainfall. Journal of Climate, 29(21), 7633-7649. https://doi.org/10.1175/JCLI-D-15-0792.1

Sung, M. K., Kwon, W. T., Baek, H. J., Boo, K. O., Lim, G. H., & Kug, J. S. (2006). A possible impact of the North Atlantic Oscillation on the east Asian summer monsoon precipitation. Geophysical research letters, 33(21). https://doi.org/10.1029/2006GL027253

Wang, B., Lee, J. Y., Kang, I. S., Shukla, J., Park, C. K., Kumar, A., ... & Zhou, T. (2009). Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Climate Dynamics, 33(1), 93-117.https://doi.org/10.1007/s00382-008-0460-0

Wang, B., Li, J., & He, Q. (2017). Variable and robust East Asian monsoon rainfall response to El Niño over the past 60 years (1957–2016). Advances in Atmospheric Sciences, 34(10), 1235-1248.https://doi.org/10.1007/s00376-017-7016-3

Wang, B., Wu, R., & Fu, X. (2000). Pacific–East Asian teleconnection: how does ENSO affect East Asian climate?. Journal of Climate, 13(9), 1517-1536.https://doi.org/10.1175/1520-0442(2000)013<1517:PEATHD>2.0.CO;2

Wang, B., Wu, R., & Li, T. (2003). Atmosphere–warm ocean interaction and its impacts on Asian–Australian monsoon variation. Journal of Climate, 16(8), 1195-1211.https://doi.org/10.1175/1520-0442(2003)16<1195:AOIAII>2.0.CO;2

Wang, H., Fan, K., Sun, J., Li, S., Lin, Z., Zhou, G., ... & Chen, H. (2015). A review of seasonal climate prediction research in China. Advances in Atmospheric Sciences, 32(2), 149-168.https://doi.org/10.1007/s00376-014-0016-7

Wang, L., & Chen, W. (2014). An intensity index for the East Asian winter monsoon.

Journal of Climate, 27(6), 2361-2374. https://doi.org/10.1175/JCLI-D-13-00086.1

Wang, Y., & Yan, Z. (2011). Changes of frequency of summer precipitation extremes over the Yangtze River in association with large-scale oceanic-atmospheric conditions. Advances in Atmospheric Sciences, 28(5), 1118. https://doi.org/10.1007/s00376-010-0128-7

Watanabe, M. (2004). Asian jet waveguide and a downstream extension of the North Atlantic Oscillation. Journal of Climate, 17(24), 4674-4691.https://doi.org/10.1175/JCLI-3228.1

Watanabe, M., Kimoto, M., Nitta, T., & Kachi, M. (1999). A comparison of decadal climate oscillations in the North Atlantic detected in observations and a coupled GCM. Journal of climate, 12(9), 2920-2940. https://doi.org/10.1175/1520-0442(1999)012<2920:ACODCO>2.0.CO;2

Wei, W., Yan, Z., & Jones, P. D. (2017). Potential predictability of seasonal extreme precipitation accumulation in China. Journal of Hydrometeorology, 18(4), 1071-1080.https://doi.org/10.1175/JHM-D-16-0141.1

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Wu, B., Li, T., & Zhou, T. (2010). Relative contributions of the Indian Ocean and local SST anomalies to the maintenance of the western North Pacific anomalous anticyclone during the El Niño decaying summer. Journal of Climate, 23(11), 2974-2986.https://doi.org/10.1175/2010JCLI3300.1

Wu, R., & Wang, B. (2002). A contrast of the East Asian summer monsoon–ENSO relationship between 1962–77 and 1978–93. Journal of Climate, 15(22), 3266-3279. https://doi.org/10.1175/1520-0442(2002)015<3266:ACOTEA>2.0.CO;2

Wu, Z., Wang, B., Li, J., & Jin, F. F. (2009). An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO. Journal of Geophysical Research: Atmospheres, 114(D18). https://doi.org/10.1029/2009JD011733

Xie, S. P., Hu, K., Hafner, J., Tokinaga, H., Du, Y., Huang, G., & Sampe, T. (2009). Indian Ocean capacitor effect on Indo–western Pacific climate during the summer following El Niño. Journal of Climate, 22(3), 730-747. https://doi.org/10.1175/2008JCLI2544.1

Yan, Z., &Yang, C. (2000). Geographic Patterns of Extreme Climate Changes in China during 1951-1997. Climatic and Environmental Research (in Chinese), 3, 004.doi: 10.3878/j.issn.1006-9585.2000.03.05

Yang, Q., Ma, Z., Fan, X., Yang, Z. L., Xu, Z., & Wu, P. (2017). Decadal modulation of precipitation patterns over eastern China by sea surface temperature anomalies. Journal of Climate, 30(17), 7017-7033. https://doi.org/10.1175/JCLI-D-16-0793.1

Yim, S. Y., Wang, B., & Xing, W. (2014). Prediction of early summer rainfall over South China by a physical-empirical model. Climate dynamics, 43(7-8), 1883-1891.https://doi.org/10.1007/s00382-013-2014-3

Yuan, Y., Gao, H., Li, W., Liu, Y., Chen, L., Zhou, B., Ding, Y.(2017). Analysis and Comparison of summer precipitation features and physical mechanisms between 2016 and 1998. Acta Meteorologica Sinica (in Chinese), 75 (1): 19-38. doi: 10.11676/qxxb2017.019

Yuan, Y., Yang, H., & Li, C. Y. (2012). Study of El Niño events of different types and their potential impact on the following summer precipitation in China. Acta Meteorologica Sinica (in Chinese), 70, 467-478.doi: 10.11676/qxxb2012.039

Zhang, J., Li, D., Li, L., & Deng, W. (2013). Decadal variability of droughts and floods in the Yellow River basin during the last five centuries and relations with the North Atlantic SST. International Journal of Climatology, 33(15), 3217-3228. https://doi.org/10.1002/joc.3662

Zhao, S., Sun, J., Chen, H., &Zhang, F. (1998). Study of Heavy Rainfall in the Changjiang River during July 1998. Climatic and Environmental Research(in Chinese),3(4):368-381.doi: 10.3878/j.issn.1006-9585.1998.04.09

Zhu, Y., & X. Yang. (2003). Relationships between Pacific decadal oscillation (PDO) and climate variabilities in China (in Chinese). Acta Meteor. Sin., 61, 641–654.

Zhu, Y., Wang, T., & Ma, J. (2016). Influence of internal decadal variability on the summer rainfall in eastern China as simulated by CCSM4. Advances in Atmospheric Sciences, 33(6), 706-714. https://doi.org/10.1007/s00376-016-5269-x

**Tables**

Table 1. Performance of the most balanced models for the two regions. The results are calculated based on the predictions of the testing sets. AUC means area under the ROC curve; BSS is the Brier skill score calculated using climatological probabilities as the reference forecasts. Accuracy, hit rate and false alarm rate are the measures corresponding to the cut-off point with maximum accuracy.

| | AUC | BSS | Accuracy | Hit rate | False alarm rate |
|---|---|---|---|---|---|
| SC_AMJ | 0.90 | 38% | 87% | 100% | 15% |
| NCP_JA | 0.97 | 51% | 93% | 67% | 0% |

# Figure captions

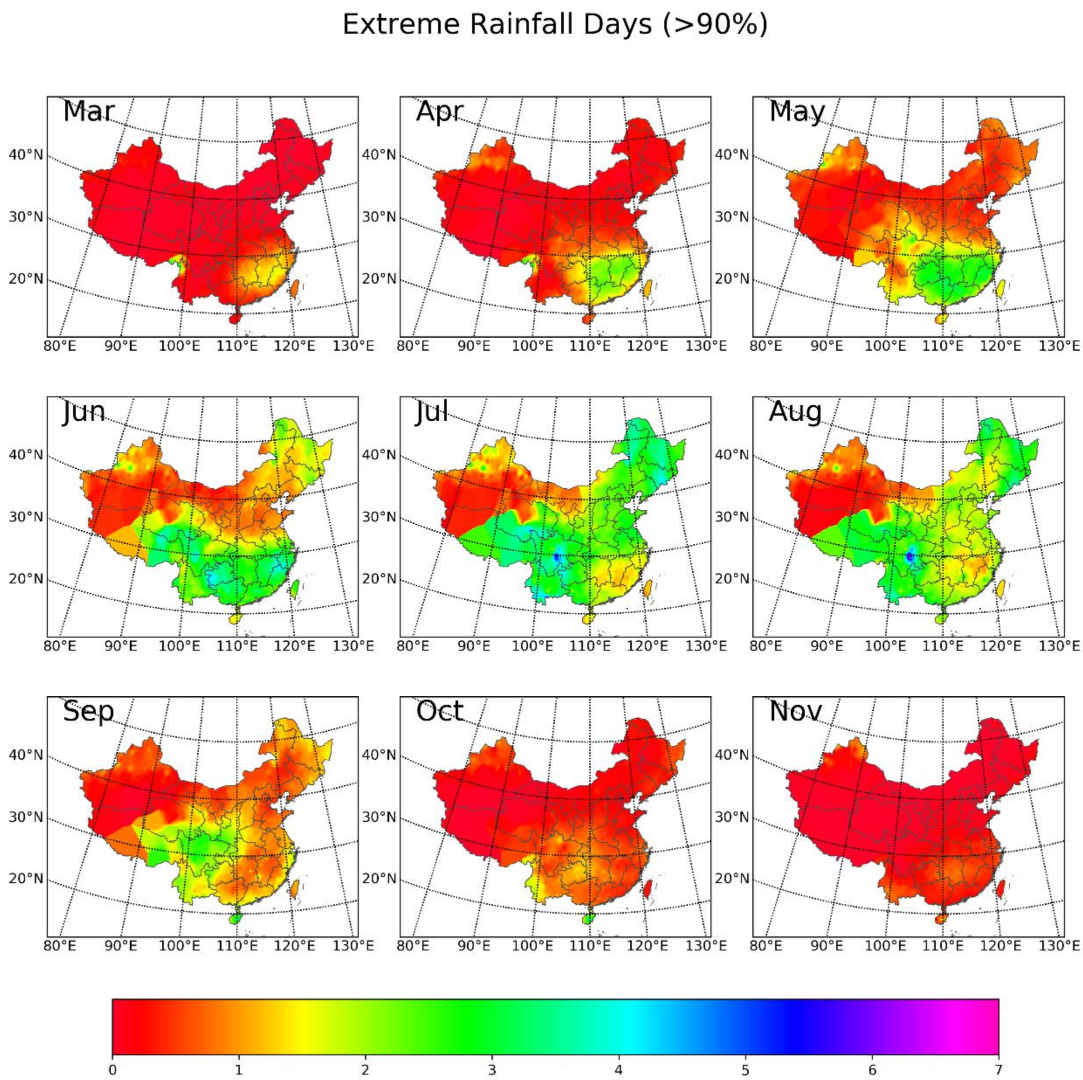


Figure 1. The climatological distribution of extreme precipitation days for each month. An extreme precipitation day is defined as the one whose daily precipitation amount is larger than the 90th percentile of the daily precipitation distribution. Results from December, January and February are not shown since there are no extreme precipitation days in these three months.

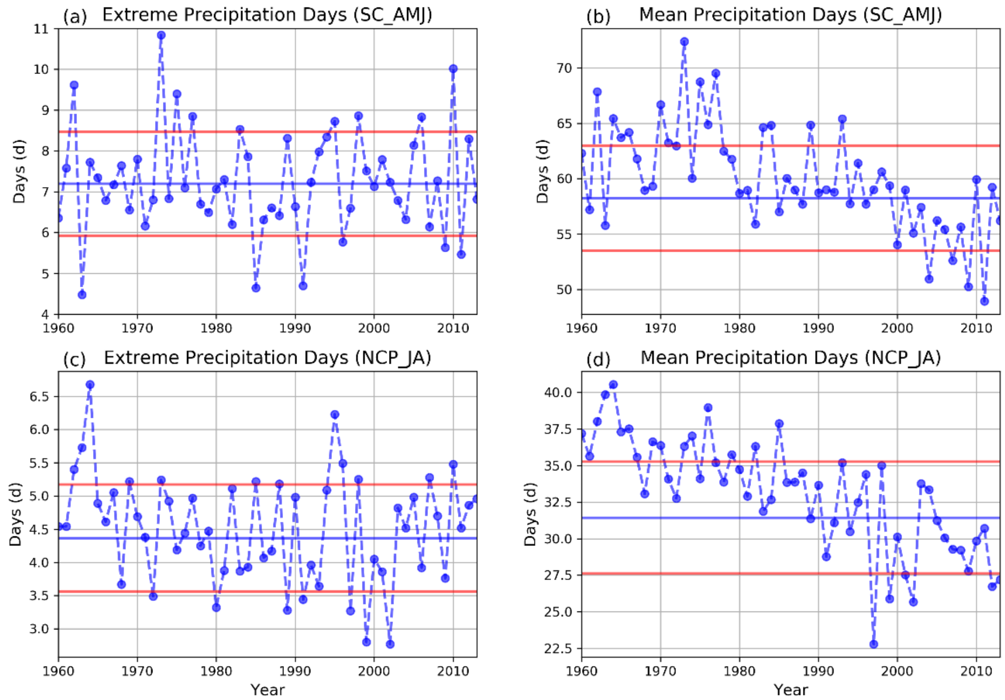


Figure 2. The spatial-temporal mean precipitation indices: a) extreme precipitation days for SC_AMJ, b) precipitation days for SC_AMJ, c) extreme precipitation days for NCP_JA and d) precipitation days for NCP_JA.

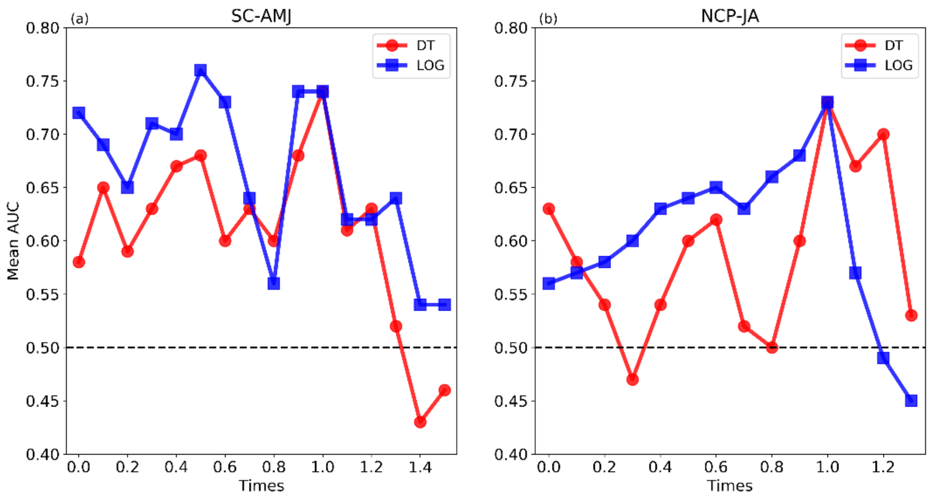


Figure 3. The 12-model mean AUC changing with varying threshold to define an extreme precipitation year, red lines for DT model while blue lines for binary logistic regression model, a) for SC-AMJ, b) for NCP-JA. A mean AUC is calculated using the

mean ROC curve from the 12 models. A threshold is represented as how many times standard deviations above the climatological mean. The black dash lines show the position of an AUC of 0.5 which indicates no prediction skill.

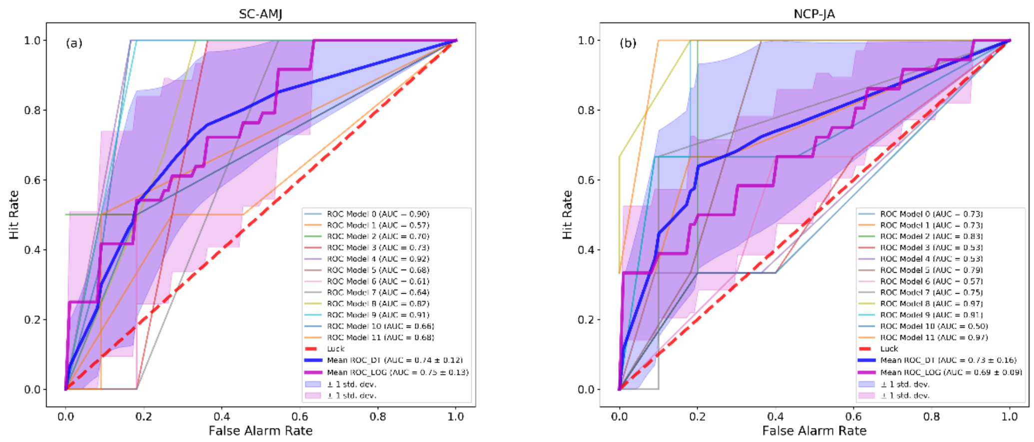


Figure 4. The ROC curves for the experiments with ENSO and NAO indices as the predictors, a) for SC-AMJ and b) for NCP-JA. Each thin line represents the ROC curve from one of the 12 DT models. The blue heavy line is the mean ROC curve of the 12 DT models, from which the mean AUC is calculated. The red heavy line is the mean ROC curve of the 12 binary logistic regression models. The ROC curves for the 12 binary logistic regression models are not shown here. The red dash line represents the ROC curve from a random prediction model of no skill. The blue and red areas show the standard errors of mean ROC for DT model and binary logistic regression model respectively.

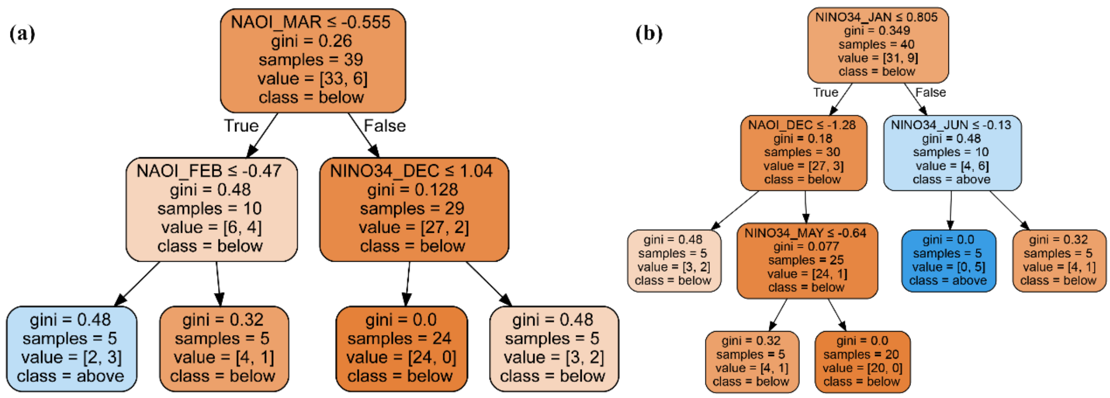


Figure 5. The most balanced decision trees corresponding to Figures 4a and 4b, named as Model 0 and Model 8, respectively. The first line in a non-leaf node (e.g. NAOI_MAR ⩽ -0.555) is the statement to generate a binary branch. A "true" answer to this statement always leads to the left branch while the right branch is arrived following a "false" answer.

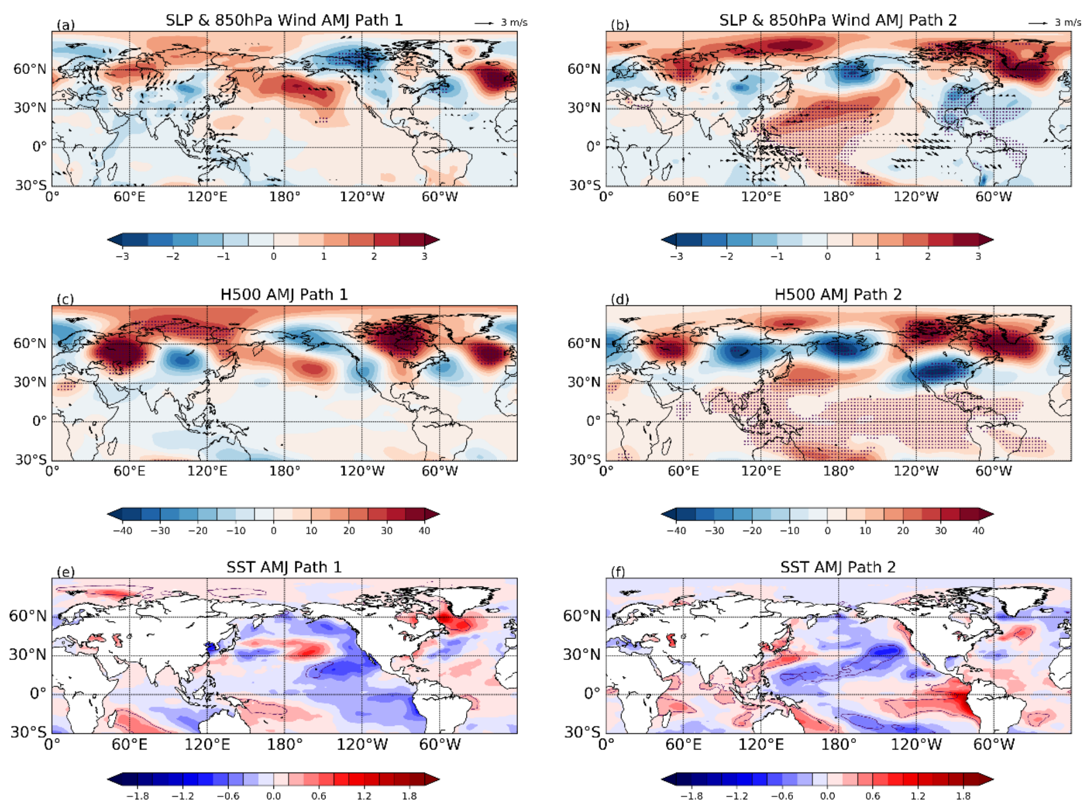


Figure 6. Simultaneous composite results for the two main paths in Model 0 for SC-AMJ. (a) and (b) for SLP anomalies (shaded area, units: hPa) and horizontal wind

anomalies at 850hPa (arrows, units: m/s); (c) and (d) for the geopotential height anomalies at 500 hPa (units: gpm); (e) and (f) for the SST anomalies (units: K). The left column for path 1 and the right column for path 2. The wind vectors, dotted areas (for SLP and H500) and areas encircled by black lines (for SST) are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the paths.

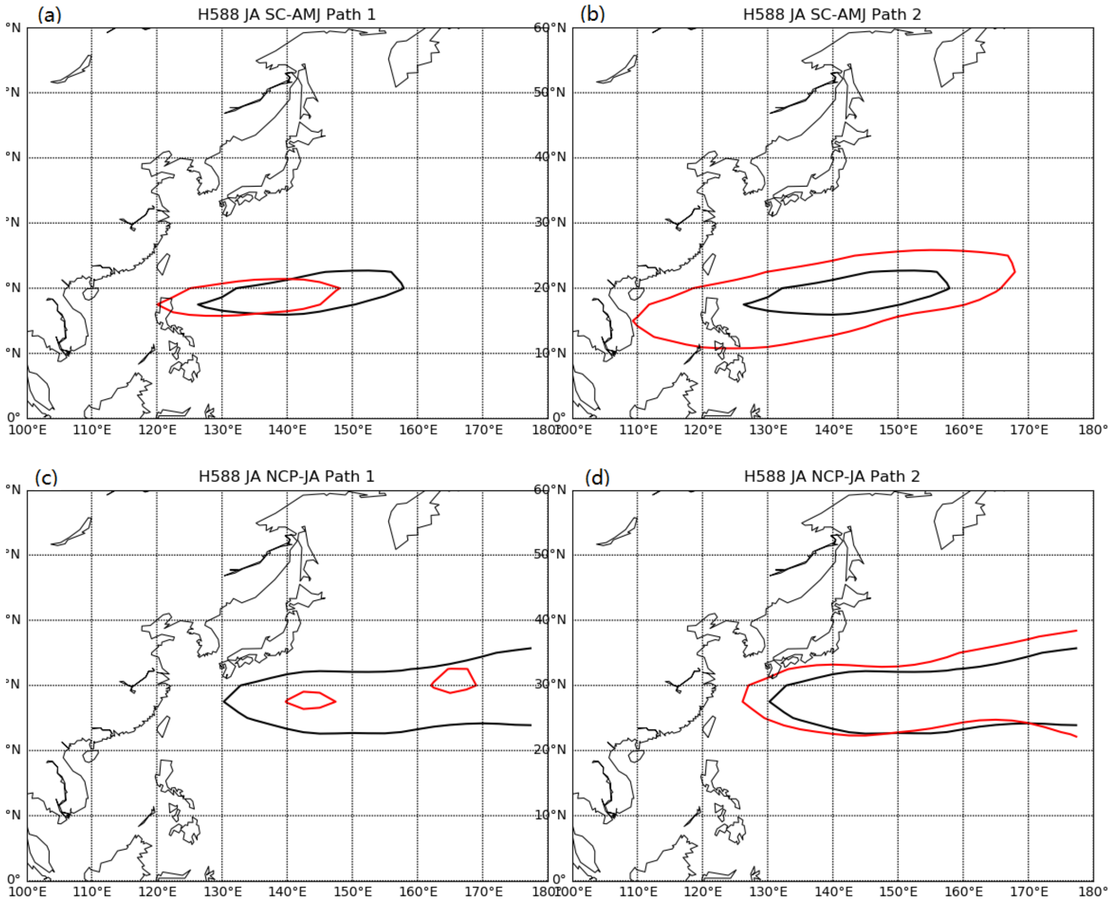


Figure 7. The simultaneous composite isopleth of 5880 gpm at 500 hPa level. The red lines represent the composite results and the black lines represent the climatology. (a) Path1_SC, (b) Path2_SC, (c) Path1_NCP, and (d) Path2_NCP.

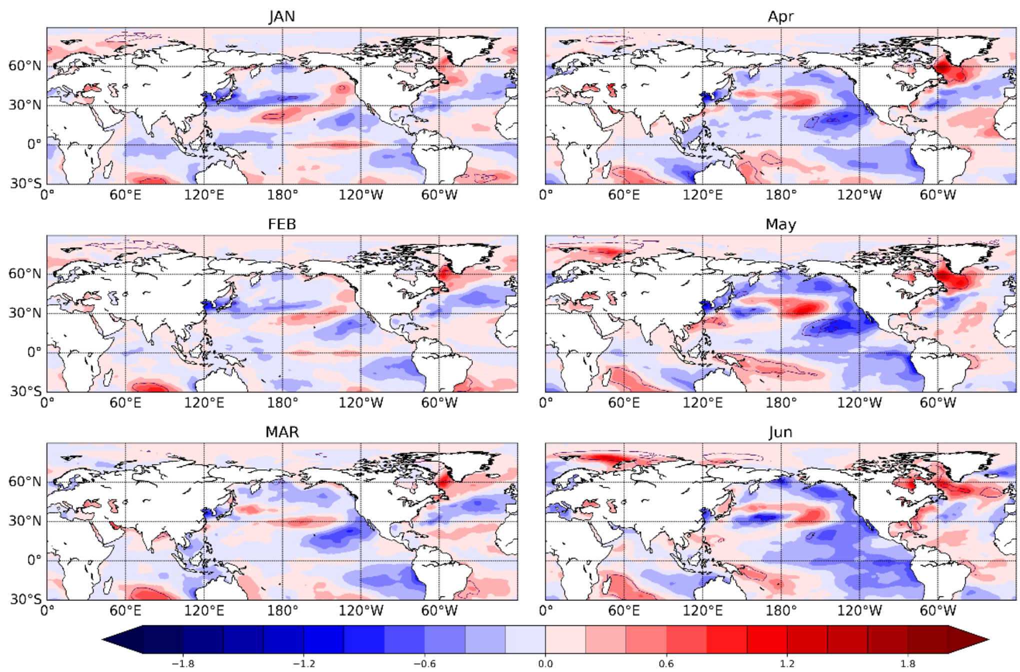


Figure 8. The evolution of monthly SST anomalies (units: K) for Path1_SC in Model 0 for SC-AMJ. Areas encircled by black lines are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the path.

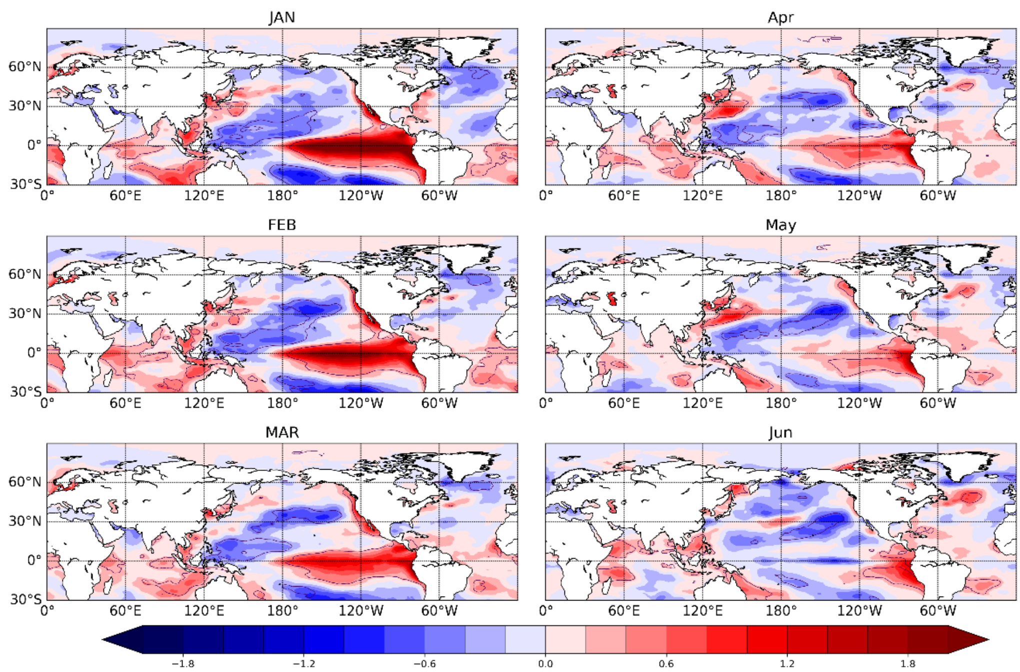


Figure 9. The evolution of monthly SST anomalies (units: K) for Path2_SC in Model 0

for SC-AMJ. Areas encircled by black lines are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the path.

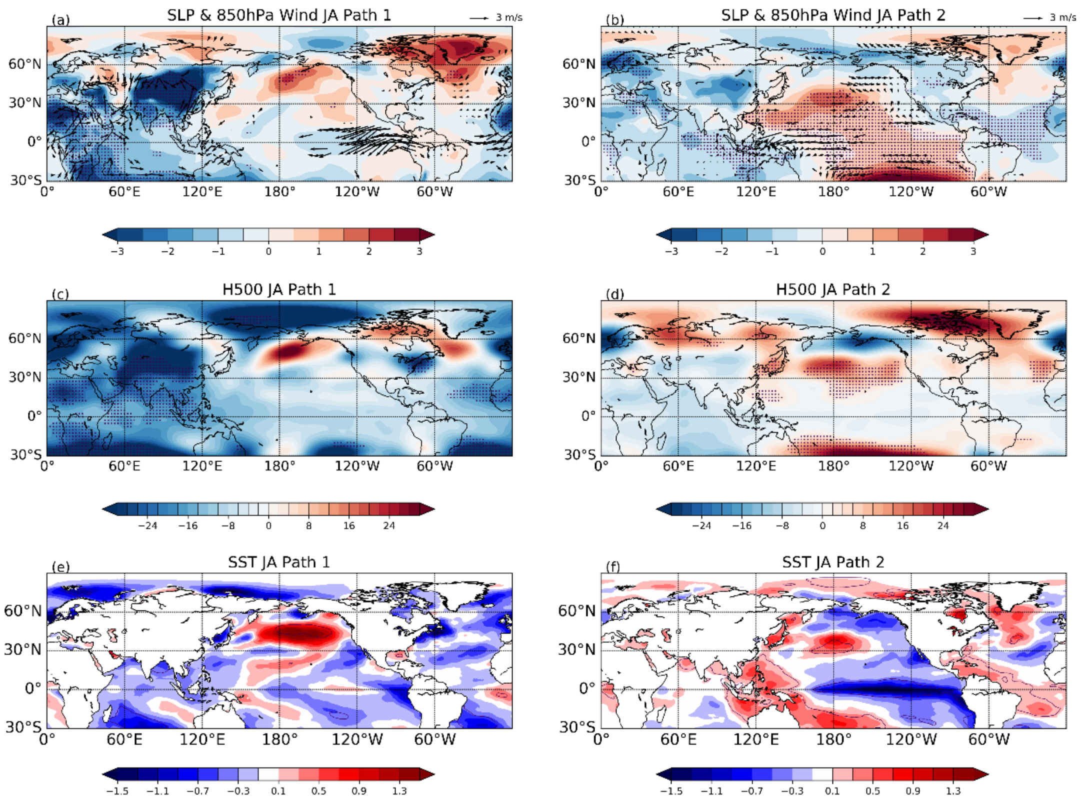


Figure 10. Simultaneous composite results for the two main paths in Model 8 for NCP-JA. (a) and (b) for SLP anomalies (shaded area, units: hPa) and horizontal wind anomalies at 850hPa (arrows, units: m/s); (c) and (d) for the geopotential height anomalies at 500 hPa (units: gpm); (e) and (f) for the SST anomalies (units: K). The left column for path 1 and the right column for path 2. The wind vectors, dotted areas (for SLP and H500) and areas encircled by black lines (for SST) are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the paths.

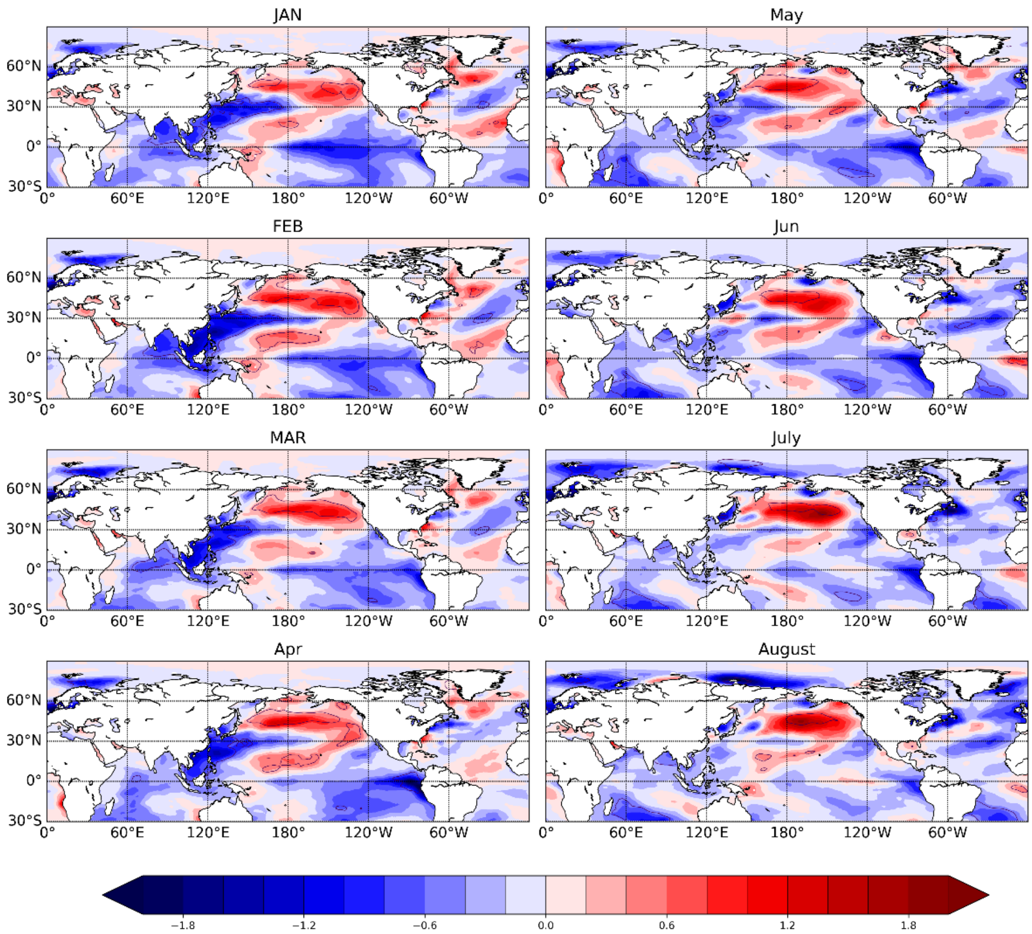


Figure 11. The evolution of monthly SST anomalies (units: K) for Path1_NCP in Model 8 for NCP-JA. Areas encircled by black lines are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the path.

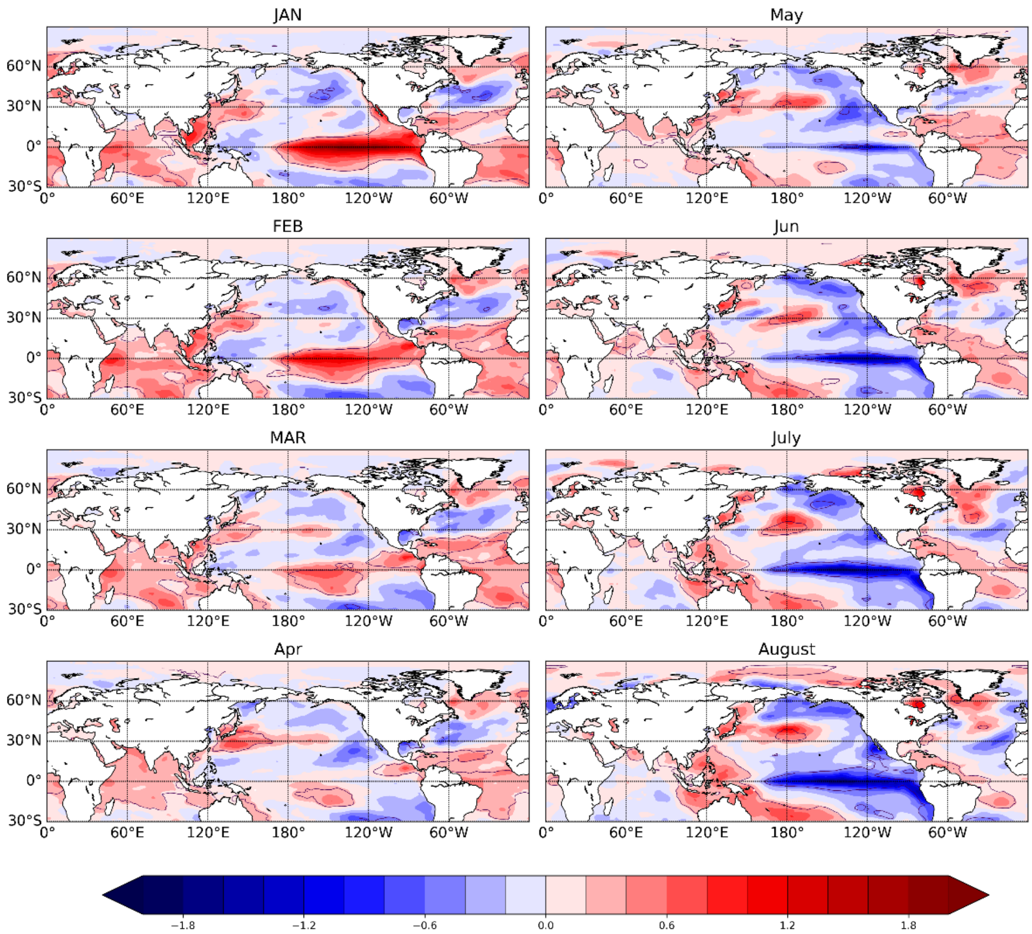


Figure 12. The evolution of monthly SST anomalies (units: K) for Path2_NCP in Model 8 for NCP-JA. Areas encircled by black lines are statistically significant using a t-test at the significance level of 0.05 for the hypothesis of no difference between the samples following and not following the path.