Machine Learning Models for the Seasonal Forecast of Winter Surface Air Temperature in North America

Qi Feng Qian1,2, Xiao Jing Jia1,2, and Hai Lin1,3

1Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou, China, 2Atmospheric Sciences Department, School of Earth Sciences, Zhejiang University, Hangzhou, China, 3Recherche en Prévision Numérique Atmosphérique, Environment and Climate Change Canada, Dorval, Quebec, Canada

Abstract In this study, two machine learning (ML) models (support vector regression (SVR) and extreme gradient boosting (XGBoost)) are developed to perform seasonal forecasts of the surface air temperature (SAT) in winter (December-January-February, DJF) in North America (NA). The seasonal forecast skills of the two ML models are evaluated via cross validation. The forecast results from one linear regression (LR) model, and two dynamic climate models are used for comparison. In the take-one-out hindcast experiment, the two ML models and the LR model show reasonable seasonal forecast skills for winter SAT in NA. Compared to the two dynamic models, the two ML models and the LR model have better forecast skill for the winter SAT over central NA, which is mainly derived from a skillful forecast of the second empirical orthogonal function (EOF) mode of winter SAT over NA. In general, the SVR model and XGBoost model hindcasts show better forecast performances than the LR model. However, the LR model shows less dependence on the size of the training data set than the SVR and XGBoost models. In the real forecast experiments during the period of 2011–2017, the two ML models exhibit better forecasting skills for the winter SAT over northern and central NA than do the two dynamic models. The results of this study suggest that the ML models may provide improved forecasting skill for seasonal forecasts of the winter climate in NA.

1. Introduction
A skillful seasonal forecast of winter temperature is of great value because large-scale seasonal surface air temperature (SAT) anomalies have a significant impact on societal and economic activities across North America (NA) (e.g., Barnston & Ropelewski, 1992; Derome et al., 2001; Lin & Wu, 2011; Seager et al., 2010; Scaife et al., 2014). The day-to-day variations in the atmosphere cannot be predicted beyond 2 weeks (Lorenz, 1963), and for both statistical and dynamical models, most of the seasonal predictability originates from coupled climate mechanisms involving lower-boundary forcing (e.g., Charney & Shukla, 1981; Cohen et al., 2019; Lin & Wu, 2011) or low-frequency, large-scale circulation patterns. Compared to the atmosphere, the ocean has relatively large thermal and mechanical inertia and correspondingly long timescales; thus, it can impart to the atmosphere a higher degree of predictability than would be present otherwise. For example, the El Niño/Southern Oscillation (ENSO) can affect the winter climate over NA through tropical diabatic forcing-related atmospheric teleconnections, which is one of the most important sources of predictability (e.g., Barnston & Ropelewski, 1992; Barnston et al., 2012; Hurrell, 1996; Ropelewski & Halpert, 1986; Shabbar & Khandekar, 1996). In addition to ENSO, the North Atlantic Oscillation (NAO) is another important factor that can affect climate variation over NA and is used as one of the predictors in seasonal forecasts for NA (e.g., Hurrell, 1995; Lin & Wu, 2011; Seager et al., 2010; Wallace & Gutzler, 1981). Coherent with the NAO, the tripole sea surface temperature anomalies (SSTAs) in the North Atlantic Ocean may also be one of the potential sources of seasonal predictability (e.g., Lin & Wu, 2011; Rodwell et al., 1999). Some studies have revealed that the SSTAs over the Indian Ocean can also act as a potential indicator of the long-term Northern Hemisphere extratropical circulation variability (e.g., Bader & Latif, 2005; Hoerling et al., 2004).

In addition to the ocean, sensible heat flux from the land surface may also provide a potential source of predictability for NA climate prediction (e.g., Gong et al., 2002; Gong et al., 2004; Sobolowski et al., 2010). For example, the autumn snow cover on the Tibetan Plateau (TP) has been revealed to affect the winter SAT over...
NA through thermally forced Rossby wave trains related to snow cover changes (Lin & Wu, 2011; Qian et al., 2019). This suggests that the TP snow cover can also serve as a useful predictor for the seasonal prediction of NA winter SAT and extreme events (e.g., Lin & Wu, 2012).

Compared to dynamical models, statistical or empirical methods are appealing for use in seasonal climate predictions due to their low computational requirements. Moreover, the current operational dynamic models do not always perform well in forecasting climate variables over certain areas. For example, the second version of the Canadian Seasonal to Interannual Prediction System (CanSIPSv2) shows limited skill in lead-zero and lead-one forecasts of the winter SAT over central and western NA. Therefore, it remains necessary to investigate statistical models, which may contribute to improving the seasonal forecast skill.

With the development of computer science, machine learning (ML) models have begun to greatly influence many fields (Reichstein et al., 2019) and have been applied to both weather forecasting (e.g., Chattopadhyay et al., 2020; Weyn et al., 2019) and climate prediction (e.g., Badr et al., 2014; Chi & Kim, 2017; Cohen et al., 2019; Ham et al., 2019; Hwang et al., 2019; Iglesias et al., 2015; Kämäräinen et al., 2019; Nooteboom et al., 2018; Richman & Leslie, 2012). Some newly developed ML models show great power in improving the monthly and subseasonal forecast skills of dynamic models and may even outperform dynamic models. For example, Nooteboom et al. (2018) proposed a hybrid ML model based on an autoregressive integrated moving average (ARIMA) model and an artificial neural network (ANN) to forecast monthly ENSO indices, and this model was found to outperform the ensemble mean forecast of the Climate Forecast System version 2 (CFSv2). Hwang et al. (2019) constructed two ML models and showed that the ensemble of the two ML models and CFSv2 exhibit significantly better subseasonal temperature and precipitation prediction skills over NA than the CFSv2 alone. Ham et al. (2019) applied the transfer learning technique to a convolutional neural network (CNN) and showed that the CNN model outperforms the current state-of-the-art dynamic forecast system in monthly ENSO predictions. Compared with the application of ML models in weather forecasting, ML models have not been applied widely in seasonal climate forecasts mainly due to the limited amount of training data available (Kämäräinen et al., 2019). Specifically, under the circumstance of seasonal climate forecasting, the available number of years (i.e., the number of samples) is insufficient. Thus, the first goal of the present work is to apply ML models for SAT prediction in winter (December-January-February, DJF) in NA using autumn (September-October-November, SON) predictors and to then show that ML models trained with even small data sets can provide improved skill compared to dynamic model forecasts. The second goal of this study is to compare the ML models to the traditionally used linear regression (LR) model, and the third goal is to assess the sensitivity of the ML model forecasting skill to the size of the training data set.

The paper is organized as follows. Section 2 describes the data and the two Canadian dynamic models used in this study. Section 3 describes the methods of constructing the two ML models in this study. Section 4 describes the climate predictor selection process. Section 5 presents the results of the hindcast and forecast experiments. Finally, a summary and discussion are presented in section 6.

2. Data and the Dynamic Models

2.1. Data

This study uses monthly SAT data on T62 Gaussian grids from 1979 to the present obtained from the National Centers for Environmental Prediction (NCEP)-Department of Energy (DOE) Reanalysis II (Kanamitsu et al., 2002). The reanalysis data are provided by the National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Division and are used to construct the ML models in this study. Additionally, monthly mean SAT data (Version TS4.03) from the Climate Research Unit (CRU) of the University of East Anglia (Harris et al., 2014), with a spatial resolution of 0.5° × 0.5°, for the period 1971–2018 are utilized to evaluate the performance of the models.

The Rutgers University Global Snow Lab Weekly Northern Hemisphere Snow Cover Extent Data (Robinson et al., 2012), with a horizontal resolution of 25 km and a time span from October 1966 to the present are used in this study. These snow data are converted to monthly mean data following Wang, Wu, and Huang (2017). Monthly mean SST data gridded at a 1° × 1° resolution were obtained from the Met Office Hadley Center (Rayner et al., 2003) and cover the period from 1870 to the present.
The Niño3.4 index, Niño4 index, Niño3 index, and Niño1 + 2 index are defined as the mean SSTAs over the Niño3.4 region (5°S to 5°N, 170°E to 150°W), Niño4 region (5°S to 5°N, 160°E to 150°W), Niño3 region (5°S to 5°N, 150°–90°W), and Niño1 + 2 region (0–10°S, 90°–80°W), respectively. The Arctic Oscillation (AO) index is constructed by projecting monthly mean 1,000-hPa height anomalies onto the loading pattern of the AO, which is defined as the first leading mode from the EOF analysis of monthly mean 1,000-hPa height anomalies. The NAO index and the Pacific-North America (PNA) index are defined by applying the rotated principal component analysis technique to monthly mean standardized 500-mb height anomalies, as described in Barnston and Livezey (1987). All these climate indices are obtained from the Earth System Research Laboratory of NOAA and have a time span from 1948 to the present.

2.2. The Canadian Dynamical Seasonal Forecast Models

The two atmosphere-ocean coupled dynamic models are from Environment and Climate Change Canada (ECCC). The two models are the Canadian Center for Climate Modelling and Analysis (CCCma) Coupled Climate Model version 4 with improved sea ice initialization (CanCM4i, hereafter) (Merryfield et al., 2013) and a newly developed global atmosphere-ocean coupled model (Smith et al., 2018) based on the Global Environmental Multiscale model (GEM), the Nucleus for European Modelling of the Ocean (NEMO) and the Community of Ice Code sea ice model (CICE). The latter is referred to as GEM_NEMO in this study. By applying a multimodel ensemble approach, CanCM4i and GEM_NEMO together form the second version of the Canadian Seasonal to Interannual Prediction System (CanSIPSv2). The atmospheric component of GEM-NEMO has a horizontal resolution of 1.4° and 79 vertical levels with the top at 0.075 hPa, while CanCM4i is T63, corresponding to a resolution of about 1.9°, with 35 vertical levels and a top at 1 hPa. Both CanCM4i and GEM_NEMO are initialized at the beginning of each month and produce forecasts of 10 ensemble members with all the integrations starting from the same date and lasting for 12 months (Lin et al., 2020). The outputs of the CanCM4i and GEM_NEMO hindcast with a resolution of 1° x 1° and a time span from 1981 to 2010 are obtained from the website (https://weather.gc.ca/grib/grib2_cansips_e.html). In addition to the hindcast outputs for 1981–2010, CanCM4i and GEM_NEMO forecast outputs with a resolution of 1° x 1° are also available from 2011 to 2018 and are downloaded from the website (https://collaboration.cmc.ec.gc.ca/cmc/CMOI/GRIB/NMME/1p0deg/). The model outputs for 1981–2010 are used in the hindcast experiments, and the model outputs for 2011–2018 are used in the real forecast experiment in this study. In the present work, forecasts initialized on 1 December, that is, DJF forecasts with a 0-month lead, are analyzed for comparison with the statistical models that use the data available before 1 December. The winter (DJF) of 1979, for example, refers to the average of the 3 months from December 1979 to February 1980.

3. ML Model

3.1. Model Construction

In seasonal forecasting, the number of samples is usually less than 100 (Kämäräinen et al., 2019). It is hard to retain both spatial and temporal information when using small data sets. Thus, EOF analysis is performed first to separate the spatial information and temporal information. By retaining a few leading modes of the EOFs, the model space is reduced, and some small-scale unpredictable features are filtered (Chen & Yuan, 2004). Forecast models built based on EOFs tend to be insensitive to the EOF modes retained (e.g., Wu et al., 2013).

The present study attempts to build ML models that use climate predictors from the previous autumn to make seasonal forecasts for winter SAT over NA. More specifically, assume that the winter SAT in the observations forms a matrix $V$. $V$ can be decomposed into EOFs (spatial patterns) and corresponding principal components (PCs) (time series):

$$V = EOF \ast PC^T$$  \hspace{1cm} (1)

where the columns of EOF are orthogonal to each other and the columns of PC are also orthogonal. The superscript $T$ represents matrix transposition. We use a covariance matrix of the SAT to perform EOF, and more information on EOF can be found in Björnsson and Venegas (1997).
Truncating Equation 1 to the first a few modes can reduce the model space (Chen & Yuan, 2004; Wu et al., 2013). In the current work, only the leading three EOFs are used in the forecast equation, as the rest of the EOFs cannot be separated from each other under the criteria of North et al. (1982). The leading three EOFs together account for 64 percentage of SAT variance (Figure 1) and represent large-scale features that are likely related to boundary forcing and thus potentially predictable on the seasonal time scale. Specifically, during operation of the forecast for winter SAT, for each grid point in space, a multiple LR model is constructed as follows:

$$SAT(x, y, t) = PC1(t) \ast EOF1(x, y) + PC2(t) \ast EOF2(x, y) + PC3(t) \ast EOF3(x, y) + \varepsilon$$

Figure 1. The spatial distributions of (a) EOF1, (c) EOF2, and (e) EOF3 of the winter SAT over NA for the period 1979–2018 obtained using regression and the corresponding principal component (PC) for (b) EOF1 (PC1), (d) EOF2 (PC2), and (f) EOF3 (PC3). The number in the top right corners of (a), (c), and (e) is the percentage of the variance explained by the three EOF modes. The dotted regions in (a), (c), and (e) represent the 95% confidence level.
\[ P_{i,j} = F(I_{i,j}) \]  

where the subscript \( i \) denotes the \( i \)th PC, \( j \) denotes the \( j \)th year, and \( I \) is the matrix of the \( i \)th PC-related climate predictors from the previous autumn.

Among the ML approaches, some deep learning techniques require vast amounts of training data to learn the nonlinear relationships between predictors and predictands. These approaches are suitable for weather forecasting, for which the data amount is usually huge. However, in seasonal forecasting, the number of samples is usually less than 100 (Kämäräinen et al., 2019). In our study, the total number of samples is 40 (1979–2018). This limitation disables the application of many deep learning models (e.g., Kämäräinen et al., 2019). Thus, only those ML models requiring small training data sets are preferred in our seasonal forecasting experiments. In this study, we use two ML methods: support vector regression (SVR) and extreme gradient boosting (XGBoost) as \( F \) due to their good performance in time series prediction, as revealed by previous studies (e.g., Hong, 2011; Pan, 2018; Pavlyshenko, 2016; Sambasivan & Das, 2017). Other ML models have also been examined, for example, the Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regression models, both of which are extensions of the traditional LR model. However, the results do not outperform those of the traditional LR model and are not presented in this study.

As the routine for obtaining a conventional LR, SVR, or XGBoost model is essentially solving an optimization problem, we summarize the basic theories of LR, SVR, and XGBoost under the framework of the optimization problem in the following subsection.

### 3.2. LR

Assume that there is a training data set \( \{(X,Y), X \in \mathbb{R}^N \times M, Y \in \mathbb{R}^N \times 1\} \), where \( X \) is the input, \( Y \) is the target, \( \mathbb{R}^N \times M \) represents the input space and \( \mathbb{R}^N \times 1 \) represents the target space. Each pair of data in the training data set is expressed as \( \{(x,y), x \in \mathbb{R}^1 \times M, y \in \mathbb{R}^1\} \). Here, the superscript \( N \) denotes the number of time points, and \( M \) denotes the number of climate predictors.

LR attempts to model the relationship between several features and a label variable by fitting a linear model to the data set. LR follows the principle of minimizing the prediction error on the training data set. LR has been widely used in seasonal climate predictions in previous studies (e.g., Lin & Wu, 2011; Wang, Ting, & Kushner, 2017; Wu et al., 2009). The optimization problem of LR can be expressed as follows:

\[ \hat{y} = x_\omega + b \]  

\[ \text{minimize } X_\omega - Y^2 \]  

where \( \omega \) and \( b \) are the coefficients determined from the training data set and \( \hat{y} \) is the model output.

### 3.3. SVR

SVR is essentially an extension of the support vector machine (SVM) technique. Different from LR, SVR follows the principle of structural risk minimization, trying to minimize the upper and lower bounds of a generalized error rather than minimizing the prediction error for the training data set. SVR attempts to generate a hypersurface that has a maximum margin to separate a high-dimensional feature space using a support vector kernel with \( \varepsilon \)-insensitive loss. The SVR optimization problem can be expressed as follows: (Smola & Schölkopf, 2004):

\[ \hat{y} = \phi(x)_\omega + b \]  

\[ \text{minimize } \frac{1}{2} \omega^2 + C \sum_{i=1}^{N} \left( \xi_i + \xi_i^* \right) \]  

subject to

\[ y_i - \phi(x_i)_\omega - b \leq \varepsilon + \xi_i \]

\[ \phi(x)_\omega + b - y_i \leq \varepsilon + \xi_i^* \]

\[ \xi_i, \xi_i^* \geq 0 \]  

where the subscript \( i \) denotes the \( i \)th pair of training data; \( \phi(x) \) denotes the high-dimensional feature space,
which is nonlinearly mapped from the input space; \( \xi_i \) and \( \xi_i^* \) denote the slack variable that measures the error of the upper and lower sides of the hypersurface; \( \omega \) and \( b \) are coefficients; and \( C \) is a penalty parameter. In this study, the radial basis function (RBF) is used as the kernel.

### 3.4. XGBoost

Different from LR and SVR, XGBoost is a tree ensemble model using \( K \) additive functions to predict the output. The XGBoost optimization problem can be expressed mathematically as follows (Chen & Guestrin, 2016):

\[
\hat{y} = \psi(x) = \sum_{k=1}^{K} f_k(x)
\]

\[
f(x) = \omega_{q(x)}
\]

\[
\text{minimize } \sum_{t} l(\hat{y}_t, y_t) + \sum_{k} \Omega(f_k)
\]

\[
\Omega(f) = YT + \frac{1}{2} \lambda \omega^2
\]

\[
\omega : \mathbb{R}^{1 \times M} \rightarrow T, \omega \in \mathbb{R}^{T}
\]

where each \( f_k \) is an independent regression tree with the structure \( q \) and leaf weight \( \omega \); \( T \) is the number of leaves in the tree; \( l \) is a differentiable convex loss function that measures the difference between the model output \( \hat{y} \) and the target \( y_t \), with the subscript \( t \) denoting the \( t \)th leaf; \( \Omega \) penalizes the complexity of the model; and \( Y \) and \( \lambda \) are coefficients.

### 4. Major Modes of Winter SAT Over NA and Related Climate Predictors

To obtain potentially useful predictors for the winter SAT over NA, EOF analysis is first applied to the observed winter SAT data over NA for the period 1979–2018. The first three EOF modes and their corresponding PCs are presented in Figure 1. The leading three EOF modes account for 32.87%, 21.71%, and 8.88% of the total variance, respectively, which are well separated from other modes based on the criteria of North et al. (1982). Since the fourth EOF mode of SAT accounts for a small fraction of the total variance (less than 6%) and cannot be separated from the other high-order EOF modes, we only retain the first three EOF modes. The positive phase of the first EOF (EOF1) (Figure 1a) has significant negative anomalies over large areas of northern NA (Figure 1a). The time series associated with EOF1 (PC1) is dominated by positive values before 1996 and by negative values during the period 1997–2018 (Figure 1b), implying that PC1 experienced an interdecadal change in the mid-1990s. The positive phase of the second EOF (EOF2) is characterized by a northeast-southeast dipole pattern with significant negative SAT anomalies extending from northeastern Canada to the southeastern United States (U.S.) and positive anomalies near the Labrador Sea (Figure 1c). The positive phase of the third EOF (EOF3) shows an east-west SAT dipole structure, with positive anomalies occupying western NA and negative anomalies dominating eastern NA (Figure 1e). Different from PC1, PC2, and PC3 feature alternating positive and negative values (Figures 1d and 1c), implying that PC2 and PC3 are dominated by interannual variations. These dominant SAT EOFs in North America are in general consistent with those obtained in Lin and Wu (2011) using a different data set covering a different period.

As we mentioned in the Introduction, the potential seasonal predictability of the atmosphere is mostly based on interactions with slow-varying components, for example, the ocean and land (e.g., Charney & Shukla, 1981), or low-frequency large-scale circulation patterns. In the present study, potential predictors, including SST, snow cover extent (SCE), and some large-scale teleconnection patterns from the previous autumn, are selected according to the correlation coefficients between these potential predictors and the three PCs. The SST and SCE indices are constructed simply by area-weighted averages over highly correlated regions. Before calculating the correlation coefficient, all the time series are normalized. Different combinations of predictors are chosen to predict the three PCs.
Figure 2 shows the correlation coefficient maps between SCE over the TP and the three PCs. The SCEs over the central-eastern TP are highly correlated with PC1, while the SCEs over the northwestern and northeastern TP are highly correlated with PC2 and PC3, respectively. Three TP snow indices are then constructed over Regions A–C for PC1, PC2, and PC3 by applying the area-weighted average of the SCE. Figure 3 presents the correlation coefficient maps of SCE over the Eurasian continent (EU) with the PCs. According to Figure 3, the EU snow cover index associated with PC1 is calculated as \((D + E)/2\), where \(D\) and \(E\) represent the area-weighted average of SCE in the respective region, that associated with PC2 is calculated as \((F - G)/2\), and that associated with PC3 is calculated as \((I - H)/2\).

Figure 4 shows the correlation coefficient maps between SST and the three PCs. The Indian Ocean index and Pacific Ocean Index for PC1 are calculated by averaging the SST over Regions J and K, respectively. The Atlantic Ocean index for PC1 is calculated as \((L + M)/2\). For PC2, the Indian Ocean index is calculated over region O. The Pacific Ocean index is calculated as \((P - Q)/2\), and the Atlantic Ocean index is calculated over Region R. The SST over the Indian Ocean is not significantly correlated with PC3; therefore, the Pacific Ocean Index for PC3 is formulated as \((U + V - S - T)/4\), and the Atlantic Ocean index is formulated as \((X - W - Y)/3\).

In addition to the above constructed SCE and SST indices, according to previous work, certain teleconnection patterns, including the AO, NAO, and PNA, as well as various ENSO indices from the previous autumn, may also be correlated with these PCs. The correlation coefficients between the abovementioned indices and the three PCs are summarized in Table 1. According to Table 1, the indices with coefficient coefficients passing the 95% confidence level (bold font in Table 1) are chosen as predictors in the ML models. Specifically, the TP index, EU index, Indian Ocean index, Pacific Ocean index, Atlantic Ocean index and NAO, AO, and Niño1 + 2 index are used to construct the ML model to predict PC1; the TP index, EU index, Pacific Ocean index, Atlantic Ocean index, AO, NAO index, and Niño3.4 index are used to predict PC2; and the TP index, EU index, Pacific Ocean index, Atlantic Ocean index, Niño3.4 index, and Niño4 index are used to predict PC3.

5. Results

5.1. Hindcast Experiments

In this section, hindcast experiments are first performed for the period from 1979 to 2018 to test the performance of the constructed ML models. To avoid artificial skill in the hindcast, a cross-validation framework is applied, in which the data used to verify the model hindcasts are not used during the model training period (Barnston & Ropelewski, 1992). In the hindcast experiment, the take-1-year-out scheme is utilized. In the take-1-year-out cross-validation scenario, 1 year of data is removed, and the LR, SVR and XGBoost models are trained using the remaining years, and the removed year is used to verify the forecast results. The 1-year window is moved forward year by year until the end of the whole time series. To take into account the possible interannual autocorrelation of the time series, the results of the take-5-years-out scheme are also examined (supporting information Figure S1). Similar to the take-1-year-out scheme, in the take-5-years-out scheme, 35-year data sets are used to construct the...
empirical forecast model, and then the model is used to produce the forecast for the removed 5 years. The performances of the take-1-year-out and take-5-years-out schemes bear many similarities to each other, implying that the ML models are robust. In the following, the hindcasts from the two ML models for the period 1981–2010 are compared with the LR model and with the results from the two dynamic models.

Figure 5 shows the cross-validated correlation coefficients between the three empirical model hindcasts and the observations (Figures 5a–5c). The correlation coefficients between the CanCM4i and GEM_NEMO hindcasts and the observations are presented in Figures 5d and 5e. The average skill over the entire domain is 0.31 (LR), 0.40 (SVR), 0.34 (XGBoost), 0.40 (CanCM4i), and 0.43 (GEM_NEMO), respectively. In general, the hindcast correlation skills of the LR, SVR, and XGBoost models are comparable to those of the two dynamic models. The three empirical models have considerable correlation skills over a large area in northwestern and central NA. However, all three empirical models have poor hindcast skills over southwestern NA.

Among the three empirical models, the hindcasts from the SVR model have the highest correlation skill (Figure 5b). The correlation skills of the hindcast from CanCM4i and GEM_NEMO differ somewhat in terms of spatial distribution from those of the three empirical models. CanCM4i and GEM_NEMO have relatively poor correlation skills in central and western NA, while they perform well over southern and northern NA (Figures 5d and 5e).

Figure 6 presents the correlation skills of the ensemble mean of the three empirical models (Figure 6a) and that of the two dynamic models (Figure 6b). Upon comparing Figure 6a to Figure 6b, the ensemble-mean

---

**Figure 3.** Correlation coefficient maps between the SON snow cover extent over the EU and (a) PC1, (b) PC2, and (c) PC3 for the period 1979–2018. Dotted areas indicate correlation coefficients significant at the 95% confidence level. The boxes denote regions used to construct snow cover indices. The blue and red alphabets (D–I) are labels used for regions.
hindcast of LR, SVR, and XGBoost clearly shows a higher correlation skill over central NA but a lower skill over southeastern NA. The ensemble-mean hindcast of the two dynamic models shows a lower correlation skill over central NA but a higher correlation skill over southern NA (Figure 6b). The above results suggest that the empirical models and the dynamic models have advantages in forecasting winter SAT over different areas of NA. The correlation skill of the multimodel ensemble means of all five models' hindcasts is presented in Figure 6c. In this figure, the five-model ensemble mean displays considerable correlation skill over most areas of NA and is clearly better than the individual dynamic climate models.

To better illustrate the improvement in the hindcast correlation skills of the empirical models relative to the dynamic models, Figure 7a shows the differences in the correlation skills between the ensemble-mean hindcast from the three empirical models and that from the two dynamic models. Again, it is evident that the three empirical model ensemble mean shows a better skill over large areas of central NA than does the two dynamic model ensemble mean but a lower skill over southern and northeastern NA. Figure 7b presents the differences in the correlation skill between the five-model ensemble-mean hindcast and the two dynamic model ensemble-mean

**Figure 4.** Correlation coefficient map between the SON SST and (a) PC1, (b) PC2, and (c) PC3 for the period 1979–2018. Dotted areas indicate correlation coefficients significant at the 95% confidence level. The boxes denote regions used to construct snow cover indices. The blue and red alphabets (J–Y) are labels used for regions.

| Table 1  | Correlation Coefficient Between the Autumn Climate Indices and PCs |
|---------|---------------------------------------------------------------|
|         | PC1    | PC2    | PC3    |
| TP      | –0.50  | 0.33   | 0.34   |
| EU      | –0.46  | –0.59  | –0.49  |
| Indian Ocean SST | –0.42 | –0.29 |    |
| Pacific Ocean SST | –0.46 | 0.35 | 0.57 |
| Atlantic Ocean SST | –0.43 | 0.36 | 0.58 |
| AO      | 0.14   | –0.43  | 0.06   |
| NAO     | 0.17   | –0.33  | 0.22   |
| FNA     | –0.33  | 0.00   | –0.11  |
| Niño3.4 | –0.10  | –0.30  | 0.35   |
| Niño4   | –0.13  | –0.22  | 0.44   |
| Niño3   | –0.08  | –0.31  | 0.26   |
| Niño1 + 2 | 0.04  | –0.35  | 0.11   |

**Note.** The correlation coefficients pass 95% are in bold font and used as predictors for PCs.
The five-model ensemble-mean hindcast shows a higher skill over most regions of NA, especially over the central NA, than the two dynamic model ensemble-mean hindcast. The above results indicate that the ensemble mean of the empirical and dynamic models can indeed improve the correlation skills of winter SAT over NA.

To compare the performance of each model, the differences in the correlation skill between individual empirical models and the dynamic models are further presented in Figure 8. This figure shows that both the ML models and the LR model have better skills than do the dynamic models, with the highest improvement appearing over the central NA. Additionally, the SVR model exhibits a larger improvement in the

Figure 5. Cross-validated correlation coefficients between the winter SAT of the (a) LR, (b) SVR, (c) XGBoost, (d) CanCM4i, and (e) GEM_NEMO hindcasts and the observed DJF SAT over NA for the period of 1981–2010.
To better understand the sources of the correlation skill for the two ML models and the LR model, three sets of winter SAT fields are constructed using Equation 2 while using the three EOFs separately. The correlation coefficients between each set of constructed SAT fields and the observed winter SAT data for the two ML models and for the LR model are presented in Figure 9. For the SAT constructed using EOF1, all three models show considerable correlation skills over northeastern Canada and the western coast of the US. In contrast, the LR, SVR, and XGBoost models show high skills over central NA for the SAT constructed using EOF2, with a pattern resembling that in Figure 8. For the SAT fields constructed using EOF3, all three models perform well along the western and eastern coast of NA. The above results suggest that the improvement in the correlation skill of the three empirical models compared to the two dynamic models mainly originates from the contributions of a skillful forecast of EOF2. As is seen from Table 1, EOF2 is likely associated with other factors, for example, the AO/NAO, Nino1 + 2, and EU. This implies that to improve dynamical seasonal SAT forecasts in NA, it is important to capture the impact of these factors. Significant correlation skills can be noticed over southwest United States for the hindcast using EOF1 (Figures 9a, 9d, and 9g). Further examination of the root-mean-square error (RMSE) shows that the CanCM4i and GEM_NEMO hindcasts have large values over these areas (not shown). The differences of the RMSE skill between the EOF1 × PC1 hindcast using the three empirical models and CanCM4i and GEM_NEMO suggest that the three empirical models perform better than CanCM4i and GEM_NEMO (supporting information Figure S6). The improved skills in the southwest United States indicates that a longer data set may also help capture the signal in the empirical models.

Another issue that we want to examine is the advantages of the two ML models over the traditionally used LR model. The differences in correlation skill among SVR, XGBoost and LR are presented in Figure 10. The SVR model outperforms LR over most parts of NA, except for a small area in southwestern NA (Figure 10a). In comparison, XGBoost shows higher correlation skill over the northern and central regions but has lower skill over some areas of the southwest than the LR model (Figure 10b). Overall, the two ML models have better correlation skills than the LR model for most NA regions, and the SVR model performs the best among the three empirical models.

In summary, compared to the two dynamic models, the two ML models and the LR model exhibit higher correlation skills for hindcasts of the winter SAT over central NA for the period of 1981–2010. The three empirical models, however, have low correlation skills over southwestern NA. A five-model ensemble-mean hindcast outperforms the two dynamic model ensemble-mean hindcast, suggesting that the ML models can be used as supplementary tools to improve the dynamic model seasonal prediction. The improvements in the correlation skills of the two ML models and the LR model mainly
contribute to a skillful forecast of the second EOF of the winter SAT. A comparison of the hindcasts of the three empirical models shows that the SVR model performs the best among the three models.

5.2. Real Forecast Experiments

In this section, an experiment mimicking a real-time forecast is conducted. The process of constructing the empirical forecast model is similar to that of the hindcast forecasts; however, only the data from the period of 1979–2010 is used. Specifically, the EOF analysis is performed on the winter SAT in North America for the period of 1979–2010, and the key regions of snow and the SST that can be related to the variation of the winter SAT in North America are determined by correlations. Snow and the SST indices are constructed by averaging the variables over the key regions and serve as predictors. The models are then used to forecast the winter SAT over NA for the years from 2011 to 2017. The results of the EOF analysis of the winter SAT in North America are presented in supporting information Figure S2. The correlation maps of the leading three PCs for the snow and SST fields are presented in Figures S3–S5. A comparison to Figures 1–4 in the manuscript shows many similarities between the results obtained using the whole period and the period of 1979–2010. However, differences can be noted for the key regions of the snow cover over the TP (Figure S5). During the period of 1979–2010, only PC1 can be significantly correlated with the snow over the TP, indicating that there exist decadal changes of the relationship between the TP snow and the climate over North America, which is consistent with the results of our previous work (Qian et al., 2019). Therefore, in the real forecast experiment, only one TP snow index is used as a predictor for the winter SAT in North America instead of three. Other predictors used for the prediction of PCs can be found in Table S1.

The SAT forecasts from CanCM4i and GEM_NEMO are compared with those from the three empirical models. It can be seen in Figure 11 that all three empirical models show considerable correlation skills over central and western NA (Figures 11a–11c). In this real forecast experiment, both the SVR and XGBoost models outperform the LR model. The SAT forecast of the SVR model shows a higher correlation skill over the central NA than that of the LR model, while the SAT forecast of the XGBoost model shows a higher skill over the southeastern NA than that of the LR model (the skill differences are calculated but not shown).

The advantage of the five-model ensemble mean in forecasting winter SAT in NA over the two dynamic model ensemble mean was then validated. The correlation skills of the five-model ensemble-mean forecast and the two dynamic model ensemble-mean forecast are presented in Figures 12a and 12b, respectively. Their differences are presented in Figure 12c. In the real forecast experiments, the five-model ensemble mean shows a higher skill than the two dynamic model ensemble over a large portion of NA, especially central NA and the western United States (Figure 12c). In general, both the forecast and hindcast experiments show that the ensemble mean of the five models yields a higher skill for winter SAT prediction in NA than the dynamic models alone. Since all the predictors can be easily monitored, the ML models can be utilized in practical forecasting. This implies that ML models have great potential in operational forecasting as supplementary tools to improve the seasonal forecast skill of the NA climate.

5.3. Sensitivity Experiments

To further investigate the performance of the ML models, two additional sensitivity experiments are conducted. The first sensitivity experiment aims to test the sensitivity of the ML model skills to the size of the training data set.
In this sensitivity experiment, in addition to the take-1-year-out hindcast shown in Figures 5a–5c, another two sets of cross-validation schemes with take-5-years-out and take-10-years-out windows are applied. Specifically, for the take-5-years-out and take-10-years-out cross-validation schemes, 35- and 30-year data sets are used to train the empirical models, respectively. The performance of the constructed empirical models is then evaluated using the removed 5 or 10 years.

Figure 13 presents the correlation skills of the winter SAT obtained using the take-1-year-out (Figure 11a), take-5-years-out (Figure 11b) and take-10-years-out schemes (Figure 11c). For the SVR model, the correlation skills of the take-1-year-out (Figure 11d) and the take-5-years-out (Figure 11e) schemes are comparable; however,
the correlation skill of the take-10-years-out scheme is significantly lower (Figure 11f). Different from the LR and SVR models, the XGBoost model exhibits a progressive decrease in correlation skill with the reduction in the training data set size (Figures 11g-11i). The results of the sensitivity experiment indicate that compared to the SVR and XGBoost models, the LR model is not very sensitive to the size of the training data set during the period under examination. For the two ML models, increasing the size of the training data set might efficiently improve their forecast skills.

As revealed by previous studies (Lin & Wu, 2011, 2012; Qian et al., 2019), autumn TP snow cover can affect the winter SAT over NA via the propagation of a snow-related Rossby wave train. These previous studies suggest that TP snow cover can be an important climate predictor for NA winter SAT. The second sensitivity experiment attempts to investigate the effect of the TP snow predictor on the seasonal forecast skill of the SAT in NA. In this sensitivity experiment, the LR, SVR, and XGBoost models are trained with and without the TP snow predictor, respectively. Their forecast results are then compared with each other under the framework of take-1-year-out cross validation.

Figure 14 shows the results of the correlation skills of the LR, SVR, and XGBoost with (Figures 14a, 14d, and 14g) and without (Figures 14b, 14e, and 14h) the TP snow predictor. The differences in the correlation skills are presented in Figures 14c, 14f, and 14i. Figure 14c shows that the additional use of the TP snow predictor
can indeed improve the forecast skill of the LR model, mainly over the western US and northwestern Canada. For the SVR model, the TP snow predictor can improve the forecast skill over northwestern Canada and large areas of central NA (Figure 14f). The TP snow predictor can also help improve the forecast skill of the XGBoost model over northern and southern NA (Figure 14i). In summary, the additional use of the TP snow predictor can indeed contribute to improving the seasonal forecast skill of the winter SAT over NA. The spatial distribution of the improved areas, however, is mode dependent.

6. Summary

In this study, two ML models are established to predict the winter SAT over NA using various autumn climate predictors. To investigate the performance of the ML models, both hindcast and real forecast experiments are conducted. The forecast skills of one LR model and two Canadian dynamic models are used for comparison.

The hindcast experiment is first performed under the framework of a take-1-year-out cross-validation scheme. Both the ML models and the LR model show reasonable forecast skill for the winter SAT in NA. The three empirical models perform well over most parts of NA, especially over northern and central NA. Limited forecast skill, however, is found over southwestern NA. In contrast, the two dynamic models have better forecast skill over northern and southern NA than over central NA. Thus, the empirical models can be used as supplementary forecast tools for the dynamic climate models to improve the forecast skill of winter SAT in NA, especially over central NA. Further analysis of the sources of the forecast skill of the empirical models suggests that the improvement in the skill over central NA is derived mainly from a skillful forecast of the second EOF mode of winter SAT over NA. To examine the real seasonal forecast skills of the two ML
models, the models are trained using the data from 1979 to 2010. Then, the ML models are used to forecast the winter SAT for the period 2011–2017. The results show that the two ML models outperform the LR model. Compared to the two dynamic climate models, the SVR and XGBoost models exhibit better seasonal forecasting skills for the winter SAT over northern and central-western NA.

A comparison of the two ML models and the LR model suggests that the two ML models exhibit better forecasting performance than the LR model. In general, the SVR model shows the highest skill among the three empirical models. Since all the predictors can be easily monitored, the ML models can be utilized in
practical forecasts, indicating that they could represent real-time supplementary forecast tools for improving the forecast skill and may operationally facilitate the seasonal forecasting of the winter climate in NA. The sensitivity of the three empirical models to the training data set size is investigated by comparing the take-1-year-out, take-5-years-out and take-10-years-out cross-validation scheme experiments. The results show that the LR models have less dependence on the training data set size than the two ML models. Increasing the size of the training data set might improve the forecast skills of the two ML models.

Figure 12. Correlation coefficients between the (a) LR, SVR, XGBoost, CanCM4i, and GEM_NEMO ensemble-mean forecast and (b) CanCM4i and GEM_NEMO ensemble-mean forecast and the observed DJF SAT over NA for the period of 2011–2017. (c) The difference in correlations between (a) and (b).
forecast experiments, it is noted that the SVR model performs better compared to the XGBoost model in forecasting the winter SAT in North America during the period under examination. One possible reason accounting for the difference might be that the XGBoost model is an ensemble ML model that improves the forecast performance of each additive function through the boosting method (Chen & Guestrin, 2016) and, therefore, is more sensitive to the training size, as shown in Figure 13. A longer training data set might be helpful for improving the forecasting skill of the XGBoost model. As an extension of previous work (Lin & Wu, 2011; Qian et al., 2019), the sensitivity of empirical model skill to the TP snow predictor is also investigated. It is shown that the additional use of the TP snow predictor can improve the forecast skill of the three empirical models but over different regions of NA.

In the current work, two ML models, that is, SVR and XGBoost, are used to perform the forecast experiments. Other ML methods are also examined and compared to the results presented in the study (not shown). For example, the LASSO and Ridge regression methods are both extensions of the traditional LR model and have also been investigated. However, the forecasting skills of these models do not perform better than those of the LR model in forecasting the winter SAT in North America. Other more complex ML models like CNN and Recurrent Neural Network (RNN) require large training data sets. In the current work, we only include 40 samples, and thus, these models are not suitable for use in the current work.

Figure 13. Cross-validated correlation between the (a, d, and g) take-1-year-out, (b, e, and h) take-5-years-out and (c, f, and i) take-10-years-out for (a–c) LR, (d–f) SVR, and (g–i) XGBoost hindcast and the observed DJF SAT over NA for the period of 1981–2010. The number in the left lower corner of each panel is the domain average skill.
Additionally, in this study, we show that the SVR and XGBoost models perform better in forecasting the winter SAT over northern and central NA than do the two dynamic models. It will also be interesting to examine whether this method is also useful for other seasons—for example, summer—or for longer-lead time predictions. Further forecast experiments will be performed to examine the above issues in future works.

In the current work, we focus on examining the potential predictors related to the low boundary conditions, that is, the SST and the land snow. Other factors, for example, the stratosphere and other land process, might also be useful and might help improve the forecasting skill. However, using all potential predictors is beyond the scope of the current work and may require further study in the future.

**Data Availability Statement**

The reanalysis data are available online (https://www.esrl.noaa.gov/psd/data/gridded/). The original weekly Climate Data Record of Northern Hemisphere Snow Cover Extent is available online (https://climate.rutgers.edu/snowcover/index.php). The Climate Research Unit land surface temperature data can be downloaded online (http://www.cru.uea.ac.uk/data). The Hadley Center Sea Surface Temperature data are from https://www.metoffice.gov.uk/hadobs/hadisst website. All climate indices are obtained from https://www.esrl.noaa.gov/psd/data/climateindices/list website. The CanCM4i and GEM_NEMO hindcast
Acknowledgments

This research is funded by the National Natural Science Foundation of China (Grants 41722503 and 41530425).

References

Bader, J., & Latif, M. (2005). North Atlantic Oscillation response to anomalous Indian Ocean SST in a coupled GCM. Journal of Climate, 18(24), 5382–5389.

Badr, H. S., Zaitchik, B. F., & Guikema, S. D. (2014). Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel. Journal of Applied Meteorology and Climatology, 53(3), 614–636.

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Monthly Weather Review, 115(6), 1083–1126.

Barnston, A. G., & Ropelewski, C. F. (1992). Prediction of ENSO episodes using canonical correlation analysis. Journal of Climate, 5(11), 1316–1345.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? Bulletin of the American Meteorological Society, 93(5), 631–651.

Björnsson, H., & Venegas, S. A. (1997). A manual for EOF and SVD analyses of climatic data. CCGCR Report, 97(11), 112–134.

Charney, J. G., & Shukla, J. (1981). In J. Lighthill, & R. P. Pearce (Eds.), Predictability of monsoons. Monsoon dynamics. England: Cambridge University Press.

Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12, e2019MS001958. https://doi.org/10.1029/2019MS001958

Chen, D., & Yuan, X. (2004). A Markov model for seasonal forecast of Antarctic sea ice. Journal of Climate, 17(16), 3156–3168.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining (pp. 785–794). ACM.

Chi, J., & Kim, H. C. (2017). Prediction of arctic sea ice concentration using a fully data driven deep neural network. Remote Sensing, 9(12), 1305.

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E. (2019). S2S reboot: An argument for greater inclusion of machine learning in seasonal to subseasonal forecasts. Wiley Interdisciplinary Reviews: Climate Change, 10(2), e06567.

Dermise, J., Brunet, G., Plante, A., Gagnon, N., Boer, G. J., Zwiers, F. W., et al. (2001). Seasonal predictions based on two dynamical models. Atmosphere–Ocean, 39(4), 485–501. https://doi.org/10.1080/07055900.2001.9649690

Gong, G., Entekhabi, D., & Cohen, J. (2002). A large-ensemble model study of the wintertime AO–NAO and the role of interannual snow perturbations. Journal of Climate, 15(21), 3488–3499.

Gong, G., Entekhabi, D., Cohen, J., & Robinson, D. (2004). Sensitivity of atmospheric response to modeled snow anomaly characteristics. Journal of Geophysical Research, 109, D06107. https://doi.org/10.1029/2003JD004160

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multiyear ENSO forecasts. Nature, 578(7775), 568–572. https://doi.org/10.1038/s41586-019-1599-7

Harris, J., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. International Journal of Climatology, 34(3), 623–642. https://doi.org/10.1002/joc.37111

Herman, G. R., & Schumacher, R. S. (2018). Money doesn’t grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. Monthly Weather Review, 146(5), 1571–1600.

Hoering, M. P., Hurrell, J. W., Xu, T., Bates, G. T., & Phillips, A. S. (2004). Twentieth century North Atlantic climate change. Part II: Understanding the effect of Indian Ocean warming. Climate Dynamics, 23(3–4), 391–405.

Hong, W. C. (2011). Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. Neurocomputing, 74(12–13), 2096–2107.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. Science, 269(5224), 676–679. https://doi.org/10.1126/science.269.5224.676

Hurrell, J. W. (1996). Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. Geophysical Research Letters, 23(8), 665–668.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., & Mackey, L. (2019). Improving subseasonal forecasting in the western US with Nooteboom, P. D., Feng, Q. Y., López, C., Hernández-García, E., & Dijkstra, H. A. (2018). Using network theory and machine learning to predict El Niño. arXiv preprint arXiv:1803.10076.
North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review, 110*(7), 699–706. https://doi.org/10.1175/1520-0493(1982)110,0699:SEITEO.2.0.CO;2

Pan, B. (2018). Application of XGBoost algorithm in hourly PM2.5 concentration prediction. In *IOP conference series: Earth and environmental science* (Vol. 113, pp. 12,127–12,133). London: IOP publishing. https://iopscience.iop.org/article/10.1088/1755-1315/113/1/012127/pdf

Pavlýshenko, B. M. (2016). Linear, machine learning and probabilistic approaches for time series analysis. In 2016 *IEEE First International Conference on Data Stream Mining & Processing (DSMP)* (pp. 377–381). Liviv, Ukraine: IEEE. https://arxiv.org/ftp/arxiv/papers/1703/1703.01977.pdf

Qian, Q., Jia, X., & Wu, R. (2019). Changes in the Impact of the Autumn Tibet Plateau Snow Cover on the Winter Temperature over North America in the mid-1990s. *Journal of Geophysical Research: Atmospheres, 124*, 10321–10343. https://doi.org/10.1029/2019JD030245

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research, 108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Reichstein, M., Camps-Varas, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204.

Richman, M. B., Leslie, L. M. (2012). Adaptive machine learning approaches to seasonal prediction of tropical cyclones. *Procedia Computer Science, 12*, 276–281.

Robinson, D. A., Estilow, T. W., & NOAA CDR Program (2012). NOAA Climate Data Record (CDR) of Northern Hemisphere (NH) Snow Cover Extent (SCE), version 1. [indicate subset used]. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5N014G9 [access date].

Rodwell, M. J., Rowell, D. P., & Folland, C. K. (1999). Oceanic forcing of the wintertime North Atlantic oscillation and European climate. *Nature, 398*(6725), 320.

Ropelewski, C. F., & Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review, 114*(12), 2352–2362.

Sambasivan, R., & Das, S. (2017). Big data regression using tree based segmentation. In 2017 14th IEEE India Council International Conference (INDICON) (pp. 1–6). IEEE.

Scala, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters, 41*, 2514–2519. https://doi.org/10.1002/2014GL059637

Seager, R., Kushnir, Y., Nakamura, J., Ting, M., & Naik, N. (2010). Northern Hemisphere winter snow anomalies: ENSO, NAO and the winter of 2009/10. *Geophysical Research Letters, 37*, L14703. https://doi.org/10.1029/2010GL043830

Shabbar, A., & Khandekar, M. (1996). The impact of el niño-Southern oscillation on the temperature field over Canada: Research note. *Atmosphere-Ocean, 34*(2), 401–416.

Smith, G. C., Bélanger, J. M., Roy, F., Pellerin, P., Ritchie, H., Onu, K., et al. (2018). Impact of coupling with an ice-ocean model on global medium-range NWP forecast skill. *Monthly Weather Review, 146*(4), 1157–1180. https://doi.org/10.1175/MWR-D-17-0157.1

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222.

Sobolowski, S., Gong, G., & Ting, M. (2010). Modeled climate state and dynamic responses to anomalous North American snow cover. *Journal of Climate, 23*(3), 785–799.

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review, 109*(4), 784–812.

Wang, L., Ting, M., & Kushner, P. J. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports, 7*(1), 279. https://doi.org/10.1038/s41598-017-00533-y

Wang, Z., Wu, R., & Huang, G. (2017). Low-frequency snow changes over the Tibetan Plateau. *International Journal of Climatology, 38*(2), 949–963. https://doi.org/10.1002/joc.5221

Weyn, J. A., Durrant, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems, 11*, 2680–2693. https://doi.org/10.1029/2019MS001705

Wu, Q., Yan, Y., & Chen, D. (2013). A linear Markov model for East Asian monsoon seasonal forecast. *Journal of Climate, 26*(14), 5183–5195.

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, 114*, D18120. https://doi.org/10.1029/2009JD011733