



Application of penalized linear regression and ensemble methods for drought forecasting in Northeast China

Zeng Li¹ · Taotao Chen¹ · Qi Wu¹ · Guimin Xia¹ · Daocai Chi¹

Received: 15 April 2018 / Accepted: 21 May 2019 / Published online: 14 June 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

Effective drought prediction can be conducive to mitigating some of the effects of drought. Machine learning algorithms are increasingly used for developing drought prediction models due to their high efficiency and accuracy. This study explored the ability of several machine learning models based on penalized linear regression and decision tree (DT)-based ensemble methods to predict drought conditions represented by the Standardized Precipitation–Evapotranspiration Index (SPEI) in Northeast China. We compared the forecasting performance of the penalized linear regression models based on ridge regression (RR) and lasso regression (LR) with the ordinary least squares (OLS) regression model. In addition, the AdaBoost and Random Forests (RF) models were also used to explore the suitability of ensemble methods for improving the forecasting performance. The SPEI was forecast at the different timescales of 3, 6, 12, and 24 months using the aforementioned machine learning models and the indices were used to predict short-term and long-term drought conditions. The prediction results indicated that the penalized linear regression models provided better prediction results and the ensemble methods consistently outperformed the DT model. Overall, the LR models were the optimum models for forecasting the SPEI at different timescales in Northeast China.

Keywords Drought forecasting · Standardized precipitation · Evapotranspiration index (SPEI) · Penalized linear regression · Ensemble methods · Machine learning

Responsible Editor: F. Mesinger.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00703-019-00675-8>) contains supplementary material, which is available to authorized users.

✉ Daocai Chi
daocaichi@syau.edu.cn; daocaichi@vip.sina.com

Zeng Li
lizeng810@hotmail.com

Taotao Chen
taotao-chen@syau.edu.cn

Qi Wu
vichine_21@126.com

Guimin Xia
gmxia@syau.edu.cn

¹ College of Water Resources, Shenyang Agricultural University, Shenyang 110866, China

1 Introduction

Drought is a recurring extreme climate event characterized by below-average precipitation in a given region over a period of months to years (Dai 2011). Drought is one of the most damaging natural disasters, has widespread and detrimental impacts on hydrology, agriculture, and the environment, and causes enormous economic loss (Botai et al. 2016; Chen et al. 2016; Li and Zhou 2015). Moreover, global warming has resulted in increased risk of drought-related stresses for natural and human systems (Touma et al. 2015).

Among natural disasters, drought is considered as the most complex because the inception and end of a drought are difficult to identify. Hence, there is confusion about whether a drought exists because it difficult to define precisely (Wilhite 2000). Furthermore, the influence of drought often accumulates gradually over time and may linger long after the drought is over. In addition, it is difficult and crucial to characterize drought. To quantify the characteristics of drought, such as the intensity, magnitude, duration, severity, and spatial extent, drought indices are regarded as valid

measures. The indices reflect different events and conditions and are easier to use than raw indicator data (Zargar et al. 2011). The researchers have developed more than 150 drought indices so far, which correspond to different types of drought, including meteorology, agriculture, and hydrological drought (Niemeyer 2008). Several of them are the most important and highly popular in global warming scenario, such as the Palmer Drought Severity Index (PDSI; Wayne 1965), the Standardized Precipitation Index (SPI; McKee et al. 1993), the Reconnaissance Drought Index (RDI; Tsakiris and Vangelis 2005), the Standardized Precipitation Evapotranspiration Index (Vicente-Serrano et al. 2010), the Water Surplus Variability Index (WSVI; Gocic and Trajkovic 2014b), etc. The PDSI is a landmark drought index and still widely used, which based on the supply and demand concepts of the water balance equation. It considers not only precipitation but also evapotranspiration and soil moisture, and computes four terms in the water balance equation: evapotranspiration, runoff, soil recharge, and moisture. The SPI is solely based on precipitation data and capable of calculating drought levels for different timescales, and that is put forward by the World Meteorological Organization (WMO) as a universal drought index. The RDI calculates the aggregated deficit between the precipitation and evaporative demand of the atmosphere based on the ratio between two aggregated quantities of precipitation and potential evapotranspiration (PET). Like the SPI, it also can be used for the estimation of drought severity at different timescales. The SPEI is another drought index that considers precipitation and PET. It is based on the monthly (or weekly) difference between precipitation and PET, which represents a simple climate water balance, and then adjusted using a three-parameter log-logistic distribution. Moreover, it also possesses the multiscalar nature similar to the SPI and RDI. The WSVI is similar to the SPEI and following the concept of the RDI, which has good agreement with the SPI, RDI and SPEI for drought monitoring, especially in humid and sub-humid locations (Gocic and Trajkovic 2014a). The drought indices are indispensable tools for explaining the severity of drought events, which extensively used in drought modeling and forecasting.

In recent years, researchers have increasingly begun to utilize data-driven techniques in hydrological phenomena modeling. Karimi et al. (2018b) established models using gene expression programming and support vector machine techniques for forecasting daily streamflow values, and evaluated local (within station) and external (cross-station) data management scenarios. For simulating Leaf Area Index, Karimi et al. (2018a) raised valid alternatives to locally trained models, which used externally trained gene expression programming and random forest models. Additionally, intelligent algorithms are beneficial to improve the performance of traditional hydrological phenomena models

(Azad et al. 2019). With the development of machine learning technology and drought index, a number of studies have employed drought index and other data (e.g., meteorological data and remotely sensed data) to establish drought prediction models, which are based on various machine learning technology, such as multivariate linear regression (Ortegren et al. 2011; Xing et al. 2016), artificial neural network (Ali et al. 2017; Byakatonda et al. 2016), support vector machine (Ganguli and Reddy 2014; Gill et al. 2006), ensemble methods (Belayneh et al. 2016; Rhee and Im 2017), etc. Deo and Şahin (2015) developed Artificial Neural Network models by optimizing hidden neurons, activation functions and different combinations of training and testing algorithms for predicting the monthly SPEI in eastern Australia. Maca and Pech (2016) found that the integrated neural network model performed better than the feed forward multilayer perceptron in predictions of the SPEI and SPI. In predicting the stream flow drought index of the Latian watershed located in Iran, the support vector machine model was superior to the artificial neural network model in terms of a better efficiency (Borji et al. 2016). Among the varied machine learning techniques, the penalized linear regression and ensemble methods are two of the most effective and widely used algorithms for the vast majority of predictive analytics (Caruana et al. 2008; Caruana and Niculescu-Mizil 2006). Drought forecasting models are a type of function approximation problem, which is a subset of supervised learning. Linear regression is pervasive in most data-driven predictive models to solve the function approximation problem. Moreover, ordinary least squares regression (OLS) is the most commonly used linear regression algorithm. However, OLS has problems such as trapping in local optimum as well as high volume of computations and therefore, scientists and engineers have rarely used the OLS algorithm to establish drought prediction models at present. The penalized linear regression represents a relatively recent development in ongoing research to improve on OLS and includes ridge regression (RR) and lasso regression (LR). In ensemble methods, which are currently some of the most effective predictive models, a set of learning algorithms is developed and combined to solve a problem; whereas in conventional learning approaches, a single learning algorithm is used and is based on training data (Zhou 2012). Bootstrap aggregation (Bagging), Boosting and Random Forests (RF) are some of the most popular ensemble methods that can be used to improve the performance of the forecasting model. Zhang et al. (2017) compared seven data mining algorithms and found that the RF and AdaBoost methods resulted in higher accuracies in most cases. Moreover, the RF performed better than the AdaBoost for an unbalanced dataset in a multi-class task. For predicting drought impacts quantified from text-based reports, Bachmair et al. (2017) tested the predictive performance of three data-driven models and found that the

RF model generally performed better than logistic regression and zero-altered negative binomial regression. RF can also be used to develop the short-term drought prediction models that can produce drought prediction maps at high resolution for a very short timescale over East Asia (Park et al. 2018). With good-impact data coverage, RF machine learning approach proved to be a suitable tool for drought monitoring and early warning in Germany and the UK (Bachmair et al. 2016).

However, there are scarcely any drought prediction models based on penalized linear regression in previously published studies. And the suitability of ensemble methods for forecasting SEPI has not yet been systematically assessed. Hence, it will be worth attempting to evaluate the performance of data-driven models which based on penalized linear regression and ensemble methods for forecasting SPEI. Notably, the exploitation of optimum drought prediction models in Northeast China is a new research step. In this study, we explore the ability of several machine learning models based on penalized linear regression and DT-based ensemble methods to predict the SPEI at the different timescales of 3, 6, 12, and 24 months in Northeast China. The objectives of this study were to (1) develop drought prediction models based on two representative algorithms of penalized linear regression (in this case the RR and LR

algorithms) and to compare their forecasting performance with the OLS model; (2) establish drought prediction models using DT-based ensemble methods (in this case the AdaBoost and RF algorithms), and to compare their performance with the DT model; (3) determine the optimum drought prediction model by comparing the forecasting performance of the penalized linear regression and DT-based ensemble methods and to investigate its performance.

2 Materials and methods

2.1 Study region

Northeast China is a vast geographical region with the longitude ranging from 111 to 135°E, and the latitude ranging from 38 to 53°N and includes Heilongjiang Province, Liaoning Province, Jilin Province, and the eastern part of the Inner Mongolia Autonomous Region (Fig. 1). Northeast China, encompasses an area of 1.45 million square kilometers, has complex landforms with the Changbai Mountain to the east, the Lesser Khingan Mountains to the north, and the Great Khingan Mountains to the west. The region is dominated by a typical temperate monsoon climate with four distinct seasons, hot and rainy summers, and cold and dry winters. The

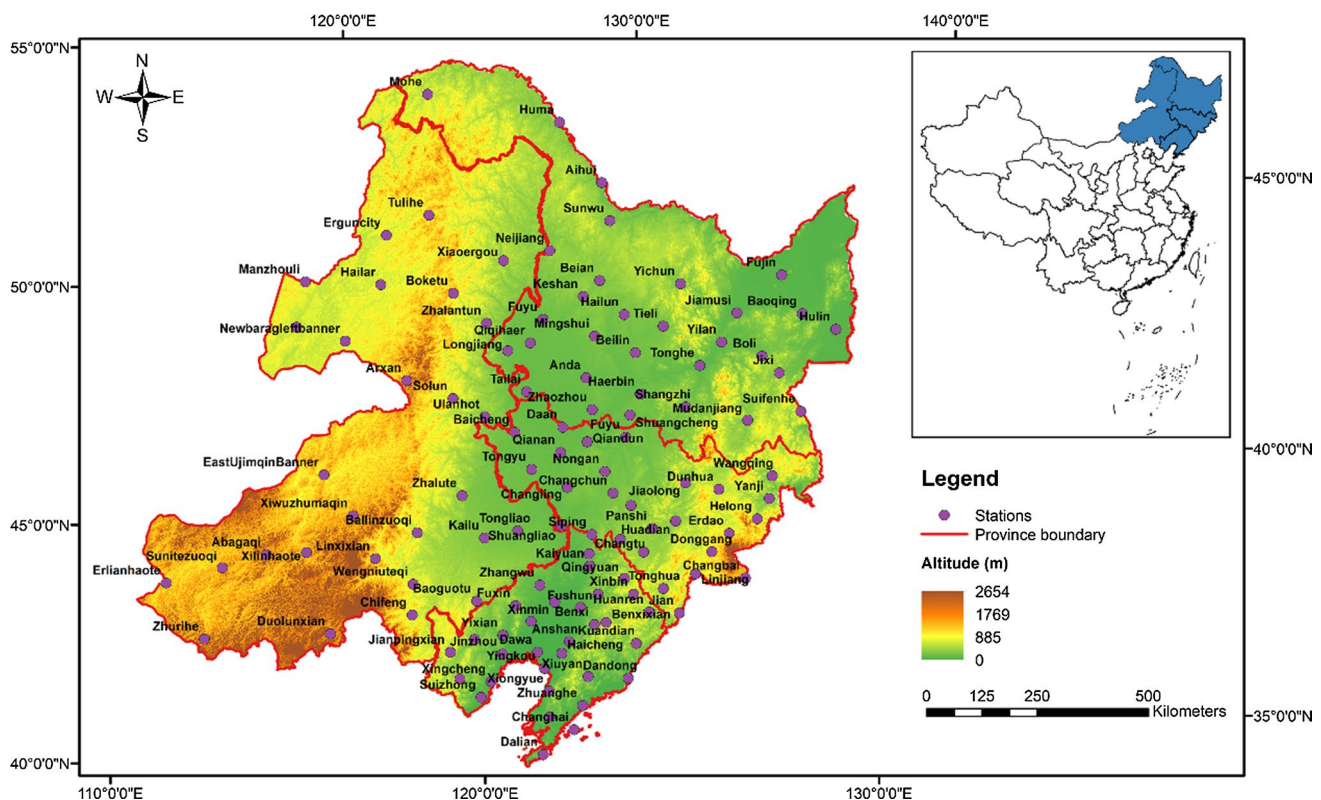


Fig. 1 Spatial distribution of the study locations for meteorological stations in Northeast China

climate zones change from a humid zone to a semiarid zone from the southeast to the northwest and the average annual precipitation is in the range of 300–1000 mm.

Northeast China is a major agricultural region and plays a critical role in maintaining national food security. Additionally, Northeast China has well-developed grassland-based animal husbandry and abundant forest resources. Drought is one of the most damaging and disastrous hazards in Northeast China (Yu et al. 2014), and the risk of drought is increasing (Kong et al. 2015). Many researchers have focused on the analysis of drought characteristics (Wang et al. 2014, 2015) and the impact of drought on agriculture (Peng et al. 2012; Yin et al. 2016) in Northeast China.

2.2 Data

2.2.1 Meteorological data

Meteorological data from 1961 to 2016 for 118 meteorological stations in Northeast China were provided by the China Meteorological Data Service Center (CMDSC; <https://data.cma.cn/>), which included daily observations of minimum, average, and maximum air temperatures, precipitation, relative humidity, sunshine duration, wind speed (at 10-m height), ground surface temperature (at 0-cm height), and atmospheric pressure (Fig. 1). We calculated the monthly accumulative meteorological data by summing the daily meteorological data and checked their qualities according to the Deo and Şahin (2015).

2.2.2 The standardized precipitation evapotranspiration index

Because of the complexity of drought, it is tough to establish a unique and universally accepted drought index for a diverse group of users. However, it is crucial to select a relevant drought index to monitor and forecast drought severity. The PDSI has several deficiencies including the strong influence of calibration period, its limited applicability in locations other than calibrated for US Great Plains' conditions, relatively sophisticated computation, noncomparability between diverse climatological regions, applicability to regions with extreme climate, etc. (Guttman 1998; Zargar et al. 2011). Although, several modified drought indices were developed to address the shortcomings of the PDSI, such as the self-calibrating Palmer Drought Severity Index (SC-PDSI; Wells et al. 2004), etc. In comparison with the other drought indices that can be calculated at different time-scales, its fixed temporal scale remains the main shortcoming of the PDSI. The SPI allows for comparison of drought severity through time and space, but it does not include the effects of temperature variability on drought severity. Under global warming scenarios, the inability of the SPI to capture

an increased evaporative demand is its significant deficiency. Both of the SPEI and the RDI are more sensitive and suitable in cases of a changing environment in that they take into account the effect of PET on drought severity and enable identification of different drought types. However, there are some differences between them. The essential difference is that they adopt different calculation approaches, that the RDI is based on the quotient between precipitation and PET; whereas the SPEI uses the difference between them. Because of using the quotient of precipitation and PET as input to standardization, the RDI gives no valid values when PET is equal to 0. Besides, the RDI shows insensitivity to variations in the magnitude of precipitation and PET by reason of its calculation approach of the drought drivers (Vicente-Serrano et al. 2015). The WSVI is a newly developed drought index, which is compared to the SPI, RDI, and SPEI with good agreement in the case of obtaining the dry and wet periods. However, the performance and limitations of the index should be further verified for the reason that few studies had evaluated drought conditions using the WSVI. In contrast to the aforementioned drought indices, the SPEI does not have distinct shortcomings, which exhibits significant advantages of combining multiscale character with the capacity to integrate potential evaporation and thereby better represent the local water balance. As global warming intensifies, the spread of drought and the loss it causes will increase in many regions (Cook et al. 2014). Effective monitoring and prediction of drought are essential tools to help reduce and mitigate the impacts on hydrology, agriculture, and the environment. The predictive models that are based on the drought index may significantly help decision-makers to achieve efficiency in risk assessments of drought occurrences and the implementation of appropriate drought mitigation strategies.

To apply drought forecasting models based on machine learning technology above, we computed the SPEI by monthly meteorological data following the methodology of (Vicente-Serrano et al. 2010), but we used the Penman–Monteith (PM) method to estimate the PET instead of the Thornthwaite method. The Thornthwaite method with fewer data requirements is the most straightforward approach to calculate PET, but the PM method incorporates the effects of solar radiation, temperature, wind speed, and relative humidity. The method used to calculate the PET is not critical for the calculation of SPEI; Beguería et al. (2014) recommend the more robust PM equation when the data needed for this equation are available. The considered stations are the official sites with complete weather data as required by the PM equation; so we selected the more robust PM method. In this study, the SPEI at the different time-scales of 3, 6, 12, and 24 months was implemented using the freely available SPEI package (version 1.7; <https://cran.r-project.org/web/packages/SPEI/index.html>) in R software.

2.3 Drought forecasting model development

In this study, we explored the different drought states ranging from short-term to long-term. Therefore, the SPEI with 3-, 6-, 12-, and 24-month timescales was used (SPEI3, SPEI6, SPEI12, and SPEI24) for analyses. In forecasting SPEI values at each timescale, a total of 10 input parameters were used to develop the drought prediction models: monthly precipitation, maximum temperature, minimum temperature, average temperature, relative humidity, sunshine duration, wind speed (at 10-m height), ground surface temperature (at 0-cm height), and atmospheric pressure and the synchronous SPEI value. The lag time of the models is one month, i.e., that the SPEI value of next month was the target variable predicted by the above ten input parameters of the current month. For example, to predict SPEI3 on a target month, the models used the meteorological parameters and SPEI3 of the previous month as input. We had retained the available input data of the 54 years (i.e., 1963–2016) for integrity and consistency of the dataset. Moreover, we partitioned the input dataset into two parts: the training dataset and the testing dataset. 74% of the dataset (i.e., 1963–2002) was the training dataset, and the final 26% of the data (i.e., 2003–2016) was the testing dataset. All of the machine learning algorithms in this study are openly accessible. The Python programming language library Scikit-learn (Pedregosa et al. 2011) was used to implement these algorithms, which is the Python package integrating most of the world's advanced machine learning algorithms for supervised and unsupervised problems. We scaled and translated each input feature individually such that it is between zero and one on the training dataset by using the MinMaxScaler function preprocessing feature within the Scikit-learn package.

2.3.1 The ordinary least squares regression models

Linear regression is a straightforward and useful approach for predicting a quantitative response. In this study, the first drought forecasting model only uses the OLS to modeling for predicting the SPEI. The purpose is to see the benefits of the penalized linear regression models to build forecasting models from data.

2.3.2 The penalized linear regression models

It is usually a difficult task to select the variables by the given response for a linear model. Researchers may mistakenly deduce the high-correlated variables because of their high p values, but they are no necessary predictors. Moreover, there would be some other irrelevant variables included in the model and leads to unnecessary complexity and interpretability. If the number of observations is not much larger than the number of variables, then there can

be much variability resulting in overfitting (increased likelihood by adding more parameters but poorer predictions on future observations not used in the model training) (Pereira et al. 2016). The OLS model has the underlying problem which is sometimes overfitting. The penalized linear regression methods can avoid the overfitting problem by shrinkage or regularization, which involves fitting a model with all the predictors. They shrink the estimated coefficients towards zero relative to the classical estimates. The penalized linear regression may improve the overall prediction accuracy by trading off a small increase in bias for a substantial decrease in variance of the predictions. The RR and LR are two of the best-known penalized linear regression.

The RR introduced by Hoerl and Kennard (1970) is very similar to the OLS, except that the coefficients are estimated by minimizing a slightly different quantity. The LR (Tibshirani 1996) is another useful algorithm, which shrinks some coefficients and sets others to 0. The difference between the RR and LR is the measure that each one uses for the vector of linear coefficients. The RR uses squared Euclidean distance but the LR uses the sum of the absolute values that is called taxicab or Manhattan distance. The different coefficient penalty functions cause some important and useful changes in the solutions. To ensure fair comparison and the generalization of each model, we made sure that the RR and LR models were estimated using the same tenfold cross-validation. We set up the default values as the parameters and found that changing these values does not make a noticeable difference in our predictions.

2.3.3 The decision trees models

The DT is a non-parametric supervised learning method used to develop either a classification or a regression model. The DT algorithms build a model in the form of a tree structure that predicts the value of a target variable by learning a set of if–then–else decision rules inferred from the data features. When using the DT, the model splits into branches that indicate the decision's choices. The procedure is repeated recursively until terminal nodes that denote the result of following a combination of decisions are reached. The DT algorithms used most frequently include C5.0, classification and regression trees (CRAT), quick unbiased efficient statistical tree (Quest) and chi-squared automatic interactive detector (CHAID) models. In this study, the DT model was based on the CART algorithm, and we used the default settings. In practice, the trees are usually grown to their maximum size before a pruning step is applied to reduce overfitting (Reiss et al. 2015) and also grouped in ensembles to improve the stability of the process.

2.3.4 The ensemble methods

Ensemble methods are effective learning algorithms that combine multiple learning algorithms to obtain better predictive performance (Dietterich 2000). The principle of ensemble methods is to create a stronger learner by combining multiple weaker learners, and there have been a large variety of ensemble methods in accordance with different weaker learners and combining types. Ensemble methods employ a hierarchy of two algorithms. The low-level algorithm is a base learner, and the upper-level algorithm manipulates the inputs to the base learners so that the models they generate are somewhat independent. There are a lot of different algorithms that can be used as base learners conceivably, but the DT is one of the base learners that gain widespread acceptance. Among various upper-level algorithms, the Bagging, Boosting, and RF are some of the most applied diffusely.

The Bagging generates some training datasets by bootstrap sampling the original training data and then trains a base learner on each of these samples. Finally, the Bagging averages out the resulting models in regression problems (Breiman 1996). The Bagging can perform quite well as long as it is used with relatively unstable learners because the unstable learners ensure the ensemble's diversity despite only minor variations between the bootstrap training datasets (Lantz 2013). Thus, the DT is often used as base learners because of its instability. Strictly speaking, the RF is an extension of the Bagging and generates its sequence of models by training them on subsets of the full training data in the same manner as the Bagging algorithm, where the principal difference with the Bagging is the incorporation of randomized feature selection (Zhou 2012). As a result of this randomness, the bias of a single non-random tree usually slightly better, however, due to averaging, the variance of the RF usually will decrease more than compensating for the increase in bias. Hence, the RF is an overall more efficient predictive model (Breiman 2001).

The Boosting is a general approach for improving the accuracy of weak learners to attain the performance of stronger learners (Freund 1995). Like the bagging, the Boosting also takes a base learning algorithm and invokes it many times with different training sets. Nevertheless, the Boosting does not involve bootstrap sampling and be explicitly constructed to generate complementary learners (James et al. 2013). Adaptive boosting (AdaBoost) that introduced by Freund and Schapire (1997) is one of the most critical Boosting algorithms since it has a solid theoretical foundation, very accurate prediction, great simplicity, and comprehensive and successful applications (Wu et al. 2008). The core principle of the AdaBoost is to fit a sequence of weak learners on repeatedly modified versions of the data. The predictions from all of them are then combined through a

weighted majority vote (or sum) to produce the final prediction (Trevor et al. 2009). We used the module of the ensemble in Scikit-learn for the RF and AdaBoost models. All of the parameter settings are defaults except the maximum number of estimators is 100 for these models.

Performance measures.

The following measures of goodness of fit were used in this study to evaluate the forecast performance of all the models above:

$$\text{The coefficient of determination } (R^2) = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (1)$$

where $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$

where \bar{y} is the mean value taken over N , y_i is the observed value, \hat{y}_i is the forecasted value and N is the number of data points. The coefficient of determination measures the degree of association among the observed and predicted values.

$$\text{The Root Mean Squared Error (RMSE)} = \sqrt{\frac{\text{SSE}}{N}}, \quad (2)$$

where SSE is the sum of squared errors, and N is the number of samples used. SSE is given by:

$$\text{SSE} = \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (3)$$

with the variables already having been defined.

$$\text{The Mean Absolute Error (MAE)} = \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{N} \quad (4)$$

The MAE is used to measure how close forecasted values are to the observed values. It is the average of the absolute errors.

3 Results

In this study, we developed the drought forecasting models based on the OLS, RR, LR, DT, AdaBoost, and RF to predict the SPEI at different timescales of 3, 6, 12, and 24 months for 118 meteorological stations in Northeast China. In the following sections, we will evaluate the forecasting performance of the models to determine if the penalized linear regression and DT-based ensemble methods can provide performance improvements. Subsequently, we will identify the optimum model among all the drought forecasting models and assess the feasibility by analyzing its forecasting performance for each station in detail.

3.1 Penalized linear regression models

Figure 2 shows the probability density distributions of the RMSE based on the LR, RR, and OLS models at the different timescales of 3, 6, 12, and 24 months; this gives a comparison of the forecasting performance of the penalized linear regression and OLS models. As can be seen from the Fig. 2, the probability density distribution of the

RMSE based on LR and RR was closer to zero than based on OLS. It indicates that the forecast deviations were lower by the penalized linear regression model than by the OLS model. In particular, the probability density distributions of the RMSE based on the LR model at each timescale were all significantly less than those based on the other models. The probability density distributions of the MAE based on the models at different timescales are similar to those shown in

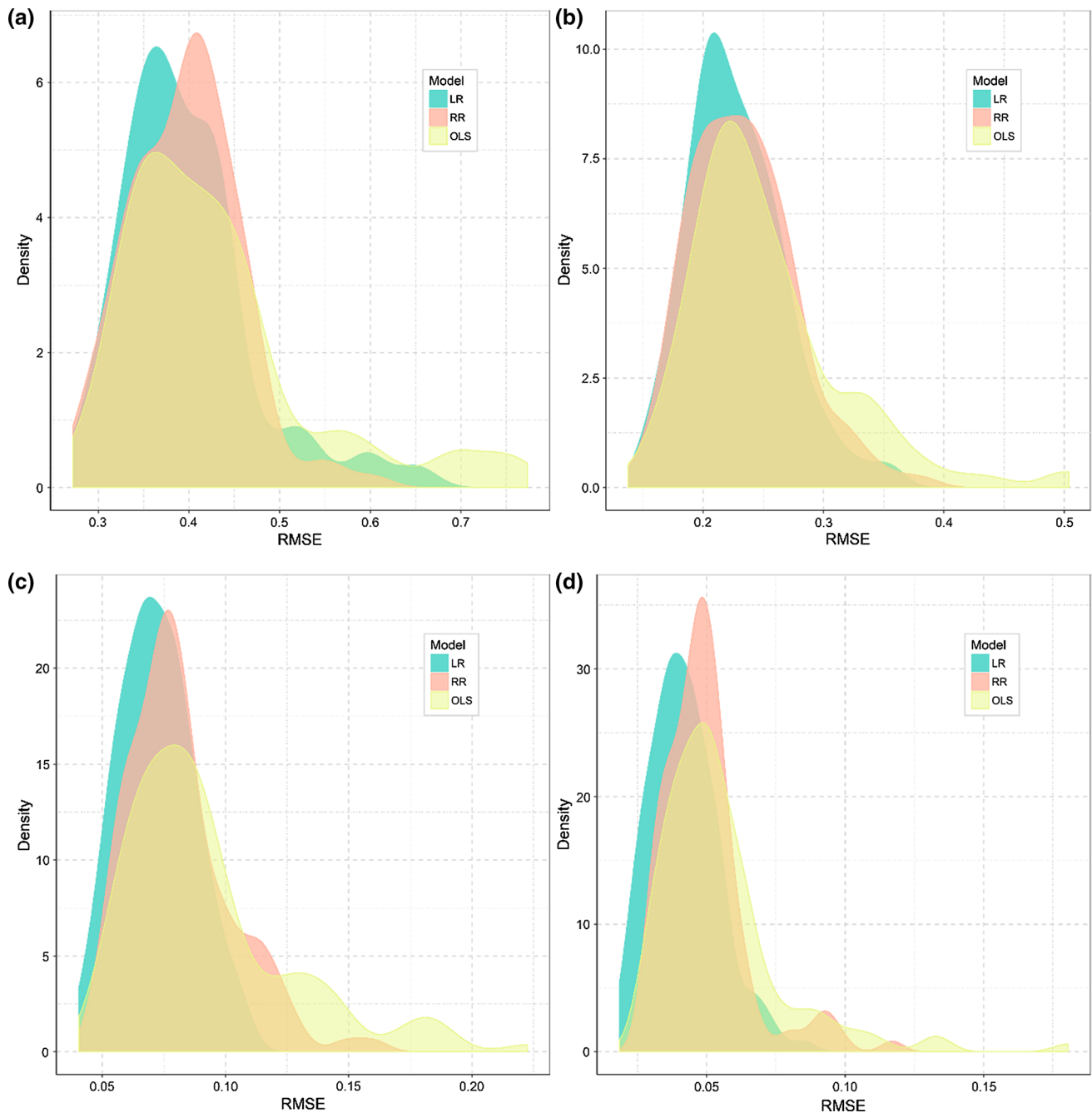


Fig. 2 Probability density distribution of the RMSE for the prediction of **a** SPEI3 **b** SPEI6 **c** SPEI12, and **d** SPEI24 using the OLS, RR, and LR models during the test period (2003–2016) for the 118 meteorological stations in Northeast China

Fig. 2, in that the forecasting performances based on the LR model were better than those of the other models. Table 1 lists the statistical properties of the performance measures of the OLS, RR, and LR models for predicting the SPEI at the different timescales of 3, 6, 12, and 24 months. For the forecasts of SPEI3, the RR model had the lowest average RMSE of 0.3960 (ranging from 0.2765 to 0.6034) and the highest average R^2 of 0.8302 (ranging from 0.4471 to 0.9112). The LR model exhibited a good performance similar to that of the RR, and had the lowest average MAE of 0.3236 (ranging from 0.2211 to 0.4814). The LR model exhibited the best forecasting performance for predicting SPEI6, SPEI12, and SPEI24 among all models and had the lowest average RMSE, the lowest average MAE, and the highest average R^2 along with a small range of the RMSE, MAE, and R^2 . In summary, the results above demonstrated that the penalized linear regression models were more efficient than the OLS model for predicting the SPEI in Northeast China at the different timescales of 3, 6, 12, and 24 months.

A comparison of the number of stations that exhibited the highest R^2 values for the LR, RR, and OLS models indicates that the LR model was the optimum model for predicting the SPEI at different timescales for most of the meteorological stations (Fig. 3). For the prediction of SPEI3, the use of the LR model resulted in 51.7% of the stations, which was a much higher proportion than for the OLS (32.2%) and RR (16.1%) models. The forecasting performances of the models for the mid- and long-term SPEIs were similar with regard to the R^2 values. For the prediction of SPEI6, SPEI12, and SPEI24, the percentages of the stations, for which the LR model was the optimum model were 47.5%, 80.5%, and 78%, respectively. These results suggested that the LR is the optimum model among the penalized linear regression models for predicting the SPEI at the different timescales of 3, 6, 12, and 24 months in Northeast China.

Interestingly, Fig. 4 shows a steady decrease in the forecast deviation as the timescale of the SPEI increased. At the same time, the range of the RMSE and MAE decreased as the timescale of the SPEI increased. For example, the ranges of RMSE for SPEI3, SPEI6, SPEI12, and SPEI24 based on the LR model were 0.3860, 0.2091, 0.0679, and 0.0664, respectively. The only exception was that the range of the MAE was slightly larger for the prediction of SPEI24 (0.1102) using the OLS model than for the prediction of SPEI12 (0.1090). These findings indicate a correlation between forecast deviation and the timescale of the SPEI.

3.2 Ensemble methods

The DT, AdaBoost, and RF models were developed to predict SPEI3, SPEI6, SPEI12, and SPEI24 for the 118 meteorological stations in Northeast China. Figure 5 provides the forecasting performance results evaluated by the RMSE. By

contrasting the probability density distribution of the RMSE and MAE between the simple DT model and the DT-based ensemble methods model, we found that the DT-based ensemble methods model had a lower forecast deviation than the simple DT model. It is apparent from these figures that the probability density distributions of the RMSE and MAE of the RF model were closest to zero, followed by the AdaBoost model and the simple DT model. For the prediction of SPEI3, the average RMSE and MAE of the RF model were 0.4745 and 0.3756, respectively; these values were lower than those of the AdaBoost (average RMSE of 0.5526 and average MAE of 0.4361) model and the DT (average RMSE of 0.8161 and average MAE of 0.6437) model. The RF model also exhibited a lower forecast deviation than the AdaBoost and the DT models in predicting SPEI6 and had an average RMSE and MAE of 0.3075 and 0.2274 whereas the average RMSE and MAE values of the AdaBoost and DT models were 0.3947, 0.3019, and 0.5725, 0.4282, respectively. For the predictions of SPEI12 and SPEI24, the model based on the RF algorithm continuously exhibited the best forecasting performance in terms of the RMSE and MAE; the average RMSE values for these predictions were 0.1674 and 0.1537 and the average MAE were 0.1120 and 0.0996, respectively. Similar to the results for the prediction of the short-term SPEIs, the average RMSE and MAE were lower for the AdaBoost model than for the DT model for the prediction of SPEI12 and SPEI24. The AdaBoost model had average RMSE values of 0.2517 and 0.2214 and average MAE values of 0.1869 and 0.1614, respectively. The DT model had average RMSE values of 0.3303 and 0.2598 and average MAE values of 0.2222 and 0.1771, respectively (Table 2). The results (Table 2) indicate that the DT-based ensemble methods provide better performances than the simple DT model.

To further compare the forecasting performances of the DT, AdaBoost and RF models, we used the R^2 value to determine the optimum model for the majority of the meteorological stations in Northeast China. Table 3 shows the summary statistics for the DT, AdaBoost, and RF models. It is apparent that the RF model was the optimum model for the majority of the meteorological stations at the different timescales of 3, 6, 12, and 24 months. For the prediction of the SPEI3 and SPEI6, the RF model was the optimum model for 92.4% and 97.5% of the meteorological stations, respectively. However, surprisingly, the percentages were 100% for the prediction of SPEI12 and SPEI24. These results demonstrate that the RF model is the optimum model among all DT-based ensemble methods in this study.

To investigate the correlation between the forecast deviation and the SPEI timescales, we compared the inter-correlations among the RMSE and MAE of the DT, AdaBoost, and RF models for the prediction of SPEI at different timescales (Fig. 6). The distribution of the RMSE and

Table 1 RMSE, MAE, and R^2 statistics of the 118 meteorological stations for the prediction of the SPEI3, SPEI6, SPEI12, and SPEI24 using the OLS and penalized linear regression models

	The OLS model			The RR model			The LR model		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
	SPEI3								
Min	0.2727	0.2174	0.0975	0.2765	0.2251	0.4471	0.2722	0.2211	0.5143
Median	0.4070	0.3344	0.8269	0.3992	0.3263	0.8401	0.3822	0.3173	0.8444
Mean	0.4297	0.3434	0.7901	0.3960	0.3239	0.8302	0.3971	0.3236	0.8274
Max	0.7733	0.5476	0.9092	0.6034	0.5198	0.9112	0.6582	0.4814	0.9072
Range	0.5006	0.3302	0.8117	0.3269	0.2947	0.4640	0.3860	0.2602	0.3929
Standard deviation	0.1075	0.0752	0.1315	0.0596	0.0507	0.0591	0.0726	0.0554	0.0730
SPEI6									
Min	0.1375	0.1128	0.6306	0.1431	0.1176	0.7846	0.1485	0.1189	0.8102
Median	0.2368	0.1947	0.9291	0.2312	0.1887	0.9343	0.2236	0.1834	0.9395
Mean	0.2521	0.2034	0.9145	0.2329	0.1919	0.9300	0.2277	0.1880	0.9327
Max	0.5037	0.3847	0.9688	0.3811	0.3301	0.9674	0.3577	0.3083	0.9692
Range	0.3662	0.2719	0.3381	0.2379	0.2124	0.1828	0.2091	0.1894	0.1590
Standard deviation	0.0648	0.0492	0.0527	0.0433	0.0376	0.0262	0.0402	0.0345	0.0269
SPEI12									
Min	0.0497	0.0396	0.9299	0.0472	0.0381	0.9536	0.0405	0.0326	0.9750
Median	0.0846	0.0671	0.9900	0.0771	0.0635	0.9916	0.0710	0.0589	0.9925
Mean	0.0925	0.0725	0.9856	0.0811	0.0663	0.9898	0.0719	0.0595	0.9922
Max	0.2226	0.1487	0.9971	0.1602	0.1347	0.9972	0.1084	0.0946	0.9975
Range	0.1729	0.1090	0.0671	0.1131	0.0966	0.0436	0.0679	0.0620	0.0225
Standard deviation	0.0339	0.0241	0.0129	0.0210	0.0185	0.0061	0.0150	0.0135	0.0036
SPEI24									
Min	0.0283	0.0213	0.9680	0.0281	0.0227	0.9856	0.0185	0.0147	0.9886
Median	0.0504	0.0408	0.9968	0.0473	0.0384	0.9974	0.0402	0.0334	0.9981
Mean	0.0549	0.0434	0.9955	0.0491	0.0404	0.9966	0.0417	0.0349	0.9975
Max	0.1806	0.1315	0.9993	0.1170	0.1061	0.9993	0.0849	0.0699	0.9995
Range	0.1523	0.1102	0.0313	0.0890	0.0834	0.0137	0.0664	0.0552	0.0110
Standard deviation	0.0233	0.0173	0.0050	0.0152	0.0133	0.0025	0.0124	0.0110	0.0018

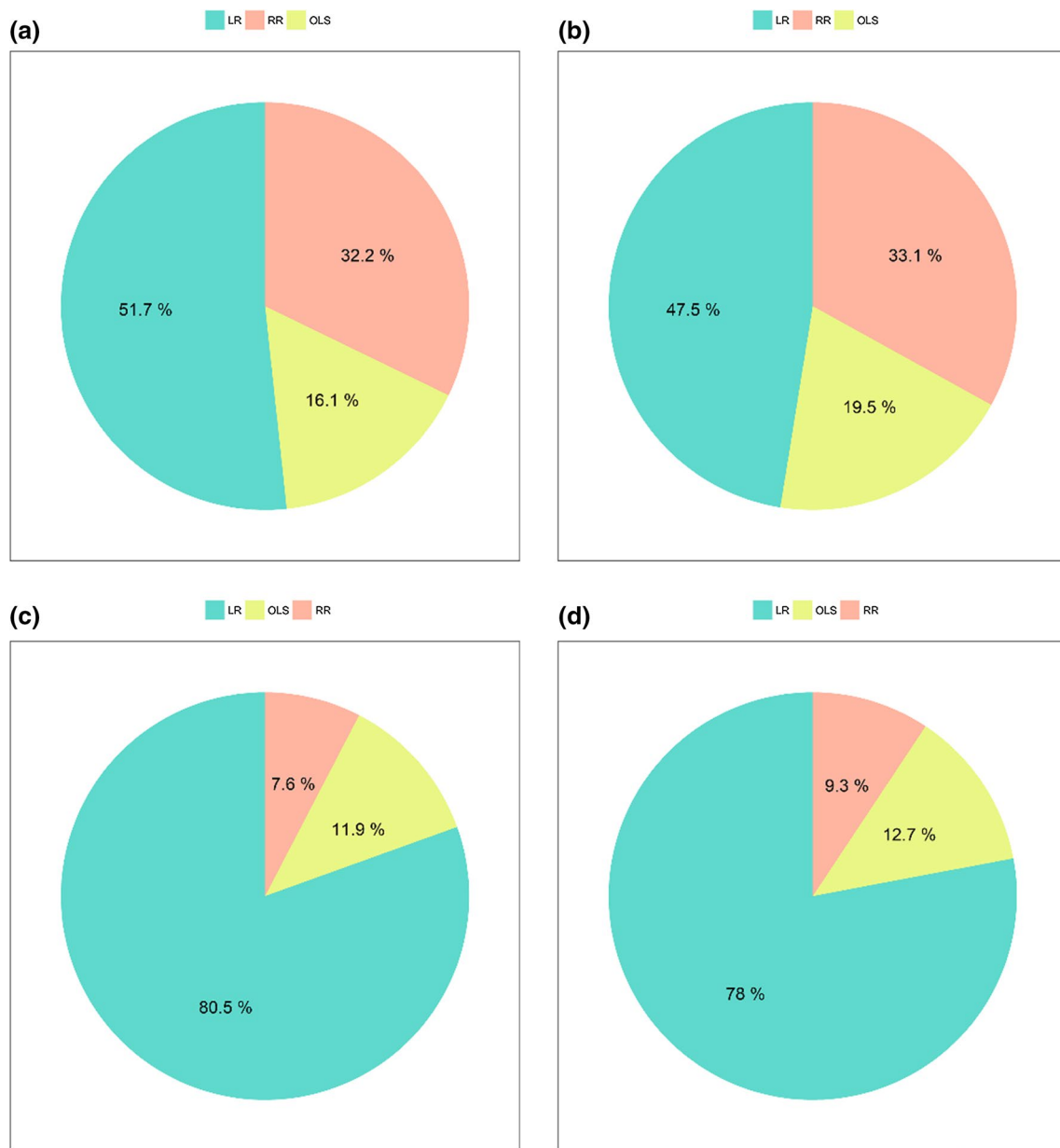


Fig. 3 Percentage of the stations with the highest R^2 for the prediction of the **a** SPEI3; **b** SPEI6; **c** SPEI12, and **d** SPEI24 using the OLS, RR, and LR models

MAE of the AdaBoost and RF models shows a decreasing trend as the timescale of the SPEI increases. However, the trend of the DT model differs from the trends of the other models in that the distributions of the RMSE and MAE do not decrease as the timescale increases and there was no significant correlation between the distribution and the timescale. Overall, these results indicate that there is a correlation between forecast deviation and the timescale of the SPEI for the DT-based ensemble methods but not for the simple DT model.

3.3 Comparison of penalized linear regression and ensemble methods.

To assess the forecasting performance of the penalized linear regression and ensemble methods in this study, we compared the LR and RF models because they were the optimum models of the two respective methods. Figure 7 shows the comparison of the distribution of the RMSE of the LR and RF models at the different timescales of 3, 6, 12, and 24 months. The violin plot shows that the distribution

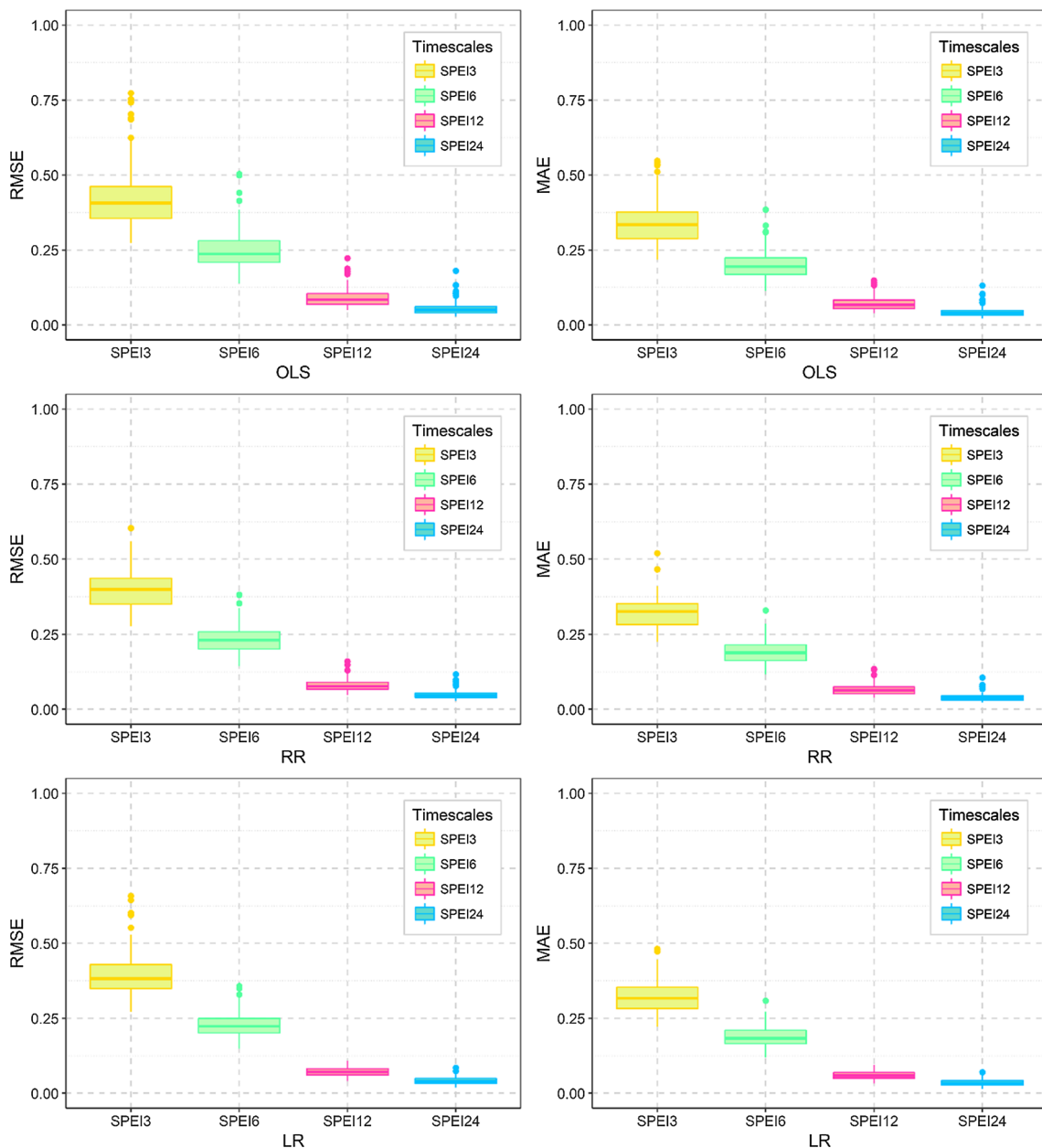


Fig. 4 Boxplot of the RMSE and MAE for the prediction of the SPEI at multiple timescales using the OLS, RR, and LR models

ranges of the RMSE are smaller for the LR model than the RF model at each timescale. For instance, the ranges of the RMSE of the LR model at the different timescales of 3, 6, 12, and 24 months were 0.2722–0.6582, 0.1485–0.3577, 0.0405–0.1084, and 0.0185–0.0849, respectively. In contrast, the ranges of the RF model were 0.3324–0.7637, 0.2146–0.4599, 0.1045–0.2924, and 0.0627–0.5807, respectively. The results indicate that the forecasting performance of the LR model is superior to that of the RF model.

The next section of the study addressed the feasibility of the LR model for the prediction of the SPEI at different timescales for the 118 meteorological stations in Northeast

China. The summary statistics for the forecasting performance of the LR model (Online Resource 1) indicates that the Xinbin station had the lowest R^2 (0.5143) value and the highest RMSE (0.6582) and MAE (0.4814) values of the 118 meteorological stations in the prediction of SPEI3. Thus, the Xinbin station had the worst performance for the prediction of the SPEI. Of the 118 meteorological stations, the Dalian station had the worst performance for the prediction of the SPEI6 and had the lowest R^2 value (0.8102), and the highest RMSE (0.3577) and MAE (0.3083) values. For the prediction of the SPEI12 and SPEI24, the R^2 values were inconsistent with the RMSE and MAE values. In terms of the highest

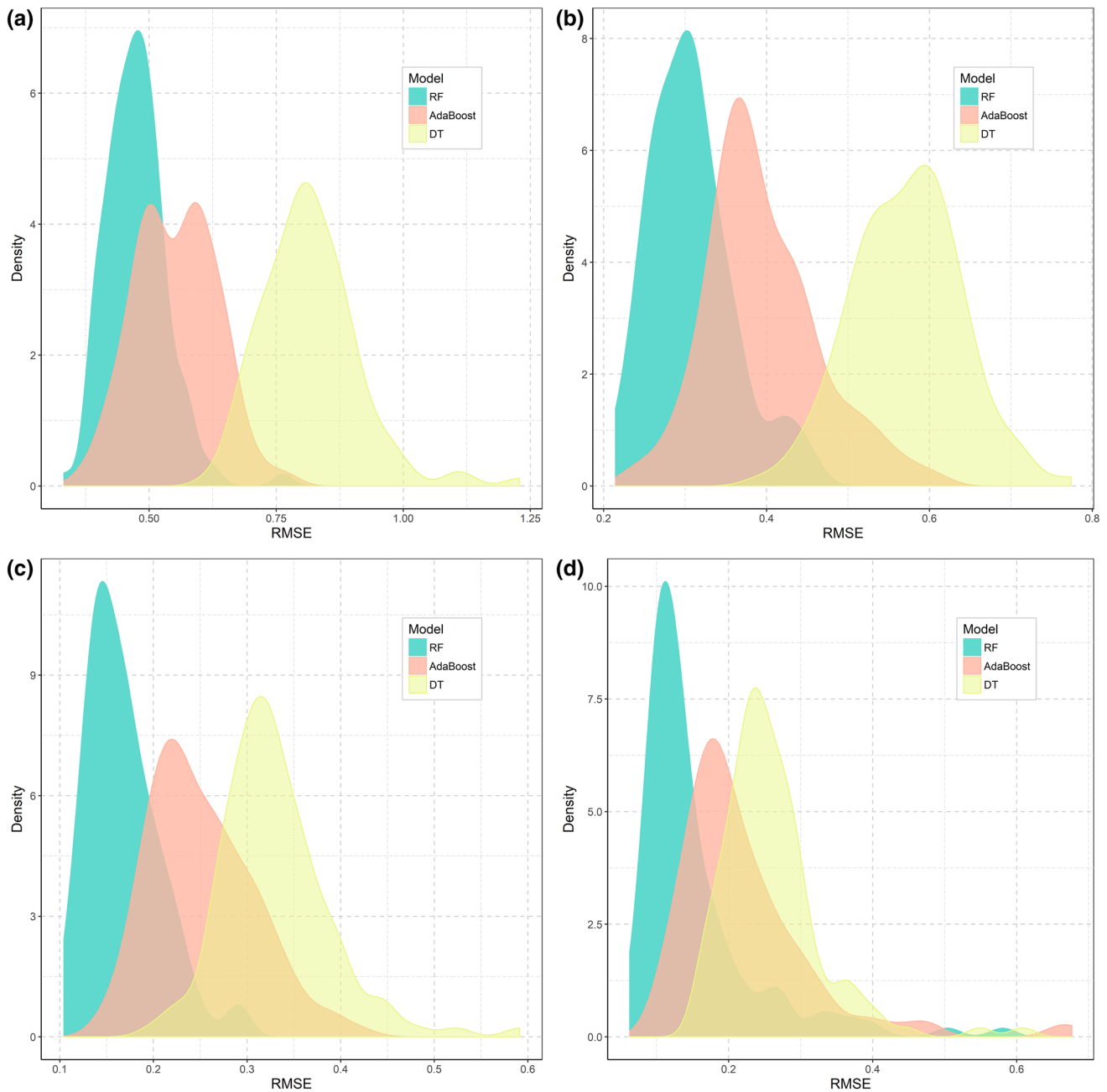


Fig. 5 Probability density distribution of the RMSE for the prediction of **a** SPEI3 **b** SPEI6 **c** SPEI12, and **d** SPEI24 using the DT, AdaBoost, and RF models during the test period (2003–2016) for the 118 meteorological stations in Northeast China

RMSE and MAE values, the Dalian Station and Zhurihe Station had the worst performance, respectively. The Dalian Station had the highest RMSE (0.1084) and MAE (0.0946) values for the prediction of the SPEI12; whereas the Zhurihe Station had the highest RMSE (0.0849) and MAE (0.0699) values for the prediction of the SPEI24. However, their R^2 values were 0.9852 and 0.9923, respectively, and these were not the lowest values among the 118 meteorological stations for the prediction of the SPEI12 and SPEI24. The Mingshui

Station had the lowest R^2 value of the 118 meteorological stations for the prediction of the SPEI12 and SPEI24 (0.9750 and 0.9886, respectively). Although the R^2 values of the Dalian Station and Zhurihe Station were higher than those of the Mingshui Station for the prediction of the SPEI12 and SPEI24, there was only a slight difference and the R^2 values were greater than 0.975. Hence, the Dalian Station and Zhurihe Station had the worst performances for the prediction of the SPEI12 and SPEI24.

Table 2 RMSE, MAE, and R^2 statistics of the 118 meteorological stations for the prediction of SPEI3, SPEI6, SPEI12, and SPEI24 using the DT, AdaBoost, and RF models

	The DT model			The AdaBoost model			The RF model		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
SPEI3									
Min	0.6177	0.4880	-0.9016	0.3653	0.2926	0.3615	0.3324	0.2559	0.3996
Median	0.8112	0.6390	0.3210	0.5567	0.4331	0.6831	0.4711	0.3744	0.7597
Mean	0.8161	0.6437	0.2797	0.5526	0.4361	0.6731	0.4745	0.3756	0.7583
Max	1.2288	1.0013	0.6382	0.7752	0.6317	0.8661	0.7637	0.6222	0.8624
Range	0.6111	0.5134	1.5398	0.4099	0.3392	0.5046	0.4314	0.3663	0.4628
Standard deviation	0.0959	0.0745	0.2087	0.0794	0.0667	0.0834	0.0600	0.0499	0.0644
SPEI6									
Min	0.3945	0.3049	0.3502	0.2345	0.1873	0.5916	0.2146	0.1588	0.7171
Median	0.5752	0.4274	0.5922	0.3816	0.2907	0.8157	0.3021	0.2237	0.8846
Mean	0.5725	0.4282	0.5821	0.3947	0.3019	0.8019	0.3075	0.2274	0.8794
Max	0.7738	0.6463	0.8047	0.6080	0.4843	0.9039	0.4599	0.3765	0.9309
Range	0.3793	0.3414	0.4545	0.3735	0.2970	0.3123	0.2453	0.2177	0.2137
Standard deviation	0.0661	0.0538	0.0927	0.0692	0.0547	0.0579	0.0520	0.0387	0.0365
SPEI12									
Min	0.1990	0.1473	0.5370	0.1494	0.1167	0.6974	0.1045	0.0739	0.8634
Median	0.3219	0.2147	0.8559	0.2404	0.1811	0.9157	0.1615	0.1071	0.9620
Mean	0.3303	0.2222	0.8328	0.2517	0.1869	0.9023	0.1674	0.1120	0.9563
Max	0.5911	0.4296	0.9483	0.4198	0.2814	0.9673	0.2924	0.2108	0.9867
Range	0.3921	0.2823	0.4113	0.2703	0.1646	0.2699	0.1879	0.1368	0.1233
Standard deviation	0.0583	0.0420	0.0782	0.0549	0.0372	0.0508	0.0387	0.0229	0.0240
SPEI24									
Min	0.1603	0.1073	0.3412	0.1014	0.0821	0.0550	0.0627	0.0465	0.4439
Median	0.2463	0.1694	0.9227	0.1968	0.1503	0.9502	0.1281	0.0854	0.9793
Mean	0.2598	0.1771	0.9037	0.2214	0.1614	0.9234	0.1537	0.0996	0.9596
Max	0.6081	0.4419	0.9733	0.6764	0.5003	0.9896	0.5807	0.4050	0.9930
Range	0.4478	0.3346	0.6321	0.5751	0.4182	0.9346	0.5180	0.3585	0.5491
Standard deviation	0.0700	0.0504	0.0815	0.0952	0.0636	0.1100	0.0848	0.0530	0.0730

Table 3 Number and percentage of the meteorological stations with the highest R^2 for the prediction of the SPEI using the DT, AdaBoost and RF models

Number and percentage of the meteorological stations			
	The DT model	The AdaBoost model	The RF model
SPEI3	0 (0%)	9 (7.6%)	109 (92.4%)
SPEI6	0 (0%)	3 (2.5%)	115 (97.5%)
SPEI12	0 (0%)	0 (0%)	118 (100%)
SPEI24	0 (0%)	0 (0%)	118 (100%)

Figure 8 shows the monthly observed and predicted SPEI values of the stations (Xinbin station, Dalian station, Dalian station and Zhurihe station) with the worst forecasting performance at the different timescales of 3, 6, 12, and 24 months during the test period (2003–2016). There was a very good agreement between the predicted and observed SPEI values. Some other stations had better goodness of fit due to lower RMSE and MAE values. Moreover, the goodness of fit between the predicted and observed SPEI increased with increasing timescales. In summary, the LR model exhibited an acceptable forecasting performance of the SPEI at the different timescales of 3, 6, 12, and 24 months for the 118 meteorological stations in Northeast China.

4 Discussion

In this study, we applied a variety of machine learning models to predict the monthly SPEI at the different timescales of 3, 6, 12, and 24 months for 118 meteorological stations in Northeast China during the period of 1963–2016. Two types of algorithms were evaluated: the penalized linear regression (the RR and LR) and the DT-based ensemble methods (the AdaBoost and RF). The goal of this study was to investigate the feasibility of using penalized linear regression and DT-based ensemble methods for forecasting drought conditions in Northeast China. The primary findings of this study are as follows. (1) The penalized linear regression achieved better forecasting performance than the OLS algorithm and the LR model had the best performance. (2) Another significant finding is that the DT-based ensemble methods had higher prediction accuracy than the simple DT algorithm for the prediction of the SPEI at different timescales. In particular, the RF model had the best forecasting performance. (3) A comparison of the optimum models of the two types of algorithms indicated that the LR model was superior to the RF model for the prediction of the SPEI at different timescales in Northeast China.

As expected, the results indicate that the penalized linear regression was more effective than the traditional OLS for the prediction of the SPEI at different timescales because of the lower forecast deviations of the stations. In the penalized linear regression, the OLS overfitting problem is solved by adding a penalty term to the least squares estimators for coefficients that are very small or zero, which improves the prediction accuracy. However, we found that the LR model had a better forecasting performance than the RR model in most cases. The best performance for the majority of the meteorological stations was achieved using the LR model. These results seem to be consistent with other studies that reported the LR resulted in significantly higher accuracy than the RR for electricity price forecasting (Uniejewski et al. 2016). However, one of the elastic net models was the best performing model in their research. We did not use the elastic net algorithm due to its penalty term, which was already included in the LR and RR. Elastic net methods introduce another parameter to adjust the ratio of the penalty for the RR and LR. Further research should be conducted to investigate the forecasting performance of other penalized linear regression models such as the elastic net model for the prediction of the SPEI. One unanticipated finding was that the forecasting performance of the models for the prediction of the SPEI improved with increasing timescales. This result is in agreement with the results reported by Park et al. (2016), who stated that the prediction accuracy was higher for long-term drought conditions than for short-term drought conditions. It is difficult to determine the specific reason for this result but it might be related to the reciprocal causal relationship between drought factors and the SPEI. Drought factors tend to represent the influence of precipitation shortages accumulated over long-term rather than short-term periods (Gessner et al. 2013).

In the current study, it is apparent that the DT-based ensemble methods had better forecasting performance than the simple DT algorithm, especially the RF model. We used the probability density distribution of the RMSE and MAE to determine the overall forecasting performance of the models. The results consistently indicated that the forecast deviations were lowest for the RF model for all timescales, followed by the AdaBoost model and the DT model. In addition, the use of the R^2 values for determining the optimum model for each station also showed that the RF model was the optimum model for most stations. It may be that the voting mechanism of the multiple tree predictors in the RF algorithm has an advantage over the overfitting problem of the DT algorithm. Also, The RF method is less time-consuming, which represents a considerable advantage for predictions task. A possible explanation for this might be that the prediction accuracy depends not only on the algorithms but also on the size, dimension, and integrity of the dataset and the degree of correlation between the variables.

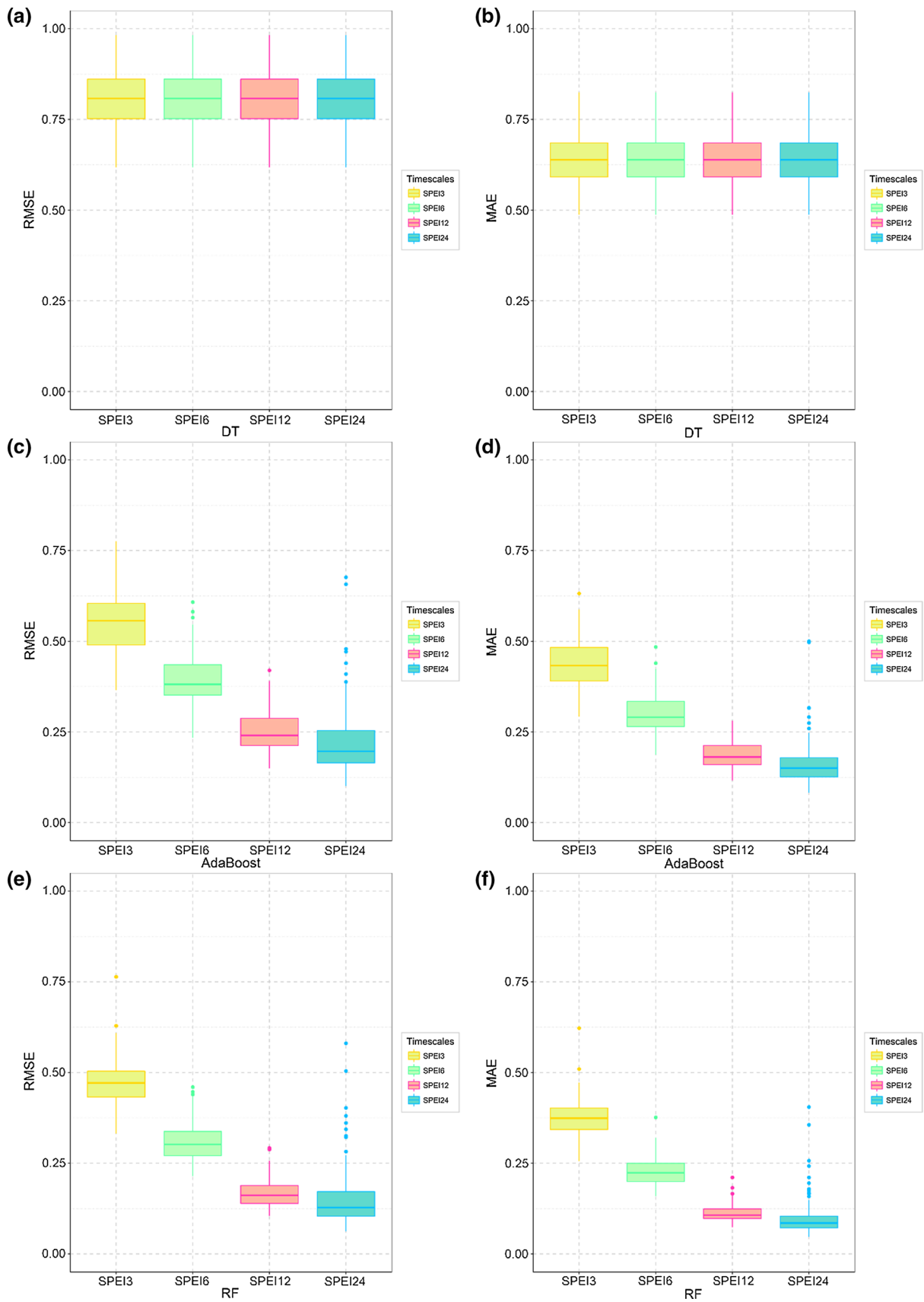


Fig. 6 Boxplot of the RMSE and MAE for the prediction of the SPEI at multiple timescales using the DT, AdaBoost, and RF models

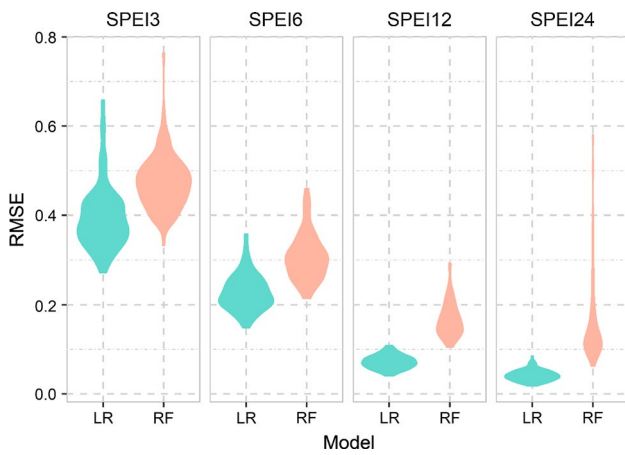


Fig. 7 Violin plot of the RMSE for the prediction of the SPEI at multiple timescales using the LR and RF models

We compared the forecasting performance of the LR and RF models to identify the optimum model. The results indicate that the LR model performed better than the RF model for all stations according to the ranges and distributions of the RMSE, MAE, and R^2 of the models. It seems that no single machine learning algorithm has outperformed other algorithms for the SPEI prediction in these all regions. The reasons may be related to the characteristic differences between the SPEI datasets in the different study regions. In addition, the timescales of the SPEI have a significant impact on the performance of the forecasting models. Therefore, the selection of the most suitable SPEI is more important than the type of machine learning algorithm. There is abundant opportunity to investigate specific models and to optimize the prediction performance. Further research should take into account the temporal characteristics of the SPEI and utilize deep learning methods to establish drought forecasting

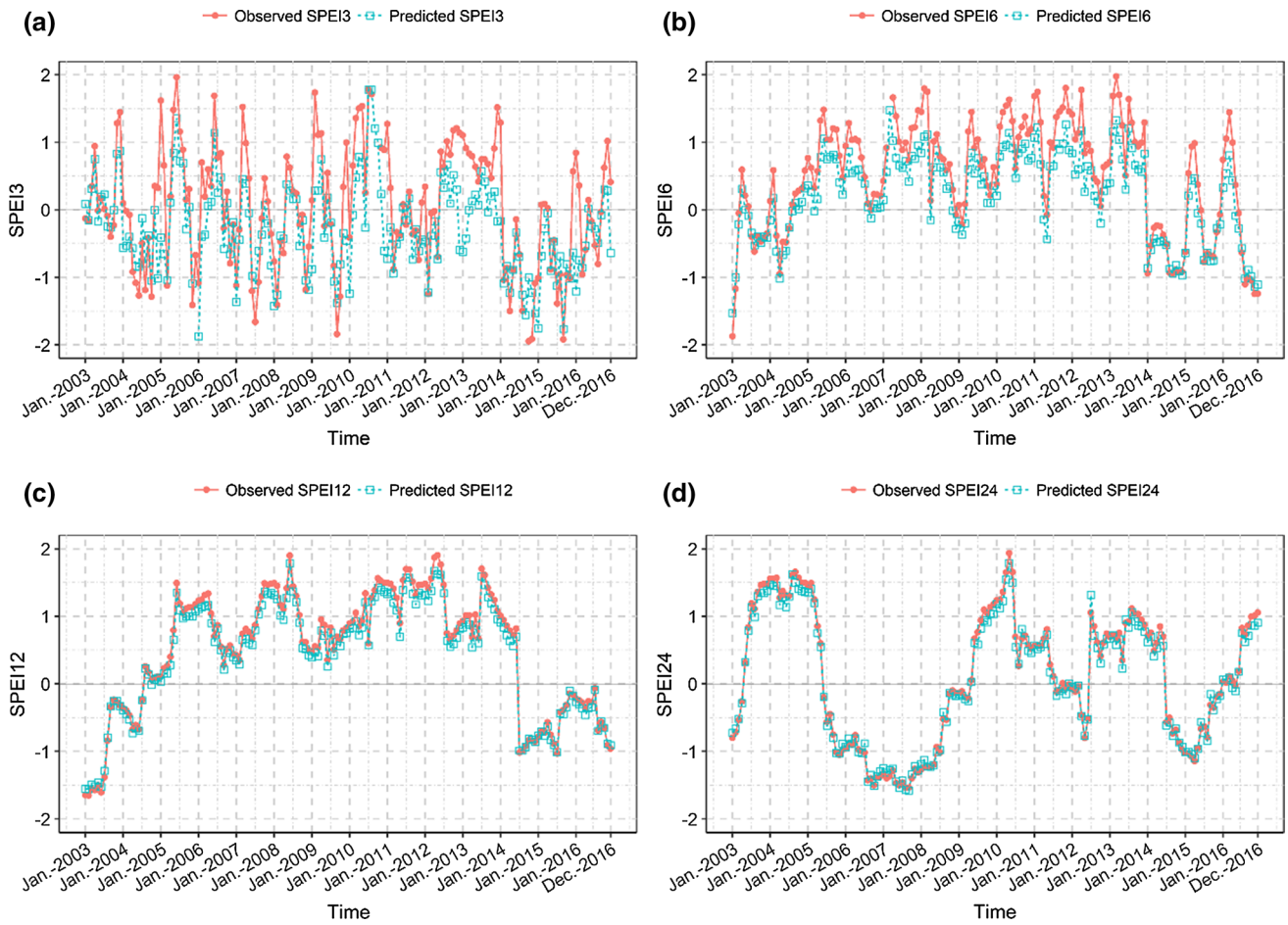


Fig. 8 Prediction results for the SPEIs at multiple timescales using the LR model at **a** Xinbin station **b** Dalian station **c** Dalian station, and **d** Zhurihe station

models. In addition, the use of different timescales of the SPEI or different drought indices may also improve the predictive performance of the model. Because the sample size was limited by the size of the study area, future research should use a dataset comprising a larger number of stations to improve the generalization ability of the drought forecasting models.

5 Conclusion

This study evaluated the ability of two machine learning methods (i.e., penalized linear regression and DT-based ensemble methods) for the prediction of the SPEI at the different timescales of 3, 6, 12, and 24 months in Northeast China. The penalized linear regression models provided better prediction results than the OLS model. Furthermore, the DT-based ensemble methods models had better forecasting performance than the simple DT model. Among all the drought forecasting models, the LR model consistently exhibited the best prediction accuracy regardless of the SPEI timescales. These findings suggest that the LR model may be applied to predict drought conditions in Northeast China. This research provides a framework for the exploration of machine learning approaches for the prediction of drought conditions. Considering the expected effect of global warming, the improvement of drought prediction models is a necessary approach to mitigate drought losses and achieve sustainable development of water.

Acknowledgements This work was supported by the National Science Foundation of China (Grants Nos. 51679142 and 51709173).

References

- Ali Z et al (2017) Forecasting drought using multilayer perceptron artificial neural network model. *Adv Meteorol*. <https://doi.org/10.1155/2017/5681308>
- Azad A, Manoochehri M, Kashi H, Farzin S, Karami H, Nourani V, Shiri J (2019) Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling. *J Hydrol* 571:214–224. <https://doi.org/10.1016/j.jhydrol.2019.01.062>
- Bachmair S, Svensson C, Hannaford J, Barker L, Stahl K (2016) A quantitative analysis to objectively appraise drought indicators and model drought impacts. *Hydrol Earth Syst Sci* 20:2589–2609
- Bachmair S, Svensson C, Prosdociimi I, Hannaford J, Stahl K (2017) Developing drought impact functions for drought risk management. *Nat Hazards Earth Syst Sci* 17:1947–1960. <https://doi.org/10.5194/nhess-17-1947-2017>
- Beguieria S, Vicente-Serrano SM, Reig F, Latorre B (2014) Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int J Climatol* 34:3001–3023. <https://doi.org/10.1002/joc.3887>
- Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmos Res* 172:37–47
- Borji M, Malekian A, Salajegheh A, Ghadimi M (2016) Multi-time-scale analysis of hydrological drought forecasting using support vector regression (SVR) and artificial neural networks (ANN). *Arab J Geosci* 9:725
- Botai C, Botai J, Dlamini L, Zwane N, Phaduli E (2016) Characteristics of droughts in South Africa: a case study of free state and north west provinces. *Water* 8:439
- Breiman L (1996) Bagging predictors machine learning 24:123–140
- Breiman L (2001) Random forests machine learning 45:5–32
- Byakatonda J, Parida B, Kenabatho P, Moalafhi D (2016) Modeling dryness severity using artificial neural network at the Okavango Delta. *Botswana Glob Nest J* 18:463–481
- Caruana R, Karampatziakis N, Yessenalina A (2008) An empirical evaluation of supervised learning in high dimensions. In: *Proceedings of the 25th international conference on Machine learning*, ACM, pp 96–103
- Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on Machine learning*, ACM, pp 161–168
- Chen T, Xia G, Liu T, Chen W, Chi D (2016) Assessment of drought impact on main cereal crops using a standardized precipitation evapotranspiration index in Liaoning Province. *China Sustain* 8:1069
- Cook BI, Smerdon JE, Seager R, Coats S (2014) Global warming and 21st century drying. *Clim Dyn* 43:2607–2627
- Dai A (2011) Drought under global warming: a review. *Wiley Interdiscip Rev Clim Change* 2:45–65
- Deo RC, Şahin M (2015) Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos Res* 161:65–81
- Dietterich TG (2000) Ensemble methods in machine learning. In: *International workshop on multiple classifier systems*. Springer, pp 1–15
- Freund Y (1995) Boosting a weak learning algorithm by majority. *Inf Comput* 121:256–285
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55:119–139
- Ganguli P, Reddy MJ (2014) Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrol Process* 28:4989–5009
- Gessner U, Naeimi V, Klein I, Kuenzer C, Klein D, Dech S (2013) The relationship between precipitation anomalies and satellite-derived vegetation activity in Central Asia. *Glob Planet Change* 110:74–87
- Gill MK, Asefa T, Kemblowski MW, McKee M (2006) Soil moisture prediction using support vector machines. *JAWRA J Am Water Res Assoc* 42:1033–1046
- Gocic M, Trajkovic S (2014) Drought characterisation based on water surplus variability index water. *Resour Manag* 28:3179–3191. <https://doi.org/10.1007/s11269-014-0665-4>
- Gocic M, Trajkovic S (2014) Water surplus variability index as an indicator of drought. *J Hydrol Eng* 20:04014038
- Guttman NB (1998) Comparing the palmer drought index and the standardized precipitation index JAWRA. *J Am Water Resour Assoc* 34:113–121
- Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67
- James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning*, vol 112. Springer, Berlins

- Karimi S, Sadraddini AA, Nazemi AH, Xu T, Fard AF (2018) Generalizability of gene expression programming and random forest methodologies in estimating cropland and grassland leaf area index. *Comput Electron Agric* 144:232–240. <https://doi.org/10.1016/j.compag.2017.12.007>
- Karimi S, Shiri J, Kisi O, Xu T (2018) Forecasting daily streamflow values: assessing heuristic models. *Hydrol Res* 49:658–669. <https://doi.org/10.2166/nh.2017.111>
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: *Proceedings of the 8th conference on applied climatology*, vol 22. American Meteorological Society Boston, MA, pp 179–183
- Kong Q, Ge Q, Zheng J, Xi J (2015) Prolonged dry episodes over Northeast China during the period 1961–2012. *Theor Appl Climatol* 122:711–719
- Lantz B (2013) *Machine learning with R*. Packt Publishing Ltd, Li Z, Zhou T (2015) Responses of vegetation growth to climate change in China. *Int Arch Photogramm Remote Sens Spat Inf Sci* 40:225
- Maca P, Pech P (2016) Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. *Comput Intell Neurosci* 2016:14
- Niemeyer S (2008) New drought indices Options. *Méditerranéennes Série A: Séminaires Méditerranéens* 80:267–274
- Ortengren JT, Knapp PA, Maxwell JT, Tyminski WP, Soulé PT (2011) Ocean–atmosphere influences on low-frequency warm-season drought variability in the Gulf Coast and southeastern United States. *J Appl Meteorol Climatol* 50:1177–1186
- Park S, Im J, Jang E, Rhee J (2016) Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. *Agric For Meteorol* 216:157–169
- Park S, Seo E, Kang D, Im J, Lee MI (2018) Prediction of drought on pentad scale using remote sensing data and MJO index through random forest over East Asia. *Remote Sens* 10:18. <https://doi.org/10.3390/rs10111811>
- Pedregosa F et al. (2011) Scikit-learn: Machine learning in Python *Journal of machine learning research* 12:2825–2830.
- Peng J, Dong W, Yuan W, Zhang Y (2012) Responses of grassland and forest to temperature and precipitation changes in Northeast China. *Adv Atmos Sci* 29:1063–1077
- Pereira JM, Basto M, da Silva AF (2016) The logistic lasso and ridge regression in predicting corporate failure. *Procedia Econ Financ* 39:634–641
- Reiss MA et al (2015) Improvements on coronal hole detection in SDO/AIA images using supervised classification. *J Space Weather Space Clim* 5:A23
- Rhee J, Im J (2017) Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data. *Agric For Meteorol* 237:105–122
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodological)* 58:267–288
- Touma D, Ashfaq M, Nayak MA, Kao S-C, Diffenbaugh NS (2015) A multi-model and multi-index evaluation of drought characteristics in the 21st century. *J Hydrol* 526:196–207s
- Trevor H, Robert T, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York
- Tsakiris G, Vangelis H (2005) Establishing a drought index incorporating evapotranspiration. *Eur Water* 9:3–11
- Uniejewski B, Nowotarski J, Weron R (2016) Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* 9:621
- Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J Clim* 23:1696–1718
- Vicente-Serrano SM, Van der Schrier G, Beguería S, Azorin-Molina C, Lopez-Moreno JI (2015) Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *J Hydrol* 526:42–54. <https://doi.org/10.1016/j.jhydrol.2014.11.025>
- Wang WX, Zuo DD, Feng GL (2014) Analysis of the drought vulnerability characteristics in Northeast China based on the theory of information distribution and diffusion. *Acta Phys Sin* 63:11. <https://doi.org/10.7498/aps.63.229201>
- Wang X, Shen H, Zhang W, Cao J, Qi Y, Chen G, Li X (2015) Spatial and temporal characteristics of droughts in the Northeast China. *Transect Nat Hazards* 76:601–614
- Wayne CP (1965) *Meteorological drought US weather bureau research paper* 58
- Wells N, Goddard S, Hayes MJ (2004) A self-calibrating Palmer drought severity index. *J Clim* 17:2335–2351
- Wilhite DA (2000) *Drought as a natural hazard: concepts and definitions*
- Wu X et al. (2008) Top 10 algorithms in data mining *Knowledge and information systems* 14:1–37.
- Yin X et al (2016) Adapting maize production to drought in the Northeast Farming Region of China. *Eur J Agron* 77:47–58
- Yu X, He X, Zheng H, Guo R, Ren Z, Zhang D, Lin J (2014) Spatial and temporal analysis of drought risk during the crop-growing season over northeast China. *Nat Hazards* 71:275–289
- Zargar A, Sadiq R, Naser B, Khan FI (2011) A review of drought indices. *Environ Rev* 19:333–349
- Zhang Y, Xin Y, Li Q, Ma J, Li S, Lv X, Lv W (2017) Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. *Biomed Eng Online* 16:125
- Zhou Z-H (2012) *Ensemble methods: foundations and algorithms*. Chapman and Hall, London

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.