Ecological Indicators 112 (2020) 106084

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Establishment of agricultural drought loss models: A comparison of statistical methods

Xiufang Zhu[a,b], Chenyao Hou[c,*], Kun Xu[c], Ying Liu[c]

[a] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[b] Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China
[c] Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Keywords:
Drought loss
Multivariate stepwise regression
Random forest model
Winter wheat

ABSTRACT

Agricultural drought loss models provide services for the rapid risk assessment of agricultural disasters, and regional disaster prevention and mitigation efforts. This paper takes wheat as an example, and chooses counties dominated by rain-fed farmland in Henan Province as the study area. Counties dominated by rain-fed farmland are determined by setting a rain-fed threshold that is related to the proportion of the effective irrigation area to the cultivated land area. Modeling samples are screened by considering both drought occurrence time and wheat yield reductions. Under different thresholds (30%, 40%, 50% and 60%), we use the yield loss ratio as the dependent variable and 24 standardized precipitation evapotranspiration index parameters as independent variables to build drought loss models using both a multivariate stepwise regression model and a random forest model. Yield loss ratio from 1990 to 2015 is calculated by decomposing historical wheat yield time series. 24 standardized precipitation evapotranspiration index variables are 1–3 months' time scale standardized precipitation evapotranspiration index during the growth period (from October to May of the following year) of winter wheat in Henan Province. The results show that the random forest-derived model outperforms the stepwise regression model in all tests. The accuracy of all the models increases with an increase of the proportion of the rain-fed threshold. When the rain-fed threshold is 60%, the R^2 values of the random forest model and the multivariate stepwise regression equation are 0.720 and 0.523, respectively. The validation results show that the mean absolute error and the root mean square error of the multivariate stepwise regression are 1.38 times and 1.31 times larger than the mean absolute error and the root mean square error from the random forests model. Moreover, both models identify that standardized precipitation evapotranspiration indices in October (sowing/planting stage) and February (overwintering stage) are important variables. However, the multivariate stepwise regression model fails to recognize the importance of standardized precipitation evapotranspiration indices during April–May (filling stage).

1. Introduction

Drought ranks as one of the most prevalent natural disasters in the world (Chen et al., 2012). The economic losses caused by drought each year far exceed those caused by other meteorological disasters. Climate change in recent decades has led to an increasing trend of drought, and models predict that the global drought risk will further intensify in the 21st century (Dai, 2013; Su et al., 2018). Research on drought has attracted the attention of scholars, government departments and the public. A large number of studies have focused on the construction of drought indices (Heim, 2002; Liu et al., 2018b; Vicente-Serrano et al., 2010; Vicente-Serrano et al., 2012; Zargar et al., 2011), drought monitoring (Bolten et al., 2010; Narasimhan and Srinivasan, 2005; Sun et al., 2018; Svoboda et al., 2002; Zhu et al., 2019), and early warning and forecasting (Cancelliere et al., 2007; Kim and Valdés, 2003; Mishra and Desai, 2005; Sheffield et al., 2014; Zhang et al., 2019). Drought risk assessment and management are important tools for mitigating the negative impacts of drought (Aboelghar et al., 2014). It is important to better understand and predict not only drought itself but also the possible consequences of drought (Bachmair et al., 2017). As such, there is an urgent need to strengthen research activities related to the drought-loss and drought-risk assessment (Ali et al., 2019a, b; Liu et al., 2018a). However, there are few studies to date that have focused on the drought risk and loss assessment rather than other drought-related subjects

* Corresponding author at: Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China.
E-mail address: houcy@mail.bnu.edu.cn (C. Hou).

https://doi.org/10.1016/j.ecolind.2020.106084
Received 17 September 2019; Received in revised form 30 December 2019; Accepted 6 January 2020
Available online 13 January 2020
1470-160X/

(Bachmair et al., 2016).

The establishment of drought loss models is the key to estimating drought loss and assessing drought risk. The direct economic losses due to agricultural drought disasters can be quantified using crop yield losses as a proxy. Two main methods have been proposed for assessing crop yield losses: mechanism models and statistical models. A mechanism model usually uses a crop growth model to set up different water scarcity scenarios to simulate crop yield losses under different degrees of water stress. Crop growth models integrate the effects of plant characteristics, soil properties and environmental conditions and see how they affect crop growth and yields (Boote et al., 1996; Jones et al., 2003; Steduto et al., 2009; Stockle et al., 2003). The establishment of a drought loss model based on a crop growth model usually needs two steps. The first step is to simulate crop yield losses under different water deficit conditions and calculate the yield loss rate. The second step is to construct a drought intensity index and establish a statistical relationship between it and the yield loss rate. For example, with different irrigation scenarios settings, Jia et al. (2012) and Yin et al. (2014) simulated the response of maize yield to water stress in China and at the global scale using the erosion-productivity impact calculator (EPIC) model with the Geographic Information System, respectively. Similarly, Wang et al. (2013) and Yue et al. (2015) simulated the yield response of wheat to drought in China with the EPIC model, while Wei et al. (2019) simulated the yield of summer maize on Huaibei Plain, China under different drought scenarios using the Crop Environment Resource Synthesis (CERES) model. In all these studies, the authors fitted the relationships between the drought intensity indices and the yield loss rates with logistical functions.

Statistical models establish a functional relationship between historical yield loss data and drought variables (such as various drought indices). Although statistical models do not consider the underlying physical mechanisms, they are easy to operate and are able to explain yield responses qualitatively and quantitatively in terms of drought disaster. Previous studies have acquired valuable results for loss assessment of crop yields using statistical models. For example, Zhang (2004) established a quadratic equation to explain the loss of maize yield on Songliao Plain of China. Ming et al. (2015) analyzed a regression relationship between the de-trended maize yield and the standardized precipitation evapotranspiration index (SPEI) in the North China plain (NCP) and found that the three-month SPEIs in August, which reflected the water conditions during June and July, had the best relationship with the de-trended maize yield. Xu et al. (2018) built a multivariate regression model between the de-trended winter wheat yield and the SPEI in Jiangsu Province, China. Wang et al. (2018) established seven aggregate drought indices to quantify the relationships between them and the anomalies of the climatic yields (that is, a standardization of climatic yield) of wheat by using two statistical regression models in the NCP. Chen et al. (2019) set up a logistic function that related a drought hazard index to the yield loss rate for maize in China.

As can be seen from the above literature studies, the quantitative relationship between yield losses and drought can described by statistical models, regardless of whether the yield loss data are derived from crop model simulations or observational data. The most popular mathematical models employ parametric regression equations (such as linear regressions, quadratic polynomials, power functions and logistic functions), which give explicit expressions of the functional relationships. However, in recent years, with the development of data mining technology, scientists have proposed some new non-parametric statistical models. The random forest (RF) algorithm is one such method, which has the advantages of good stability, high prediction accuracy and it does not easy produce over-fitting (Breiman, 2001). The RF algorithm has been applied to flood risk analyses (Wang et al., 2015; Zhao et al., 2018), floodplain mapping (Woznicki et al., 2019), landslide susceptibility assessments (Chen et al., 2017; Goetz et al., 2015; Trigila et al., 2015; Youssef et al., 2016), drought monitoring (Park et al., 2016), forecasting (Chen et al., 2012; Seibert et al., 2017) and risk evaluation (Deng et al., 2018). However, little research has been done regarding drought disaster loss model construction using RFs. Therefore, in this paper we attempt to quantitatively analyze the relationship between yield losses and drought using both a traditional regression model and the newly developed RF algorithm. Our specific aims are: 1) to compare the performance of the regression model with the RF algorithm in terms of quantifying the relationship between yield loss and drought, and 2) look for the most influential drought index variables.

2. Study area and data

2.1. Study area

Henan Province, bounded between 31°23′–36°22′N and 110°21′–116°39′E, is located in the middle-eastern part of China and the middle and lower reaches of the Yellow River. Its climate is mainly affected by monsoons. The average annual precipitation in the province is about 500–900 mm. The spatial and temporal distribution of precipitation is not balanced, where 60% of the annual precipitation is concentrated between July–September, and interannual variations in precipitation are great. The average annual temperature of the whole province is generally between 12 and 16 °C, which meets the growth and developmental needs of two-year three-cropping and one-year two-cropping farming practices. Henan Province is a major grain province in China. To date, the grain yield of winter wheat has ranked first in China for 12 consecutive years, making outstanding contributions to ensuring the food security in China. Drought has a great impact on the winter wheat yield in Henan Province, especially in the rain-fed areas, where grain crops rely on natural precipitation as the main water source. There are 174 counties and municipal districts in Henan Province, of which 106 counties have historical yield data of winter wheat grain. Therefore, we chose these 106 counties as our study area and winter wheat as the research object (Fig. 1).

2.2. Data

The data used in this study include: 1) The sown area and production of winter wheat from 1990 to 2015 in each county from the Statistical Yearbook of Henan Province. These data were mainly used to calculate the yield per unit area of wheat in each county. 2) The effective irrigated area of counties and districts from the Statistical Yearbook of Henan Province in 2000 were used to divide all the counties into two parts: counties dominated by rain-fed crops and counties dominated by irrigated crops. 3) A land cover land use map of China in 2000 from the Resource and Environment Science Data Center of the Chinese Academy of Sciences (http://www.resdc.cn). The map was used to extract the spatial distribution of the cultivated land in Henan Province in 2000 and make a farmland mask. 4) The SPEIbase v2.5 data set from the Global SPEI database (http://sac.csic.es/spei/database.html). This data set provides monthly SPEI values (whose time scales range from 1 to 48 months) with a resolution of 0.5° from 1901 to 2015. Based on this data set, we extracted SPIE values at timescales of 1–3 months during the growth period (from October to May of the following year) of winter wheat in Henan Province from 1990 to 2015. Therefore, there were totally 24 SPEI variables. These variables were recorded as $SPEI_{i\text{-}j}$, where i refers to the month (ranging from October to May of the following year) and j refers to the monthly time scale (ranging from 1 to 3). For example, $SPEI_{Oct\text{-}1}$ is the one-month timescale SPIE value in October. The combined data from these four sources were used to construct the wheat drought loss models.

3. Study method

A technical flowchart of the method used in our study is shown in

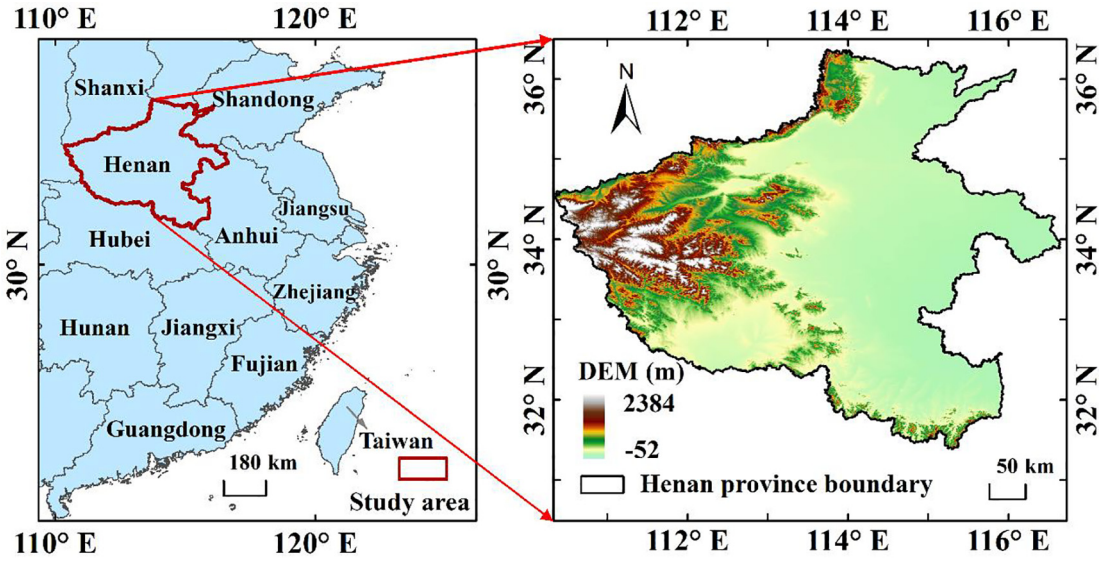

Fig. 1. Location of the study area.

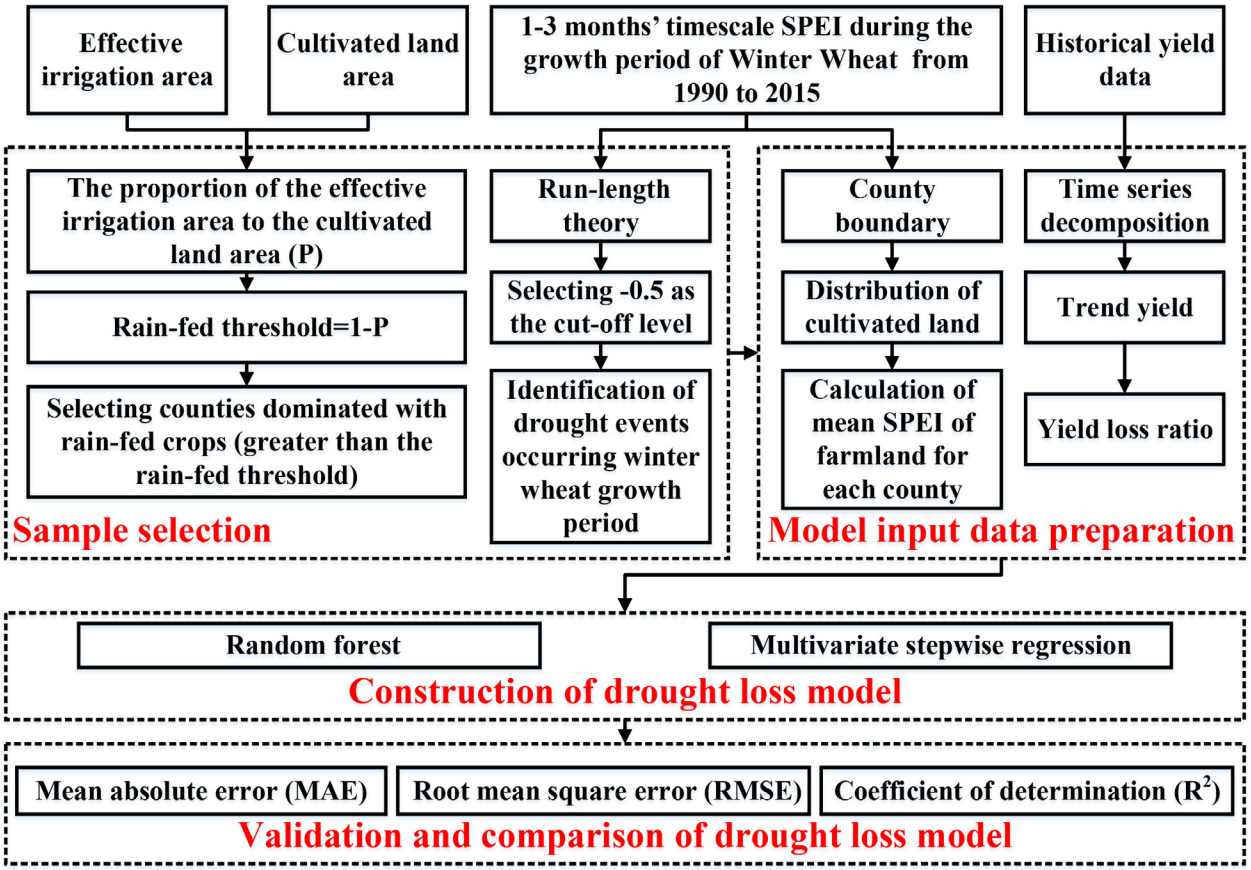

Fig. 2. Technical flowchart.

Fig. 2. The basic steps include: 1) Setting a threshold to divide all counties into two parts: the counties dominated by rain-fed crops and the counties dominated by irrigated crops according to the proportion of the effective irrigation area to the cultivated land area. 2) Calculating the mean SPEI value at different time scales in the cultivated land county-by-county in Henan Province. 3) Decomposing the historical wheat yield into trend yields and climatic yields, and calculate the yield loss ratios. 4) Select drought events affecting winter wheat growth as samples for the modeling based on run-length theory and the winter wheat growth calendar in Henan Province. 5) Quantitatively analyze the relationship between the yield losses and droughts using both the multivariate stepwise regression (MSR) and the RF models.

3.1. Extraction of counties dominated by rain-fed crops

Because counties dominated by irrigated crops are less affected by droughts, and winter wheat yields in rain-fed areas are more likely to be reduced by climate change, this paper only considers counties

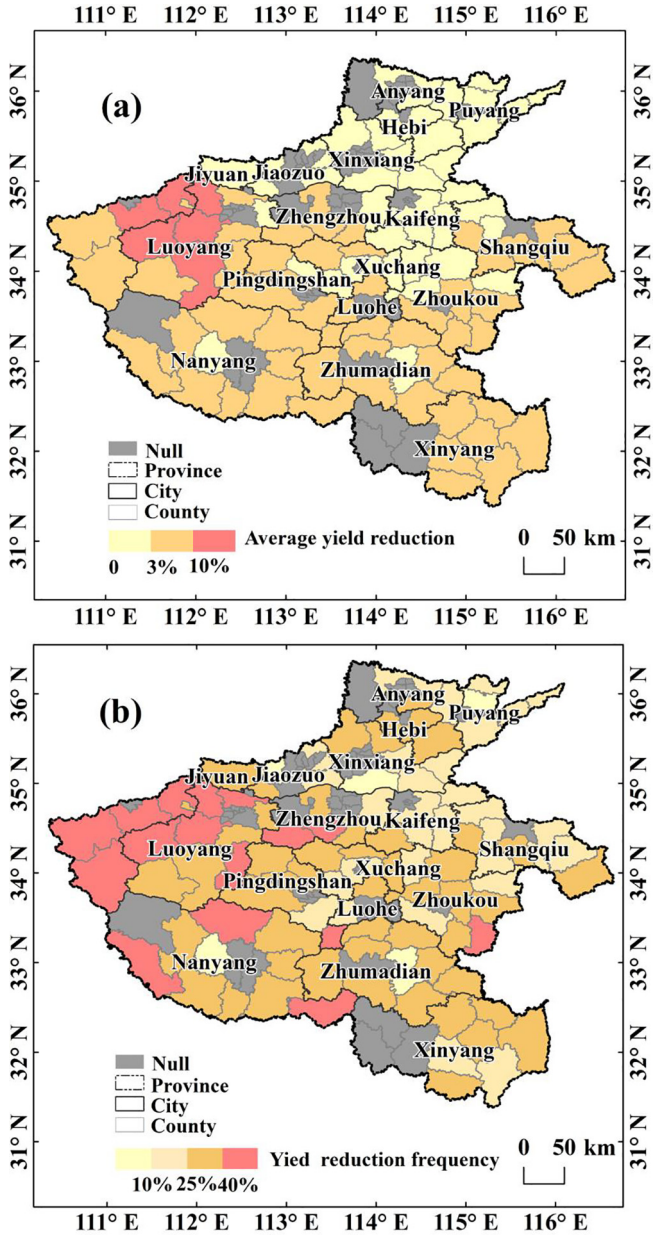

Fig. 3. (a) Average yield reduction, and (b) yield reduction frequency of wheat in the Henan counties.

dominated by rain-fed crops. We calculated the proportion of the effective irrigation area to the cultivated land area (defined as p) county-by-county, and divided all counties in Henan Province into the two types by setting a rain-fed threshold (defined as 1-p, referring to the proportion of the rain-fed area to the cultivated land area): counties dominated by rain-fed crops (greater than the rain-fed threshold) and counties dominated by irrigated crops (less than the rain-fed threshold). In order to reduce the impact of the threshold setting on the results, four different thresholds were considered: 30%, 40%, 50% and 60%. Under different rain-fed thresholds, the selected counties are different. That is to say, the samples involved in the modeling are different. When the threshold was set to 30%, 40%, 50% and 60%, the number of counties dominated by rain-fed crops was 49, 36, 25 and 16, respectively.

3.2. Calculation of the drought index at different time scales at the county level

We resampled the SPEI raster data with 0.5-degree resolution into 1-km resolution, extracted the SPEI values of farmland based on the land cover land use map of Henan Province and calculated the mean SPEI values of the farmland for each county. We repeated the above process for 24 SPEI variables ($SPEI_{Oct-1}$, $SPEI_{Oct-2}$, $SPEI_{Oct-3}$, …, $SPEI_{May-3}$). Therefore, for each county, we obtained 24 mean values of SPEI variables and used them as the model input.

3.3. Calculation of the yield loss ratio of winter wheat

Grain yield is affected by many factors, including natural factors and non-natural factors. Climate is the most important natural factor and it always fluctuates. Non-natural factors such as cultivation techniques and field management improve over time. Therefore, the annual crop yield per unit area (Y) can be divided into two parts: climatic yield (Y_c) and the trend yield (Y_t) (Eq. (1)). Climatic yield is determined by short-time climate variations, while the trend yield is influenced by long-term factors, which is also called the technical yield. In this paper, the Gompertz curve (Eq. (2)) was used to fit the trend yield of winter wheat in Henan Province:

$$Y = Y_t + Y_c \tag{1}$$

$$Y_t = Le^{-ae^{-bt}} (a > 0, b > 0) \tag{2}$$

here t denotes the number of years, where the initial value of 1 represents the year 1990, and a and b are two parameters to be estimated.

When there is autocorrelation between the residual series of the actual yield and the trend yield fitted by the Gompertz curve, the classical hypothesis of stochastic perturbations in econometrics is not

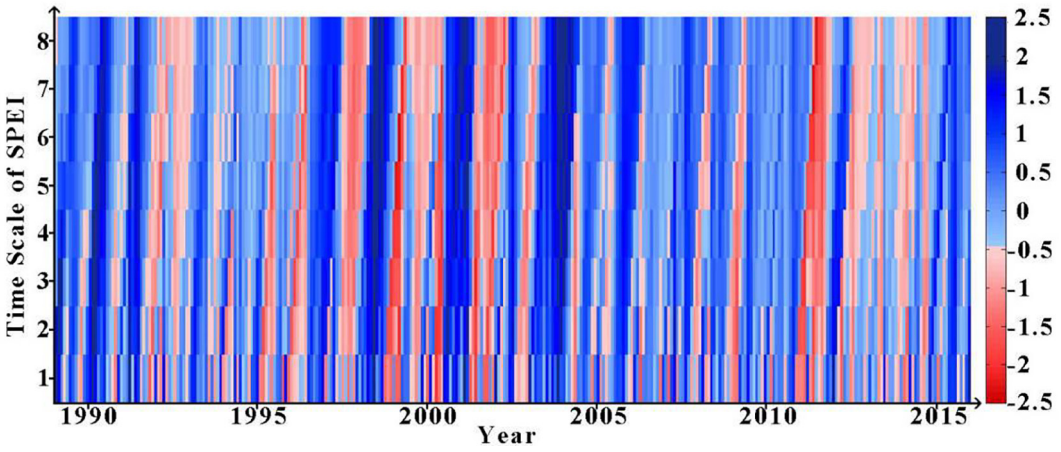

Fig. 4. SPEI intensity variation at different time scales from 1990 to 2015.

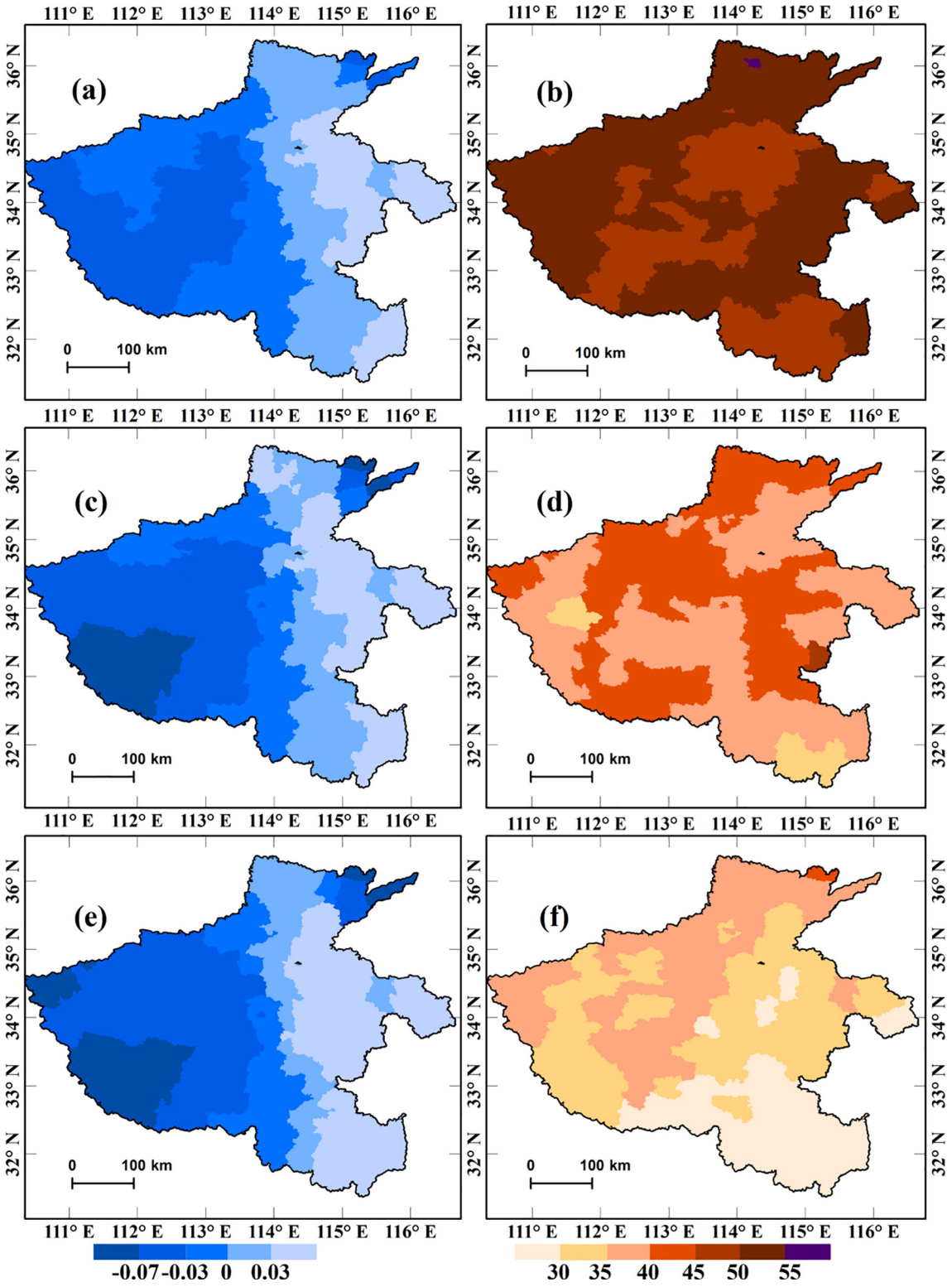

Fig. 5. The mean SPEI value and the number of droughts identified based on the SPEI at (a–b) the one-month time scale, (c– d) the two-month time scale, and (e–f) the three-month time scale in each 0.5° grid in Henan Province during 1990–2015.

satisfied, which will lead to a distortion of the fitted model results. For this situation, we further used the autoregressive moving average model (ARMA) to simulate the residual sequence, and we obtained the final trend yield (u_t) by adding the residual prediction results from ARMA with the trend yield obtained from the Gompertz curve:

$$u_t = Y_t + uc \tag{3}$$

$$\varphi(B)uc = \theta(B)\varepsilon_t \tag{4}$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p \tag{5}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \tag{6}$$

where uc is the trend correction value of the Gompertz fitting residuals, B is a lag operator and ε_t is a white noise sequence. Assuming that the

Table 1
The optimum parameters of the RF models.

Rain-fed Threshold (1-p)	N	M
30%	40	3
40%	25	4
50%	45	10
60%	55	6

year in which the actual yield did not reach the trend yield during the study period is a year of climate yield reduction, the drought yield reduction rate (rt) was calculated using Eq. (7), which ranges from 0 to 1. In agrometeorological research, a year where the yield reduction less than −3% is considered to be a "disaster", otherwise the loss is considered as a normal fluctuation (He et al., 2014).

$$rt = \max\{0, (ut - Y)/u_t\} \quad (7)$$

3.4. Identification of drought events occurring the winter wheat growth period

Climatic yield variations are not only influenced by drought but also by other climate disasters such as floods and freezing weather. Here, our aim is to identify drought-prone years to participate in the modelling process. Therefore, based on the SPEI values (of one- to three-month time scales), we used run-length theory to identify drought events in Henan Province. The specific procedure is as follows: First, referring to the classification standard of SPEI, we selected −0.5 as the cut-off level to set the threshold, recorded any negative runs which were continuously less than the threshold in one or more SPEI sequences, and then calculated the number of drought occurrences. For a specific year i, if at least one drought event was identified during the growth period of winter wheat (from October of the year i-1 to May of the year i), it was used in the modeling.

3.5. Construction of the wheat drought loss model

We first chose the sample data to participate in the modeling. Samples involved in the modeling needed to satisfy four conditions: 1) Sample data had to come from counties dominated by rain-fed crops, 2) drought occurred during the wheat growing season, 3) the climate yield was negative, and 4) the climate yield change was more than 3%. After identifying the relevant data under different rain-fed thresholds (30%, 40%, 50% and 60%), we used the yield loss ratio as the dependent variable and the 24 SPEI variables as independent variables to build the drought loss models using both the MSR and RF techniques.

The idea of MSR modelling is to screen the modeling factors according to the importance of the independent variables on the dependent variables and the correlations between the independent variables. The number of independent variables was increased one-by-one and any unimportant variable was eliminated. The idea of the RF algorithm is to generate N sets of new training samples based on sampling with the replacement method, where the unsampled samples constitute out-of-bag data (OOB). For each training set, a regression tree was grown. At each node of the tree, m variables were selected from the 24 independent variables. According to the principle of minimum node impurity, the tree branches grew from the m variables without pruning. The above steps were repeated n times to generate N regression trees to form a RF. In the process of RF construction, the number of regression trees N and the number of tree nodes M in the RF are important parameters affecting the prediction ability of the RF model, and m should be less than the number of variables in the model. The optimal parameters of the RF model were selected using the grid search method in which M ranged from 1 to 24 with a step size of 1, and N ranged from 10 to 400 with a step size of 5. By comparing the OOB errors of the models with different combinations of parameters, the M and N values corresponding to the minimum OOB errors were defined as the optimal parameters.

Table 3
Accuracy evaluation of the RF and MSR models.

Rain-fed Threshold (1-p)	Data	RF			MSR		
		MAE	RMSE	R^2	MAE	RMSE	R^2
30%	Training	0.030	0.044	0.911	0.093	0.133	0.206
	Validation	0.086	0.128	0.255	0.096	0.137	0.155
40%	Training	0.032	0.046	0.923	0.104	0.146	0.221
	Validation	0.083	0.104	0.373	0.094	0.119	0.220
50%	Training	0.033	0.049	0.915	0.095	0.138	0.328
	Validation	0.087	0.126	0.510	0.114	0.175	0.107
60%	Training	0.032	0.047	0.925	0.094	0.137	0.380
	Validation	0.055	0.073	0.720	0.076	0.096	0.523

3.6. Model validation

The performance of each MSR and RF model was evaluated using the same procedure as described in the preceding section. Eighty percent of the complete dataset was randomly selected to calibrate each model, and the remaining data were used to validate the model. Three validation measurements were used: the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|r_i - R_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - R_i)^2} \quad (9)$$

Table 2
The MSR parameters.

Rain-fed Threshold(1-p)							
30%		40%		50%		60%	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
$SPEI_{Oct-1}$	0.020	$SPEI_{Oct-1}$	0.046	$SPEI_{Oct-1}$	0.028	$SPEI_{Oct-1}$	0.055
$SPEI_{Feb-1}$	−0.060	$SPEI_{Feb-1}$	−0.033	$SPEI_{Feb-1}$	−0.080	$SPEI_{Feb-1}$	−0.080
$SPEI_{Oct-3}$	−0.027	$SPEI_{Mar-1}$	0.049	$SPEI_{Oct-3}$	−0.067	$SPEI_{Dec-2}$	−0.033
$SPEI_{Mar-3}$	0.046	$SPEI_{Oct-3}$	−0.026	$SPEI_{Nov-3}$	0.075	$SPEI_{Mar-2}$	0.075
Intercept	0.161	Intercept	0.186	$SPEI_{Jan-3}$	−0.064	$SPEI_{Apr-2}$	−0.038
				$SPEI_{Mar-3}$	0.046	$SPEI_{Oct-3}$	−0.055
				$SPEI_{May-3}$	0.024	Intercept	0.147
				Intercept	0.193		

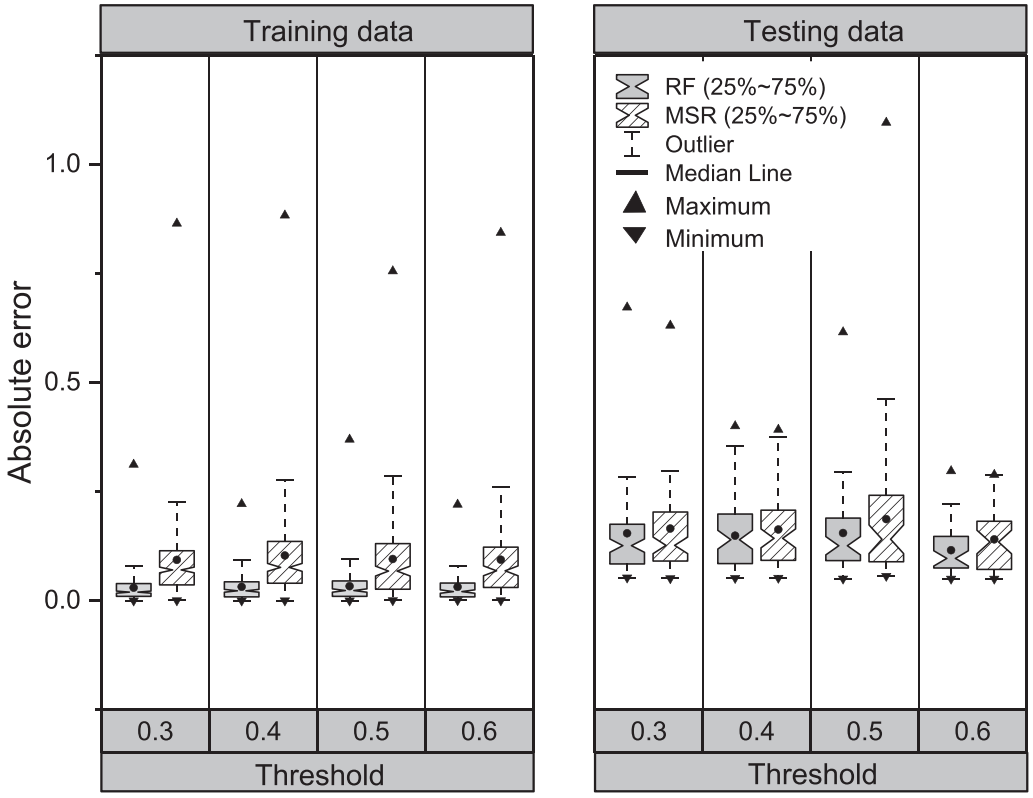

Fig. 6. Boxplots of the absolute errors of the RF and MSR models under different rain-fed thresholds.

$$R^2 = \left(\frac{\sqrt{\sum_{i=1}^{n} (r_i - \bar{r})(R_i - \bar{R})}}{\sqrt{\sum_{i=1}^{n} (r_i - \bar{r})^2} \sqrt{\sum_{i=1}^{n} (R_i - \bar{R})^2}} \right)^2 \quad (10)$$

where r_i and R_i refer to the true and estimated values of wheat yield reduction in county i, respectively, $\bar{r}$ and $\bar{R}$ are the mean values of the true and estimated values of wheat yield reduction, respectively, and n is the number of samples. When the threshold was set to 30%, 40%, 50% and 60%, n was 49, 36, 25, and 16, respectively. The MAE measures the average estimation bias, and the RMSE represents the sample standard deviation of the differences between the estimated and observed values. Results become increasingly more accurate as the MAE and RMSE approach 0 and R^2 approaches 1.

4. Results

4.1. Yield loss ratios

We calculated the average yield reduction and the frequency of yield reduction of winter wheat in Henan counties during 1990–2015 (the number of years that yielded a reduction occurred in all 26 years). The results are shown in Fig. 3. It should be noted that the yield reduction here refers to a yield loss rate greater than 3%. Generally speaking, the average reduction rate in the north of Henan Province is the lowest. The frequency of yield reduction is higher in the west than in the east, and higher in the south than in the north. The frequency of yield reduction in most counties is more than 25%. Counties with average reduction rates of more than 10% are concentrated in Sanmenxia City and Luoyang City in western Henan Province and they are also high-incidence areas of drought. Moreover, the average reduction rate and the frequency of yield reduction are relatively high in Nanyang, Zhumadian, Xinyang, Pingdingshan, and Zhoukou, where the average reduction rate of most counties is over 3% and the yield reduction frequency is over 25%.

4.2. SPEI variation

Fig. 4 shows the intensity changes of the SPEI at different time scales in Henan Province from 1990 to 2015. The SPEI at the one-month scale mainly shows monthly changes, while the SPEI at the three-month scale mainly shows seasonal changes. The SPEI at the eight-month scale represents the change of the whole growth period of winter wheat in Henan Province. From the figure we can see that the SPEI values show consistent trends, where the frequency of drought decreases with an increase of the SPEI time scale. The drought events identified in 1992, 1997, 1999, 2000, 2002 and 2011–2014 are more serious; these correctly identified results are consistent with the actual occurrence of drought in Henan Province. The drought in 2014 lasted for a long time, while the droughts in 2000 and 2011 had high intensities.

We calculated the mean SPEI values at different time scales for each 0.5-degree grid in Henan Province from 1990 to 2015 and the number of drought events identified in each grid based on the SPEIs at different time scales. The results are shown in Fig. 5. The mean SPEI values at different time scales are consistent in terms of their spatial distributions. The drought intensity in western Henan Province is greater than that in eastern Henan Province. Overall, the drought intensity and yield reductions are spatially consistent (Figs. 3a and 5a, c, and e). As for drought frequency, similar to Fig. 4, the number of identified drought episodes decreases with an increase of the SPEI time scale. On the whole, the frequency of drought in Henan Province is very high. In most counties, drought occurred more than once a year during the growing season of winter wheat between 1990 and 2015.

4.3. Fitted models

The optimum parameters of the RF algorithm were determined using the grid search method. The results are shown in Table 1, the MSR

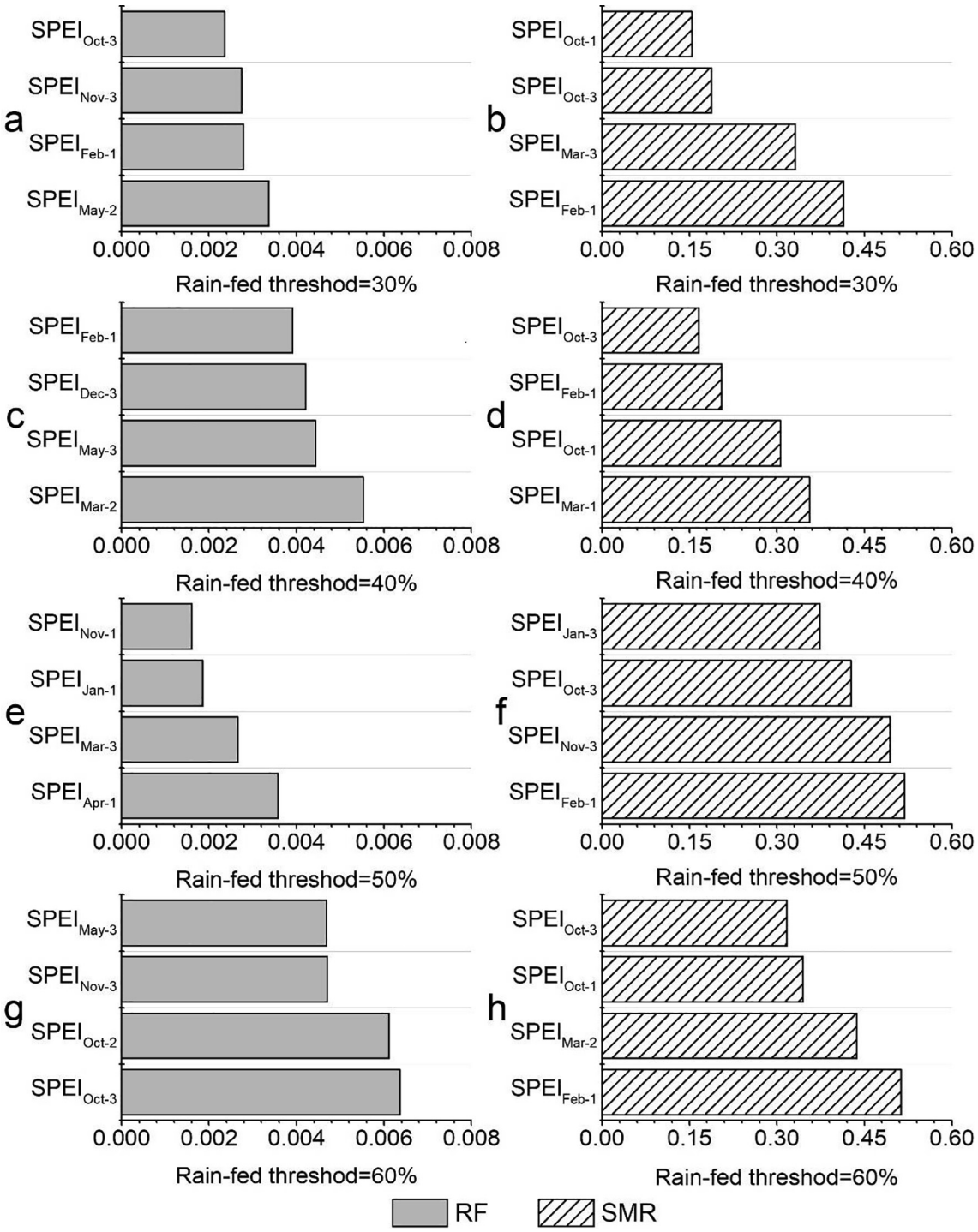

Fig. 7. Top four most important SPEI variables recognized by the RF and MSR models under the different rain-fed thresholds.

equations are shown in Tables 2, and 3 shows the accuracy evaluation of the RF and MSR models. Regardless of the rain-fed threshold, the R^2 values of the RF models are above 0.9 for the training data. For the validation data, the R^2 values of the RF models range from 0.255 to 0.720 and they increase with an increase of the rain-fed threshold. The MAE and RMSE values of the validation data are the smallest when the rain-fed threshold is 60%. The R^2 values of the MSR models increase with an increase of the rain-fed threshold. The highest R^2 value of the MSR model is 0.523, which is lower than any of those of the RF models, and the MAE and RMSE values of the MSR model are higher than those of the RF models regardless of the rain-fed threshold. When the rain-fed threshold is equal to 60%, for the training data, the MAE and RMSE values of the MSR model are 2.94 times and 2.91 times larger than their respective values for the RF model. For the validation data, the MAE and RMSE values of the MSR model are 1.38 times and 1.31 times larger than those of the RF model. Fig. 6 shows a boxplot of the absolute errors of the RF and MSR models under different rain-fed thresholds. The maximum, median and mean values, and the range of the absolute errors of the MSR models are all higher than those of the RF model for both the training and testing data. Therefore, the fitting results of MSR models are inferior to those of the RF models.

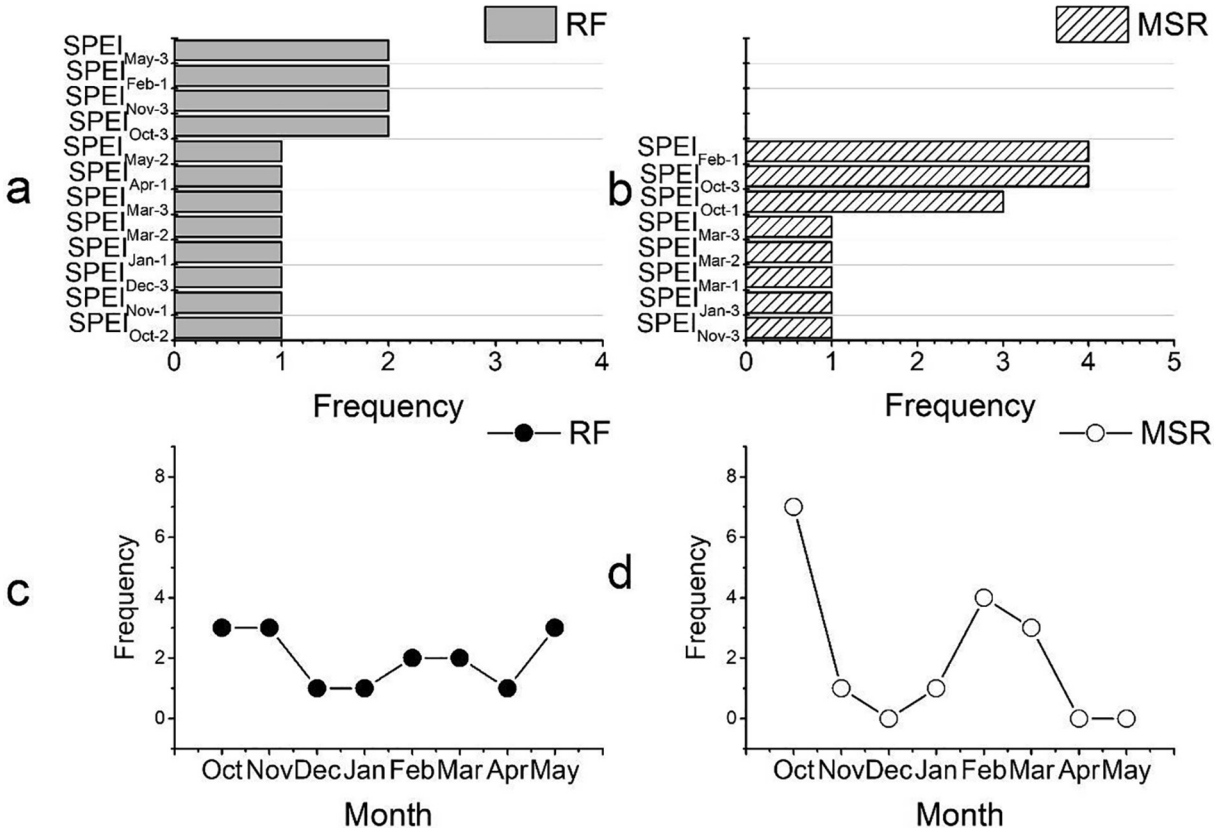

Fig. 8. The top four most important SPEI variables and their cumulative recognition numbers.

5. Discussion

5.1. Importance of different SPEI variables to modeling

There were 24 SPEI variables involved in the modeling; however, not all of them were useful. Under different rain-fed thresholds, the number of SPEI variables involved in the RF models ranged from 3 to 10 (Table 1), and those in the MSR models ranged from 4 to 7 (Table 2). Fig. 6 shows the top four most important SPEI variables recognized by the RF and MSR models under the different rain-fed thresholds. In the RF models, we estimated the importance of the variables with the permutation importance method, in which the importance of a variable was calculated as the difference between the baseline and the drop in overall accuracy after its removal from the model. In the MSR models, the importance of a variable was expressed by the absolute value of the normalized coefficient of the multivariate linear regression. From Fig. 7, we can see that the importance of the SPEI variables assessed with the different methods is different, and the importance of the SPEI variables assessed by the same method under different threshold conditions is also different.

For each of the top four most important SPEI variables under each threshold condition, we counted the cumulative number of times that it was identified as one of the top four. The results are shown in Fig. 8a and b, in which the frequency ranges from 1 to 4. For example, for $SPEI_{Oct\text{-}2}$ in the RF modeling, $SPEI_{Oct\text{-}2}$ was identified as one of the top four most important variables only when the threshold was 60%. So, the frequency of $SPEI_{Oct\text{-}2}$ in Fig. 8a is one. A variable with a frequency of four (such as $SPEI_{Feb\text{-}1}$ and $SPEI_{Oct\text{-}3}$ in the MSR models) indicates it was always recognized as one of the top four most important variables under the four different thresholds (Fig. 8b). $SPEI_{May\text{-}3}$, $SPEI_{Feb\text{-}1}$, $SPEI_{Nov\text{-}3}$ and $SPEI_{Oct\text{-}13}$ were recognized as the most important variables in the RF models, while $SPEI_{Feb\text{-}1}$, $SPEI_{Oct\text{-}3}$ and $SPEI_{Oct\text{-}1}$ were recognized as the most important variables in the MSR models.

Additionally, for each month of the growing season (October to next May), we counted the cumulative number of times that the multi-time scale SPEI values of each month were identified as one of the top four most important SPEI variables (Fig. 8c and d). For example, the number of times that $SPEI_{Oct\text{-}2}$ and $SPEI_{Oct\text{-}3}$ in the RF models were recognized as one of the top four most important SPEI variables was one and two, respectively (Fig. 8a). Therefore, the total number of times that the SPEI values in October were recognized as one of the most important variables was three in the RF models (Fig. 8c).

For Henan province, winter wheat is in the seedling stage during October–November, in the over-wintering stage during December–February, and in the late growth stage during April–May. Soil moisture (before sowing/planting) in Henan Province is mainly formed by precipitation accumulating in the soil from July to early October of that year. Adequate soil moisture (before sowing/planting) can promote root systems that effectively absorb soil moisture and nutrients, improve soil water use efficiency, reduce the probability of water deficits during the jointing stage, and ensure crop yields. Overwintering precipitation is conducive to regulating fertilizer by water, improving fertilizer efficiency, promoting tillering and secondary root growth, and meeting the water requirements of young panicle development. Water consumption of winter wheat from the heading to maturity stages (i.e. the filling stage) is the largest. If water deficit occurs during these stages, the 1000-grain weight will be reduced, resulting in yield reductions. Both the RF and MSR models recognized SPEIs in October (sowing/planting stage) and February (overwintering stage) as important variables (Fig. 8 c and d). However, the MSR models failed to recognize the importance of SPEIs during April–May (filling stage).

5.2. Comparison of the regression and RF models

There are multiple collinearities among the SPEI variables (Fig. 9).

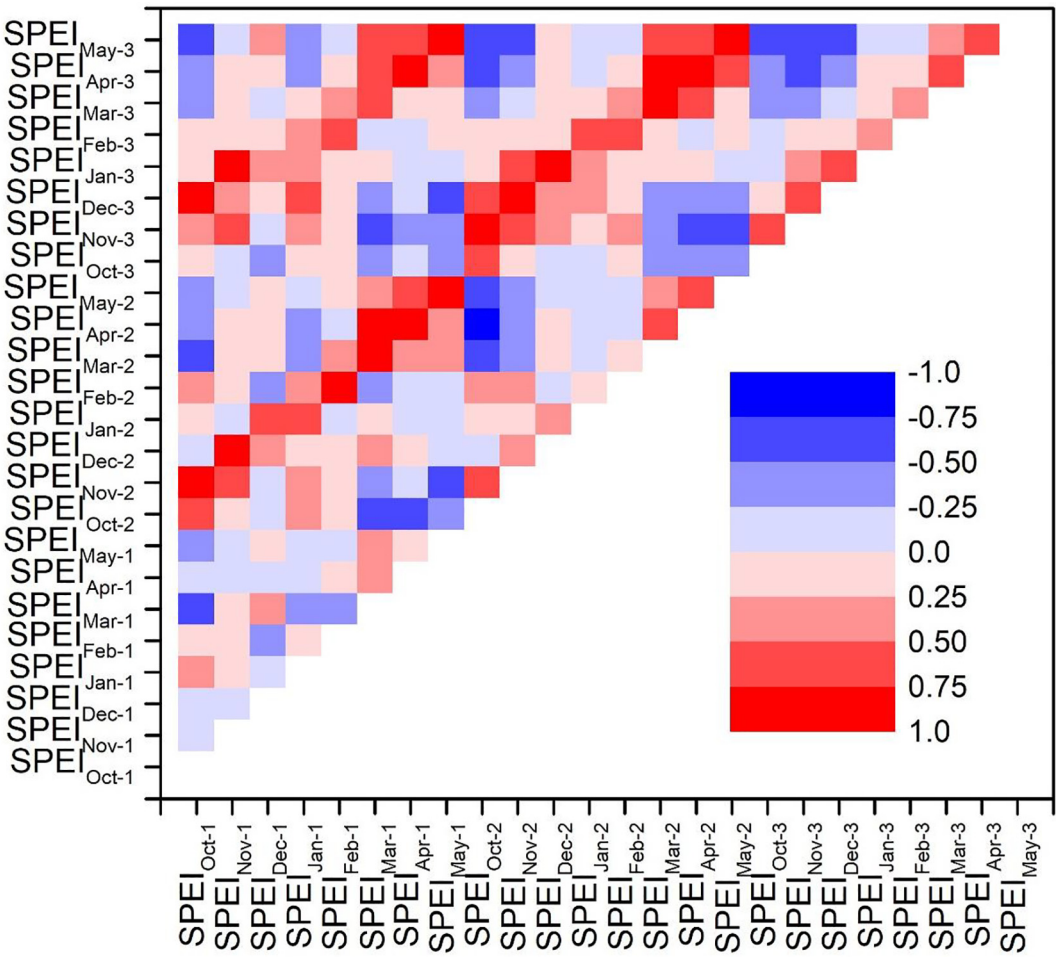

Fig. 9. Correlation coefficients between the SPEI variables (Rain-fed threshold = 60%).

Multivariate stepwise regression is a common method for eliminating multiple collinearity and selecting the optimal regression equation. Random forest is a nonlinear algorithm that builds a model without considering the impact of collinearities among the SPEI variables. Like other machine learning techniques, RF has the advantage of automatically detecting complex interactions between independent variables. It can measure the importance of each SPEI variable in the drought loss model.

The comparison between the results obtained from the MSR and RF models show that the RF models outperformed the MSR models in all tests based on the three validation measurements. However, the MSR models have better explicability than the RF models. Although interpretability may be a useful criterion for choosing a suitable model in real applications, predictive accuracy is still the most important criteria for drought loss assessment. Besides, the MSR models did not recognize the importance of the SPEI variables at the filling stage as the RF models did. Therefore, we give priority to constructing drought loss models with the RF algorithm.

5.3. Limitations of the study

The limitations of this study are mainly reflected in two aspects:

First, this paper focused on the drought loss models of rain-fed wheat. The selection of the rain-fed samples will inevitably affect the construction of the model. Traditional wheat yield data in China are reported by administrative units, and they do not distinguish between rain-fed and irrigated wheat yields. Statistical data of effective irrigation areas are available for all counties of Henan Province. Thus, we set different thresholds to distinguish between counties dominated by rain-fed crops and those dominated by irrigated crops. A low rain-fed threshold may cause a bad model fitting degree because of the impact of the irrigation samples, while a high threshold may lead to insufficient samples sizes, also causing a decrease of the model's accuracy. The results in Section 4.3 show that the accuracy of the model increases with an increase of the proportion of rain-fed areas. This also indicates that the method used in this study cannot be used to establish a high-precision loss model in irrigated areas. This might be because SPEI cannot describe agricultural drought in irrigation areas very well even though it takes into account both precipitation and potential evapotranspiration when determining drought. Agricultural drought is the result of a combination of factors, including the atmosphere, crops, soil, hydrology and human activities (such as irrigation). In contrast, SPEI mainly reflects climate conditions.

Second, we used run-length theory to identify drought events. For a specific year *i*, if at least one drought event was identified during the growth period of winter wheat, it was used as a sample in the modeling. We did not distinguish between the occurrence time, intensity and duration of each drought event. However, the response of winter wheat to drought stress at different growth stages is inconsistent (Meinke and Stone, 2005). For example, a certain degree of drought during the overwintering period might promote yields, while drought during the flowering and milk-ripening stages can enhance leaf senescence and reduce photosynthesis, reducing the final yields (Yang et al., 2001; Yu et al., 2019). The effects of different drought intensities and durations on wheat yield are also different. For example, mild drought during the milk-ripening stage and moderate drought during the flowering stage

can both have a significant impact on the yields (Yu et al., 2019). When using run-length theory to identify drought events, only a single threshold (-0.5) was set to identify the occurrence of drought events. In the future, multi-threshold can be set dynamically, and drought events can be identified according to different drought grades when using SPEI. Moreover, the combined distribution of drought duration and drought intensity of identified drought events can be used to screen the samples.

5.4. Contributions of the study

There are three main contributions of this paper.

First, most parts of the world just like China, where traditional crop yield data are reported by administrative units, and do not distinguish between rain-fed and irrigated crop yields. Our study shows that the rain-fed threshold has great impact on the modeling accuracy. This suggests that researchers need to eliminate the impact of irrigation and select rain fed areas as research areas when building drought loss models based on historical yield statistics. Crop model simulation is a more effective method to study the effect of drought on crop yield under different irrigation conditions.

Secondly, this paper proves that the nonparametric stochastic forest modeling method is better than the traditional linear stepwise regression modeling method. The advantages of random forest are reflected in two aspects: first, SPEIs in different months and time scales are highly correlated, but the contribution to modeling varies greatly. Choosing appropriate SPEI variables to participate in modeling is essential to ensure the accuracy of the model. SPEI selected by random forest has stronger physical significance and is not sensitive to variables with high correlation. Secondly, the accuracy of RF is higher than that of MSR.

Third, little research has been done regarding drought disaster loss model construction using RFs. This study enriches the case of drought loss evaluation model based on nonparametric model, especially RF, and it also provides a reference for other disaster loss evaluation models building.

6. Conclusions

This study built wheat drought loss models using both the RF and MSR methods and compared their performances. Our results show that RF models are better at estimating wheat yield losses. The RF models outperformed the MSR models under four rain-fed threshold conditions in terms of three validation measurements (MAE, RMSE and R^2). When the rain-fed threshold was 30%, the validation results show that the MAE, RMSE and R^2 values from the RF model were 0.086, 0.128 and 0.255, respectively, while the MAE, RMSE and R^2 values from the MSR model were 0.096, 0.137 and 0.155, respectively. When the rain-fed threshold was 40%, the validation results show that the MAE, RMSE and R^2 values from the RF model were 0.083, 0.104 and 0.373, respectively, while the MAE, RMSE and R^2 values from the MSR model were 0.094, 0.119 and 0.220, respectively. When the rain-fed threshold was 40%, the validation results show that the MAE, RMSE and R^2 values from the RF model were 0.087, 0.126 and 0.510, respectively, while the MAE, RMSE and R2 values from the MSR model were 0.114, 0.175 and 0.107, respectively. When the rain-fed threshold was 60%, the validation results show that the MAE, RMSE and R^2 values from the RF model were 0.055, 0.073 and 0.720, respectively, while the MAE, RMSE and R^2 values from the MSR model were 0.076, 0.096 and 0.523, respectively. The rain-threshold had a big impact on the modeling results. The accuracy of the models built by both RF and MSR increased with an increase of the rain-fed threshold. Our results prove that the stochastic forest model is an effective method for establishing crop drought disaster loss models.

CRediT authorship contribution statement

Xiufang Zhu: Conceptualization, Funding acquisition, Writing - original draft, Writing - review & editing. **Chenyao Hou:** Data curation, Methodology, Project administration, Writing - review & editing. **Kun Xu:** Formal analysis, Supervision. **Ying Liu:** Validation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Key R&D Program of China (Grant No. 2019YFA0606900), the National Natural Science Foundation for Distinguished Young Scholars of China (Grant No. 41401479), and the Project Supported by State Key Laboratory of Earth Surface Processes and Resource Ecology (Grant No. 2017-FX-01(1)).

References

Aboelghar, M., Ali, A.-R., Arafat, S., 2014. Spectral wheat yield prediction modeling using SPOT satellite imagery and leaf area index. Arabian J. Geosci. 7, 465–474. https://doi.org/10.1007/s12517-012-0772-6.

Ali, R., Kuriqi, A., Abubaker, S., Kisi, O., 2019a. Hydrologic Alteration at the Upper and Middle Part of the Yangtze River, China: Towards Sustainable Water Resource Management Under Increasing Water Exploitation. Sustainability 11, 5176. https://doi.org/10.3390/su11195176.

Ali, R., Kuriqi, A., Abubaker, S., Kisi, O., 2019b. Long-Term Trends and Seasonality Detection of the Observed Flow in Yangtze River Using Mann-Kendall and Sen's Innovative Trend Method. Water 11, 1855. https://doi.org/10.3390/w11091855.

Bachmair, S., Svensson, C., Hannaford, J., Barker, L.J., Stahl, K., 2016. A quantitative analysis to objectively appraise drought indicators and model drought impacts. Hydrol. Earth Syst. Sci. 20, 2589–2609. https://doi.org/10.5194/hess-20-2589-2016.

Bachmair, S., Svensson, C., Prosdocimi, 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-2017-187.

Bolten, J.D., Crow, W.T., Zhan, X., Jackson, T.J., Reynolds, C.A., 2010. Evaluating the Utility of Remotely Sensed Soil Moisture Retrievals for Operational Agricultural Drought Monitoring. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 57–66. https://doi.org/10.1109/jstars.2009.2037163.

Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential uses and limitations of crop models. Agron. J. 88, 704–716. https://doi.org/10.2134/agronj1996.00021962008800050005x.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. https://doi.org/10.1023/a:1010933404324.

Cancelliere, A., Di Mauro, G., Bonaccorso, B., Rossi, G., 2007. Drought forecasting using the standardized precipitation index. Water Resour. Manage. 21, 801–819. https://doi.org/10.1007/s11269-006-9062-y.

Chen, F., Jia, H., Pan, D., 2019. Risk Assessment of Maize Drought in China Based on Physical Vulnerability. J. Food Qual. 2019. https://doi.org/10.1155/2019/9392769.

Chen, J., Li, M., Wang, W., 2012. Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast. Math. Probl. Eng. 2012, 1–12. https://doi.org/10.1155/2012/915053.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151, 147–160. https://doi.org/10.1016/j.catena.2016.11.032.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nat. Clim. Change 3, 52–58. https://doi.org/10.1038/nclimate1633.

Deng, M., Chen, J., Huang, J., Niu, W., 2018. Agricultural Drought Risk Evaluation Based on an Optimized Comprehensive Index System. Sustainability 10, 3465. https://doi.org/10.3390/su10103465.

Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007.

He, Y., Tang, Y., Zhang, J., 2014. Evaluation method for effects of drought disaster on yields of maize in Southwest China. Transactions of the Chinese Society of Agricultural Engineering 30, 185–191. https://doi.org/10.3969/j.issn.1002-6819.2014.23.023.

Heim, R.R., 2002. A review of twentieth-century drought indices used in the United States. Bull. Am. Meteorol. Soc. 83, 1149–1165. https://doi.org/10.1175/1520-0477-83.8.1149.

Jia, H., Wang, J., Cao, C., Pan, D., Shi, P., 2012. Maize drought disaster risk assessment of China based on EPIC model. Int. J. Dig. Earth 5, 488–515. https://doi.org/10.1080/17538947.2011.590535.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A.,

Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235–265. https://doi.org/10.1016/s1161-0301(02)00107-7.

Kim, T.W., Valdes, J.B., 2003. Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J. Hydrol. Eng. 8, 319–328. https://doi.org/10.1061/(asce)1084-0699(2003)8:6(319).

Liu, X., Pan, Y., Zhu, X., Yang, T., Bai, J., Sun, Z., 2018a. Drought evolution and its impact on the crop yield in the North China. Plain. Journal of Hydrology 564, 984–996. https://doi.org/10.1016/j.jhydrol.2018.07.077.

Liu, X., Zhu, X., Pan, Y., Bai, J., Li, S., 2018b. Performance of different drought indices for agriculture drought in the North China Plain. Journal of Arid Land 10, 507–516. https://doi.org/10.1007/s40333-018-0005-2.

Meinke, H., Stone, R., 2005. Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Clim. Change 70 221–253. https://doi.org/ 10.1007/s10584-005-5948-6.

Ming, B., Guo, Y., Tao, H., Liu, G., Li, S., Wang, P., 2015. SPEIPM-based research on drought impact on maize yield in North China Plain. Journal of Integrative. Agriculture 14, 660–669. https://doi.org/10.1016/s2095-3119(14)60778-4.

Mishra, A.K., Desai, V., 2005. Drought forecasting using stochastic models. Stoch. Env. Res. Risk Assess. 19, 326–339. https://doi.org/10.1007/s00477-005-0238-4.

Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. Agric. For. Meteorol. 133, 69–88. https://doi.org/10.1016/j.agrformet.2005.07.012.

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. https://doi.org/10.1016/j.agrformet.2015.10.011.

Seibert, M., Merz, B., Apel, H., 2017. Seasonal forecasting of hydrological drought in the Limpopo Basin: a comparison of statistical methods. Hydrol. Earth Syst. Sci. 21, 1611–1629. https://doi.org/10.5194/hess-21-1611-2017.

Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Abou, A., Ali, A., Demuth, S., Ogallo, L., 2014. A drought monitoring and forecasting system for Sub-Sahara African water resources and food security. Bull. Am. Meteorol. Soc. 95, 861–882. https://doi.org/10.1175/bams-d-12-00124.1.

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop-The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. Agronomy Journal 101, 426–437. https://doi.org/10.2134/agronj2008.0139s.

Stockle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18, 289–307. https://doi.org/10.1016/s1161-0301(02)00109-0.

Su, B., Huang, J., Fischer, T., Wang, Y., Kundzewicz, Z.W., Zhai, J., Sun, H., Wang, A., Zeng, X., Wang, G., Tao, H., Gemmer, M., Li, X., Jiang, T., 2018. Drought losses in China might double between the 1.5 degrees C and 2.0 degrees C warming. Proc. Natl. Acad. Sci. U.S.A. 115, 10600–10605. https://doi.org/10.1073/pnas.1802129115.

Sun, Z., Zhu, X., Pan, Y., Zhang, J., Liu, X., 2018. Drought evaluation using the GRACE terrestrial water storage deficit over the Yangtze River Basin, China. Sci. Total Environ. 634, 727–738. https://doi.org/10.1016/j.scitotenv.2018.03.292.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., Stephens, S., 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83, 1181–1190. https://doi.org/10.1175/1520-0477-83.8.1181.

Trigila, A., Iadanza, C., Esposito, C., Scarascia-Mugnozza, G., 2015. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). Geomorphology 249, 119–136. https://doi.org/10.1016/j.geomorph.2015.06.001.

Vicente-Serrano, S.M., Begueria, S., Lopez-Moreno, J.I., 2010. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. J. Clim. 23, 1696–1718. https://doi.org/10.1175/2009jcli2909.1.

Vicente-Serrano, S.M., Begueria, S., Lorenzo-Lacruz, J., Julio Camarero, J., Lopez-Moreno, J.I., Azorin-Molina, C., Revuelto, J., Moran-Tejeda, E., Sanchez-Lorenzo, A., 2012. Performance of Drought Indices for Ecological, Agricultural, and Hydrological Applications. Earth Interact. 16. https://doi.org/10.1175/2012ei000434.1.

Wang, S., Mo, X., Hu, S., Liu, S., Liu, Z., 2018. Assessment of droughts and wheat yield loss on the North China Plain with an aggregate drought index (ADI) approach. Ecol. Ind. 87, 107–116. https://doi.org/10.1016/j.ecolind.2017.12.047.

Wang, Z., He, F., Fang, W., Liao, Y., 2013. Assessment of physical vulnerability to agricultural drought in China. Nat. Hazards 67, 645–657. https://doi.org/10.1007/s11069-013-0594-1.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. J. Hydrol. 527, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008.

Wei, Y., Jin, J., Jiang, S., Ning, S., Cui, Y., Zhou, Y., 2019. Simulated Assessment of Summer Maize Drought Loss Sensitivity in Huaibei Plain, China. Agronomy-Basel 9, 78. https://doi.org/10.3390/agronomy9020078.

Woznicki, S.A., Baynes, J., Panlasigui, S., Mehaffey, M., Neale, A., 2019. Development of a spatially complete floodplain map of the conterminous United States using random forest. Sci. Total Environ. 647, 942–953. https://doi.org/10.1016/j.scitotenv.2018.07.353.

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., 2018. Estimating the responses of winter wheat yields to moisture variations in the past 35 years in Jiangsu Province of China. PLoS ONE 13, e0191217. https://doi.org/10.1371/journal.pone.0191217.

Yang, J.C., Zhang, J.H., Wang, Z.Q., Zhu, Q.S., Wang, W., 2001. Remobilization of carbon reserves in response to water deficit during grain filling of rice. Field Crops Research 71, 47–55. https://doi.org/10.1016/S0378-4290(01)00147-2.

Yin, Y., Zhang, X., Lin, D., Yu, H., Wang, J.a., Shi, P., 2014. GEPIC-V-R model: A GIS-based tool for regional crop drought risk assessment. Agric. Water Manag. 144, 107–119. https://doi.org/10.1016/j.agwat.2014.05.017.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856. https://doi.org/10.1007/s10346-015-0614-1.

Yu, H., Zhang, Q., Sun, P., Song, C., 2019. Impacts of drought intensity and drought duration on winter wheat yield in five provinces of North China plain. Acta Geographica Sinica 74, 87–102. https://doi.org/10.11821/dlxb201901007.

Yue, Y., Li, J., Ye, X., Wang, Z., Zhu, A., Wang, J., 2015. An EPIC model-based vulnerability assessment of wheat subject to drought. Nat. Hazards 78, 1629–1652 https://doi.org/ 10.1007/s11069-015-1793-8.

Zargar, A., Sadiq, R., Naser, B., Khan, F.I., 2011. A review of drought indices. Environmental Reviews 19, 333–349. https://doi.org/10.1139/A11-013.

Zhang, J.Q., 2004. Risk assessment of drought disaster in the maize-growing region of Songliao Plain, China. Agric. Ecosyst. Environ. 102, 133–153. https://doi.org/10.1016/j.agee.2003.08.003.

Zhang, R., Chen, Z., Xu, L., Ou, C., 2019. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Sci. Total Environ. 665, 338–346. https://doi.org/10.1016/j.scitotenv.2019.01.431.

Zhao, G., Pang, B., Xu, Z., Yue, J., Tu, T., 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. Sci. Total Environ. 615, 1133–1142. https://doi.org/10.1016/j.scitotenv.2017.10.037.

Zhu, X., Pan, Y., Wang, J., Liu, Y., 2019. A Cuboid Model for Assessing Surface Soil Moisture. Remote Sens. 11, 3034. https://doi.org/10.3390/rs11243034.