RESEARCH

Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content

Juanjuan Zhang[1,3], Wen Zhang[1,2,3], Shuping Xiong[1,2,3], Zhaoxiang Song[4], Wenzhong Tian[5], Lei Shi[1,3] and Xinming Ma[1,2,3]*

Abstract

Background: The leaf water content estimation model is established by hyperspectral technology, which is crucial and provides technical reference for precision irrigation.

Methods: In this study, two consecutive years of field experiments (different irrigation times and seven wheat varieties) in 2018–2020 were performed to obtain the canopy spectra reflectance and leaf water content (LWC) data. The characteristic bands related to LWC were extracted from correlation coefficient method (CA) and x-Loading weight method (x-Lw). Five modeling methods, spectral index and four other methods (Partial Least-Squares Regression (PLSR), Random Forest Regression (RFR), Extreme Random Trees (ERT), and K-Nearest Neighbor (KNN)) based characteristic bands, were employed to construct LWC estimation models.

Results: The results showed that the canopy spectral reflectance increased with the increase of irrigation times, especially in the near-infrared band (750–1350 nm). The prediction accuracy of the newly developed differential spectral index DVI (R1185, R1307) was higher than that of the existing spectral index, with R^2 of 0.85 and R^2 of 0.78 for the calibration and validation, respectively. Due to a large amount of hyperspectral data, the correlation coefficient method (CA) and x-Loading weight (x-Lw) were used to select the water characteristic bands (100 and 28 characteristic bands, respectively) from the full spectrum. We found that the accuracy of the model based on the characteristic bands was not significantly lower than that of the full spectrum-based models. Among these models, the ERT- x-Lw model performed the best (R^2 and RMSE of 0.88 and 1.46; 0.84 and 1.62 for the calibration and validation, respectively). In addition, the accuracy of the LWC estimation model constructed by ERT-x-Lw was higher than that of DVI (R1185, R1307).

Conclusion: The two models based on ERT-x-Lw and DVI (R1185, R1307) can effectively predict wheat leaf water content. The results provide a technical reference and a basis for crop water monitoring and diagnosis under similar production conditions.

Keywords: Winter wheat, Leaf water content, Spectral index, Characteristic band, Modeling method, Inversion model

Background

Wheat is the main food crop in North China. Due to the imbalance between precipitation and water demand during the growing period, reasonable irrigation has become a necessary condition for a high yield of wheat [1]. The leaf is an essential component of plant canopy structure and the site where a substantial number of important

*Correspondence: xinmingma@126.com
[1] Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University, #63 Nongye Road, Zhengzhou 450002, Henan, People's Republic of China
Full list of author information is available at the end of the article

Zhang *et al. Plant Methods*   (2021) 17:34

Page 2 of 14

biochemical processes occur; thus, leaf water content is an important indicator that reflects the overall crop water status and indirectly indicates the input and output of soil water [2, 3]. Therefore, the leaf water content can be used as a reliable reference index for making feasible irrigation decisions [4]. Hyperspectral remote sensing technology has the advantages of being fast, economic, and nondestructive. It can be used to monitor the growth of crops by obtaining reflectance information. Hence, the development of a diagnostic model of water status by hyperspectral remote sensing technology is of substantial significance for precision irrigation and water-saving irrigation.

Currently, hyperspectral remote sensing technology is widely used in crop water monitoring scenarios. In previous studies, the relationship between wheat leaf water content and hyperspectral data has been analyzed, and the spectral index has been used to estimate leaf water content. For example, the leaf water content of rice, peanut, soybean, and wheat can be well predicted by the ratio of the water index to the normalized vegetation index $(WI(R_{900}/R_{970})/NDVI(R_{900} - R_{680})/(R_{900} + R_{680})$, where R is the wavelength of the original spectral reflectance) [5]. Zhao et al. [6] established the presence of a significant correlation between leaf water content and the normalized difference water index (NDWI), the simple ratio (SR), and the shortwave infrared perpendicular water stress index (SPSI). Moreover, Rapaport et al. [7] developed a water balance index $(WABI = (R_{1500} - R_{531})/(R_{1500} + R_{531}))$ for monitoring the plant water status in grapevine under field conditions. In recent years, machine learning methods have been increasingly utilized for modeling and analysis of wheat growth information and water index. In this respect, Zhang et al. [8] used reflectance data in the range of 859–1640 nm, partial least squares (PLSR), artificial neural network (ANN), and support vector machine (SVN) algorithms to construct models for estimating wheat leaf water content and the equivalent water thickness content, and compared the prediction accuracy with the vegetation index model, employed as a reference model. Additionally, Das et al. [9] established a prediction model of wheat relative water content using a PLSR algorithm, multiple linear regression, SVN and random forest regression (RFR), based on hyperspectral and relative water content data of 10 wheat varieties in different periods. When full-band data are used for modeling, some issues arise, such as interference of redundant band information and long operation time, caused by a large amount of data, which exerts a certain negative impact on the model accuracy [10]. Therefore, it is highly important to select sensitive bands related to

agronomic parameters. Sun et al. [11] utilized continuous wavelet transform (CWT) to decompose and transform the canopy spectra under different irrigation treatments, and established that the PLSR model constructed using 2400, 1596, and 2397 nm bands effectively estimated wheat equivalent water thickness (EWT). In another investigation, Krishna et al. [12] found that the whole spectral reflectance band of rice was reduced to 32 by the loading weight method, which did not decrease the accuracy of the model. In addition to the aforementioned modeling methods, random forest, extreme random tree and k-nearest neighbor algorithms have also been used to estimate biomass, nitrogen content, leaf area, and SPAD [13–15]. However, these methods have been rarely applied for the assessment of wheat leaf water content.

In this study, five different models were established separately based on the canopy reflectance of seven wheat varieties under different irrigation treatments from 2018 to 2020. The main purpose of this investigation was to provide a future reference for hyperspectral monitoring of winter wheat leaf moisture under similar production conditions. To achieve this purpose, three major experimental sub-goals were defined: (i) to analyze the effect of different irrigation times on wheat leaf water content and spectral reflectance; (ii) to evaluate the performance of newly developed spectral indices and existing spectral indices for leaf water content; and (iii) to compare the ability of four the regression models established by the four algorithms to monitor leaf water content and to identify the optimal model among them.

Experimental materials and methods
Design of field experiments
Experiment 1
The experiment was conducted in the field of Luoyang Academy of Agricultural and Forestry Sciences, Luoyang City, Henan Province, China (112° 49′ 18″ E, 36° 64′ 14″ N from 2018 to 2019). The soil type is loamy cinnamon soil (Organic Matter: 14.4 g kg^{-1}, Total N:1.83 g kg^{-1}, P_2O_5:24.6 mg kg^{-1}, K:126.9 mg kg^{-1}). The experimental variables included different irrigation period treatments (w0: irrigation at the bottom stage w1: irrigation at the bottom and jointing stage, w2: irrigation at bottom, jointing and grain-filling stage) and seven different winter wheat cultivars (Luomai 27, Zhengmai 136, Zhengmai 22, Zhengmai 16, Zhonngyu 1211, Luomai 34 and Zhoumai 18). A randomized complete block design was used in all experiments, in which winter wheat was planted in 42 plots (each with a size of 2.6 m × 5 m) with two replicates.

Zhang *et al. Plant Methods*   (2021) 17:34

Page 3 of 14

Experiment 2

The experimental site was the same as that of experiment 1 (112° 49′ 18″ E, 36° 64′ 14″ N from 2019 to 2020). The soil type in this area is loamy cinnamon soil (Organic Matter: 13.2 g kg^{-1}, Total N:1.05 g kg^{-1}, P$_2$O$_5$:18.6 mg kg^{-1}, K:116.3 mg kg^{-1}). The area of the sub-region was 10.4 m^2 (2.6 m × 4 m). The cultivation management measures and the sampling period were the same as those in experiment 1. Additional details regarding the experimental design and the sample collection time are presented in Table 1 and Fig. 1.

Table 1 Experimental treatments and sampling periods

| Exp.no.	Year	Cultivars	Irrigation treatments	Irrigation date	Sampling time and date
Exp. 1	2018–2019	Luomai 27 Luomai 34 Zhengmai 16 Zhongyu 1211 Zhengmai 136 Zhoumai 18 Zhengmai 22	w0	Irrigation at the bottom stage/October 18, 2018	Mar 15
			w1	Irrigation at the bottom stage/October 18, 2018	Mar 30
				Irrigation at the jointing stage/March 20, 2019	Apr 18
			w2	Irrigation at the bottom stage/October 18, 2018	Apr 25
				Irrigation at the jointing stage/March 20, 2019	May 10
				Irrigation at grain-filling stage/May 1, 2019	May 20
Exp. 2	2019–2020	Same as above	W0	Irrigation at the bottom stage/October 22, 2019	Mar 13
			W1	Irrigation at the bottom stage/October 22, 2019	Mar 28
				Irrigation at the jointing stage/March 18, 2020	Apr 15
			W2	Irrigation at the bottom stage/October 22, 2019	Apr 23
				Irrigation at the jointing stage/March 18, 2020	May 8
				Irrigation at grain-filling stage/April 25, 2020	May 18
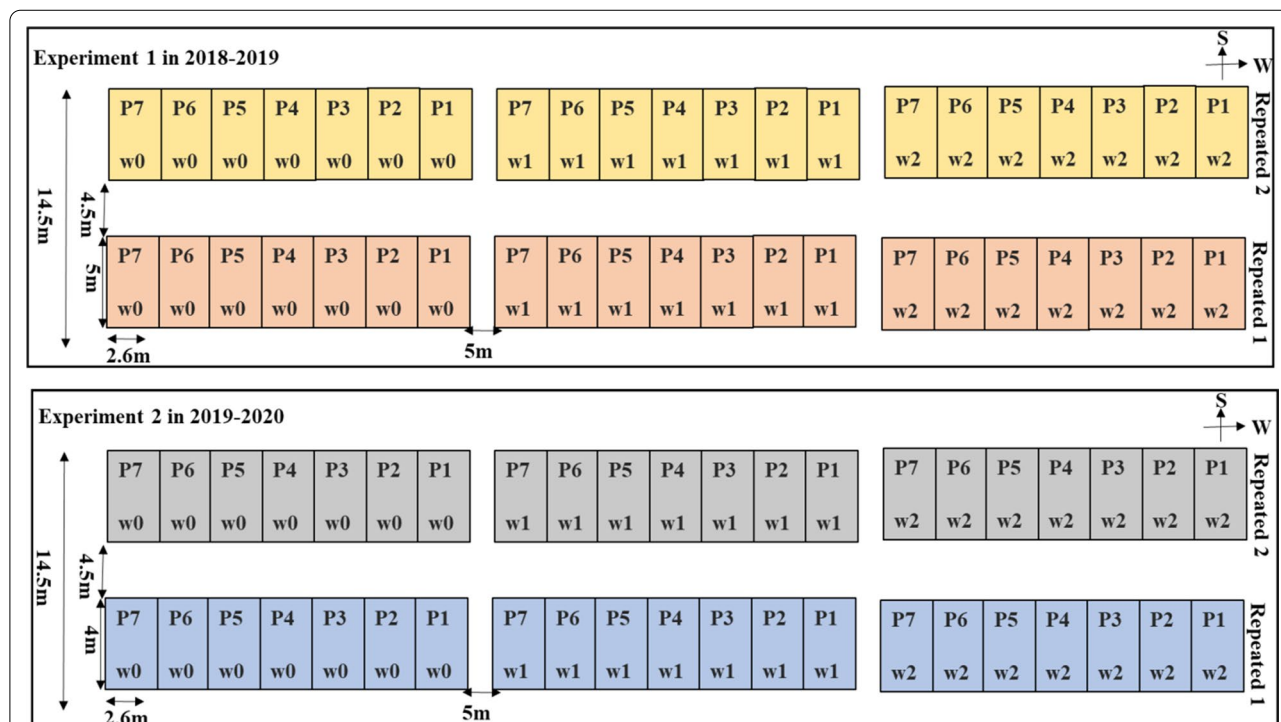
Fig. 1 Layout of the experimental design in 2018–2019 and 2019–2020. P1, P2, P3, P4, P5, P6 and P7 represented wheat cultivars, P1: Luomai 27, P2: Zhengmai 136, P3: Zhengmai 22, P4: Zhengmai 16, P5: Zhongyu 1211, P6: Luomai 34, P7: Zhoumai 18. w0, w1, and w2 represent different irrigation treatments: w0 is the irrigation during the bottom stage, w1 is the irrigation during the bottom and jointing stages, and w2 is the irrigation during the bottom, jointing, and grain-filling stages

Determination method and index
Measuring plant hyperspectral data
All canopy spectral reflectance measurements were obtained from a height of 1.0 m above the vertical canopy under clear sky conditions between 10:00 and 14:00 (Beijing local time), using ASD Field Spec Pro spectrometer (Analytical Spectral Devices, Boulder, CO, USA) with a 25° field of view fiber optics. Reflectance values in the 350–2500 nm range were collected with a sampling interval of 1.4 nm and spectral resolution of 3 nm between 350 and 1050 nm, and 2 and 10 nm, respectively, between 1050 and 2500 nm. Moreover, the black and baseline reflectance was calculated using a 0.40×0.40 m^2 white calibration panel made of BaSO4.To minimize the effects caused by sky and field conditions, spectral measurements were obtained from 10 sites in each plot and averaged into a single spectral sample. For each experiment, measurements were obtained on several dates that reflected the major growth stages of wheat. Using ViewSpec Pro version 6.0 software to assemble and interpret date.

Determination of the leaf water content
After the canopy spectra was measured, wheat plants were collected at corresponding points, and all leaves were extracted. The water content of the wheat leaves was determined by the drying method. The fresh weight of the leaves was weighed with an analytical balance (accuracy of 0.01 g). The leaves were then placed into an oven at 105 °C for 30 min. Finally, the samples were overdried at 80 °C until a constant weight was reached, which took approximately 3–5 days, and the dry weight was recorded. The LWC was calculated using the following formula [16]:

$$LWC = \frac{(m_f - m_d)}{m_f} \times 100 \qquad (1)$$

where LWC is the water content of leaf, g.g^{-1}; m_f is the fresh weight of leaves, g; m_d is the dry weight of the leaves, g.

Data analysis and utilization
During the 2 years of the experiment, a total number of 252 wheat samples were collected. The data from 2018–2019 were used for model calibration (n = 126), whereas the test data from 2019–2020 were utilized as validation samples (n = 126). The statistical parameters of the leaf water content determined in each sample set are shown in Table 2. A flow chart of the development and application of the winter wheat LWC estimation modeling method can be seen in Fig. 2.

Characteristic band screening
Due to the effect of the atmosphere on reflectance measurements, the ranges 1350–1400, 1800–1950, and 2450–2500 nm were excluded in this study. The assessment of the input of all spectral bands in the model would cause "dimension disaster". Therefore, the correlation coefficient and the load factor methods are used to screen the characteristic bands.

(1) Correlation coefficient (CA)

This method determines the characteristic bands based on a coefficient of the correlation between the spectral band and the parameters. The correlation between the leaf water content and the canopy reflectance under different irrigation treatments was analyzed in this investigation. The characteristic wavelength was determined by selecting the maximum absolute value of the correlation coefficient and the position of the wave crest and trough.

(2) x-Loading weight (x-Lw)

The loading weight based on the PLSR model can establish the proportion of the influence of different dependent variables of that of the total independent variables, which is of great significance for the rapid screening of characteristic bands [17]. The peak and valley values are extremum and can reflect the influence of a specific waveband on water content. In this study, the peak and trough were selected as the characteristic bands.

Table 2 Statistical parameters of the leaf samples used for determination of wheat leaf water content (LWC %)

Sample sets	Experimental year	Sample size	Maximum	Minimum	Mean	Standard deviation	Coefficient of variation
Total	2018–2020	252	88.61	55.87	78.82	4.76	6.02
Modeling set	2018–2019	126	88.61	55.87	78.80	5.71	7.25
Validation set	2019–2020	126	83.75	67.54	78.84	3.92	4.98

Fig. 2 Flow chart of the overall research work performed
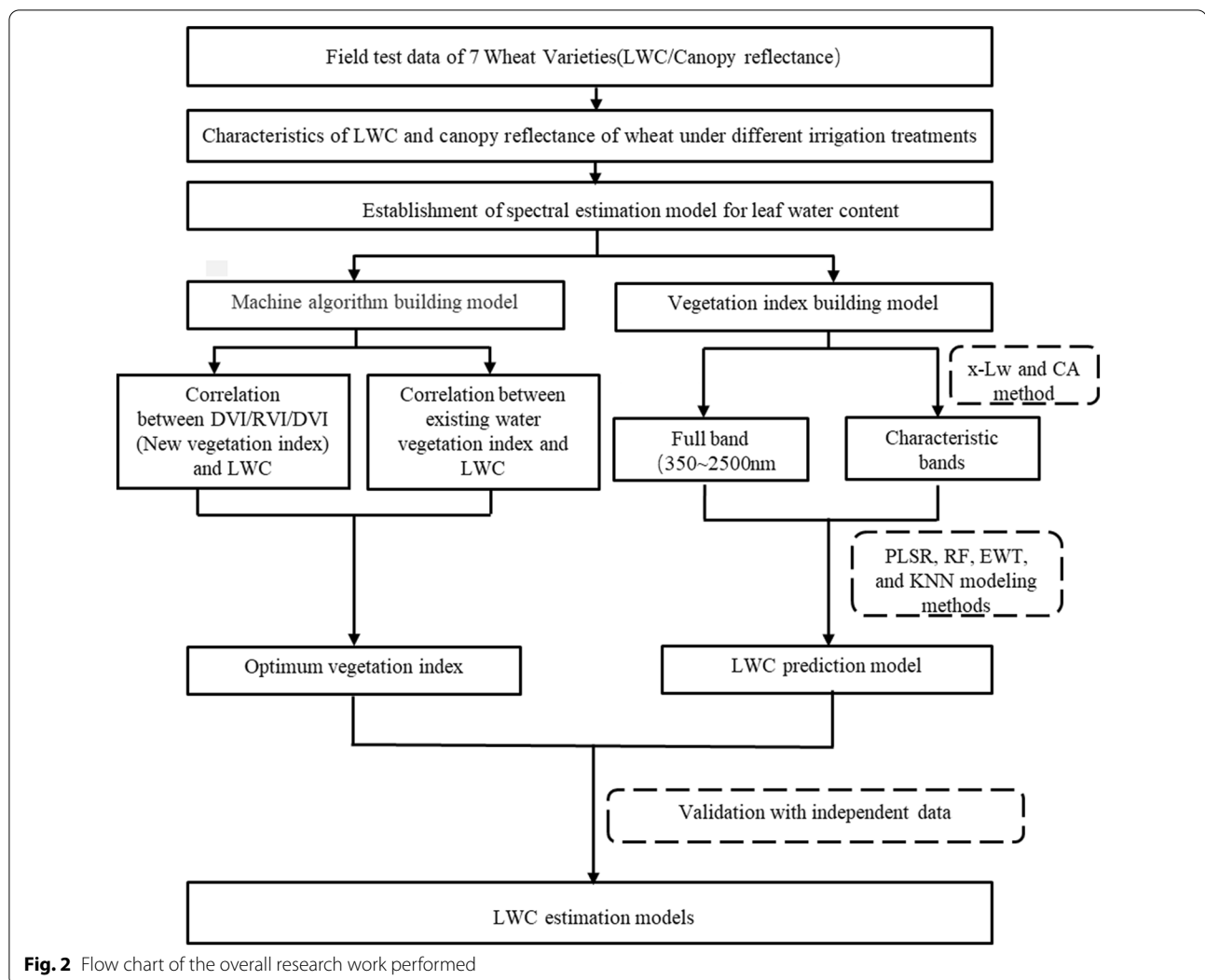
Based on the feature bands selected by the aforementioned two methods as independent variables, the inversion models for leaf water content estimation of winter wheat were constructed by PLSR, RFR, ERT, and KNN.

Modeling method

(1) Spectral index

In line with the existing research, the spectral index related to crop water status has been selected as shown in Table 3, and the relationship with leaf water content has also been analyzed. To obtain better spectral parameters, the normalized vegetation index (NDVI), the ratio vegetation index (RVI) and the difference vegetation index (DVI) were calculated in the range of 350–2500 nm, which were shown in the following formula, and the relationship between them and leaf water content was analyzed, so as to determine the optimal spectral estimation of leaf water content parameters were measured.

$$NDVI\,(\lambda 1,\ \lambda 2) = (R\lambda 1 - R\lambda 2)/\,(R\lambda 1 +\ R\lambda 2) \qquad (2)$$

$$RVI\,(\lambda 1,\ \lambda 2) =\ R\lambda 1/R\lambda 2 \qquad (3)$$

$$DVI\,(\lambda 1,\ \lambda 2) =\ R\lambda 1 - R\lambda 2 \qquad (4)$$

R λ 1 and R λ 2 represent the reflectivity of any two bands in the range of 350–2500 nm, respectively, and were selected by using self-developed code in MATLAB software.

(2) Partial Least-Squares Regression (PLSR)

Zhang *et al. Plant Methods*   (2021) 17:34

Page 6 of 14

Table 3 The calculation method and reference of spectral index related to crop moisture status

Spectral index	Definition or equation	References
Ratio Index	R1650/R2220	Elvidge et al. [18]
Normalized differential water index, NDWI	(R860−1240)/(R820+1240)	Gao et al. [19]
Moisture stress index, MSI	R1600/R820	Hunt et al. [20]
Maximum water difference index, MDWI	(Rmax1500−1750)-(Rmin1500−1750)/(Rmax1500−1750)+(Rmin1500−1750)	Eitel et al. [21]
Hyperspectral normalized difference vegetation index, hNDVI	(R900−R680)/(R900−R680)	Rouse et al. [22]
Water index, WI	R900/R970	Penuelas et al. [23]
Simple ratio water index, SRWI	R820/R1200	Zarco-Tejada et al. [24]
Normalized Difference Infrared Index, NDII	(R820−R1649)/(R820+R1679)	Hardisky et al. [25]
WI/hNDVI	(R900/R970)/[(R900−R680)/(R900−R680)]	Zhang et al. [8]
FD730-955	FD730/FD955	Liang et al. [26]

A model can be effectively established by PLSR in cases of significant multiple correlations of independent variables [27]. In the process of modeling, the large number of measured collinear spectral variables was reduced to a few non-correlated latent variables [14]. Therefore, the cross-correlation among multiple hyperspectral features can be explained by this method [28]. Here, we calculated and used the number of latent vectors yielding the smallest root mean squared error. We determined the number of latent vectors by the "Loo" cross-validation method using the Python 2.7 package "sklearn.cross decomposition".

(3) Random Forest Regression (RFR)

RFR can establish the relationship between multiple independent variables and a dependent variable. Importantly, this method improves the prediction accuracy of the individual model through fitting many regression trees [29]. To apply this technique, we used "sklearn. tree" in Python 2.7, with " n_estimators = 500, max_features = 'sqrt'".

(4) Extreme Random Trees (ERT)

ERT is a top-down method that is very similar to the random forest approach, but is different from the latter in two points: first, it does not adopt bootstrap sampling with a replacement strategy but directly uses the original training samples to reduce the deviation; second, the bifurcation value is completely random, which can achieve the bifurcation of a decision tree. The result is smaller and more stable than that of the random forest [30]. To apply this technique, we used library — "sklearn. ensemble" in Python 2.7, with the parameters of "n_estimators = 500, max_features = 'sqrt'".

(5) K-Nearest Neighbor (KNN).

KNN was proposed by Cover and Hart, which is a classification algorithm based on the proximity of similar samples in the pattern space [31]. Euclidean distance is used to measure the similarity between samples. A larger distance lowers the similarity. In this study, the k-nearest neighbor algorithm was employed in "sklearn. neighbors" of Python 2.7. A cross-validation method was used to determine K-values, K = 3.

Model validation

The coefficient of determination (R^2), root mean square error (RMSE) and the ratio of performance to deviation (RPD) were used to evaluate the accuracy of the model. The calculation formula is as follows:

$$R^2 = 1 - \sum_{i=1}^{n} \left(y_i - \widehat{y_i}\right)^2 / \sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2 \qquad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (y_i - \widehat{y_i})^2} \qquad (6)$$

$$\text{RPD} = SD/RMSE \qquad (7)$$

where y and $\widehat{y}$ represent the measured and predicted values, respectively. $\overline{y}$ represents the average of measured values. n is the number of samples. SD is the standard deviation of the measured data. The larger R^2 with smaller RMSE values indicate good model prediction accuracy. The RPD values > 1.4 indicate that the prediction ability of the model is acceptable, and the model can be applied.
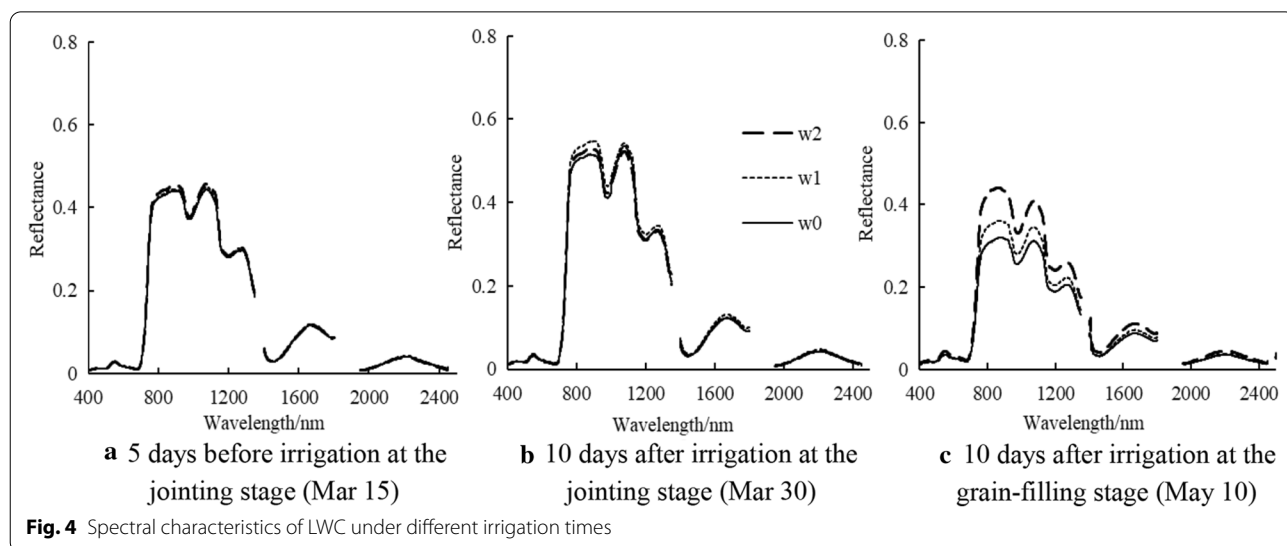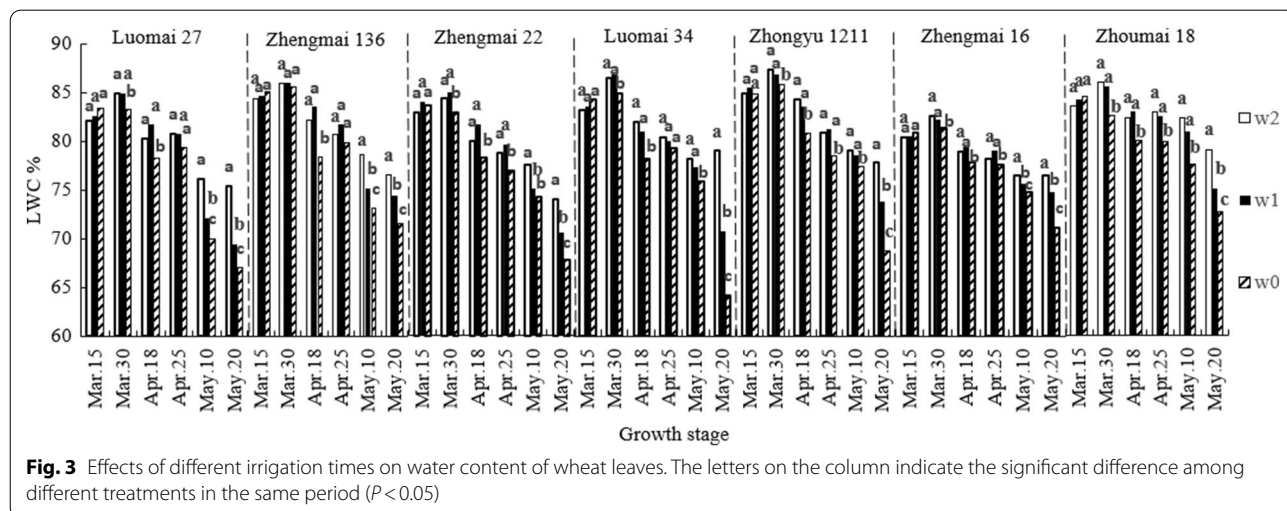
Results

Effects of the irrigation times on the leaf water content (LWC) and canopy spectral reflectance of wheat

To explain the effects of different irrigation times on LWC and canopy spectra, the data of experiment 1 were taken as an example. As can be seen from Fig. 3, plant growth of all varieties was accompanied with an initial increase in LWC, followed by a decrease. The highest LWC value was reached 10 days after the irrigation at the jointing stage (Mar 30). At the early wheat growth stage, an insignificant difference was observed in LWC among the different irrigation time treatments. However, it decreased rapidly in the late growth period, which gradually increased the difference. The LWC in w0 was the lowest, which was significantly lower than those in w2 and w1. Our results showed that the

irrigation at the grain-filling stage delayed leaf senescence and prolonged the green color-retaining period of leaves.

We analyzed the effect of irrigation times on wheat spectral reflectance in Luomai 27, considering it a representative example (Fig. 4). Five days before the irrigation at the jointing stage (Mar 15), the canopy reflectance was slightly lower than that detected 10 days after the irrigation at the jointing stage (Mar 30) at the same irrigation treatment times. The reflectance decreased again after the point of 10 days after the irrigation at the grain-filling stage (May 10). Therefore, the canopy reflectance was initially augmented but then decreased with the advancement in growth. The canopy reflectance in the visible region (350–750 nm) did not change significantly with the prolongation in the irrigation times, whereas

Fig. 3 Effects of different irrigation times on water content of wheat leaves. The letters on the column indicate the significant difference among different treatments in the same period ($P < 0.05$)

a 5 days before irrigation at the jointing stage (Mar 15)

b 10 days after irrigation at the jointing stage (Mar 30)

c 10 days after irrigation at the grain-filling stage (May 10)

Fig. 4 Spectral characteristics of LWC under different irrigation times

Zhang *et al. Plant Methods*   (2021) 17:34

Page 8 of 14

the canopy reflectance in the near-infrared region (750–1350 nm) gradually rose. The main reason for that tendency is that the sufficient water supply received by plants through the irrigation accelerated the growth of leaf area index and biomass. In return, the canopy reflectance increased. In particular, there was significant difference in canopy reflectance among different irrigation treatments after filling water, the reflectance of w2, w1, and w0 reached 43.45%, 35.28%, and 31.01% at 820 nm, respectively. The values of the canopy reflectance of w2, w1, and w0 were 32.27%, 28.08%, and 25.58% at 970 nm, respectively. The canopy reflectance of w2, w1, and w0 were 24.00%, 20.47%, and 18.85% at 1266 nm, respectively. It is more evident after filling water, that was w2 > w1 > w0. The different irrigation times did not influence the variations in the canopy spectral reflectance, which were basically the same in all wheat varieties.

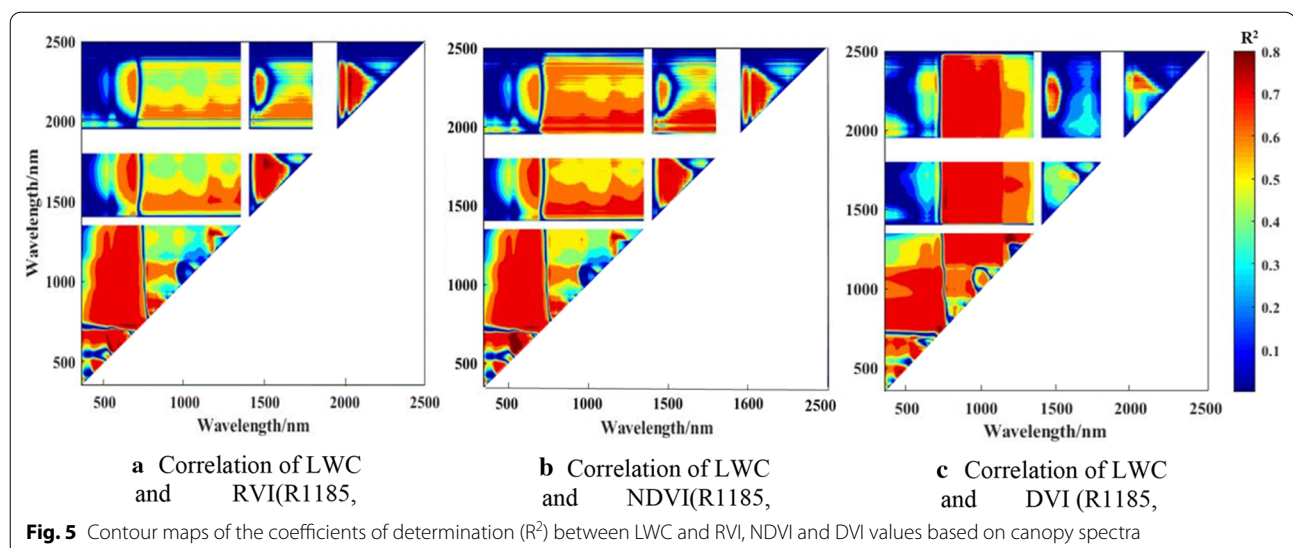### Correlation between LWC and spectral index of wheat leaves

To establish a new spectral vegetation index, the correlations between LWC and all possible two-band combinations of NDVI (normalized spectral index), RVI (ratio spectral index, and DVI (difference spectral index) were analyzed. Then, the contour maps of coefficient of determination between LWC and new spectral vegetation index were plotted (Fig. 5). According to Fig. 5, the highest R^2 band combination was extracted from the hot spot area as the best spectral index of leaf water content. The results showed that the sensitive regions of the three spectral index combinations were consistent, and the combination of 800–1300, 1600–1900, and 1950–2200 nm was better. The best result was obtained that
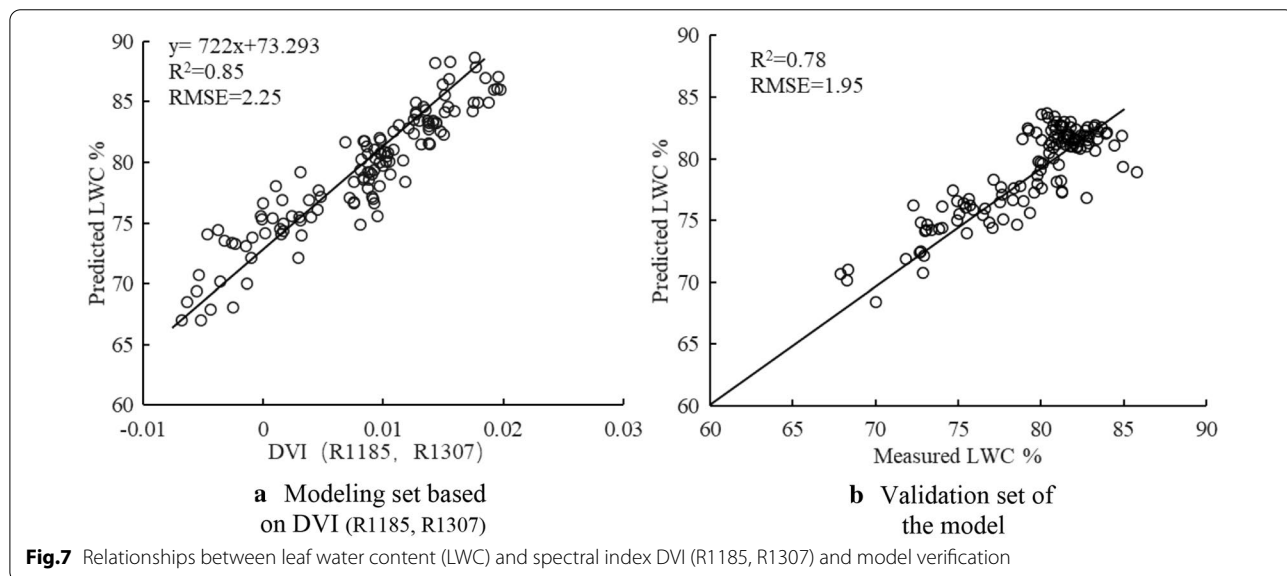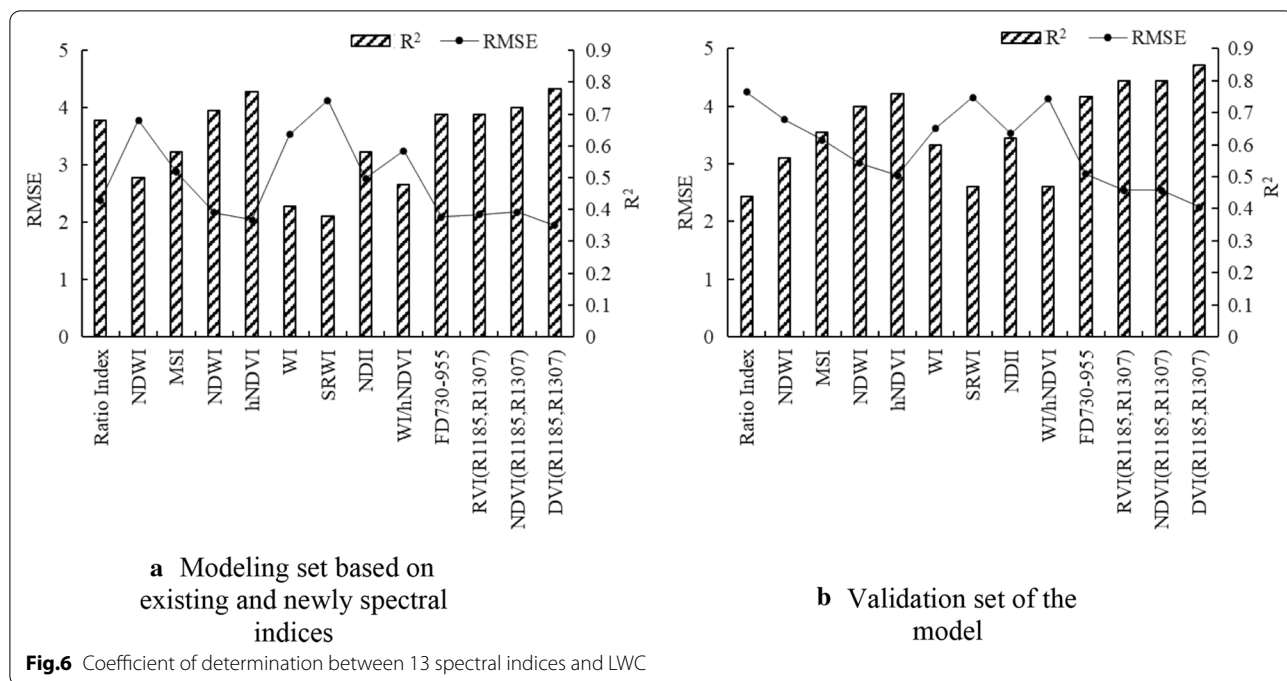
NDVI, RVI, and DVI consisting of 1185 nm and 1307 nm performed the best for LWC estimation.

Analyses of the correlations between wheat leaf water content under different irrigation treatments and the existing water-related vegetation indexes (Ratio index, NDWI, MSI, MDWI, hNDVI, NDII, WI, SRWI, WI/hNDVI, and $FD_{730-955}$), as well as of the newly developed three vegetation indexes were conducted, and the prediction performance of the models constructed by the 13 spectral indices were compared and analyzed (Fig. 6). In the modeling set, MSI, NDWI, hNDVI, WI, NDII, and $FD_{730-955}$ were the spectral indices with R^2 higher than 0.6. The models were validated by the data obtained in Experiment 2, where R^2 ranged from 0.38 to 0.78. Our results revealed that the proposed DVI (R1185, R1307) had the best performance with high R^2 and low RMSE values. The best linear equation of LWC predicted by DVI value was illustrated in Fig. 7, with calibration R^2 of 0.85 and RMSE of 2.25, validation R^2 of 0.78, and RMSE of 1.95. Therefore, the newly developed indices can be used for accurate estimation of the changes in RWC caused by irrigation times in wheat.

Extraction of the characteristic bands of LWC

The correlation between LWC and the original spectral reflectance (350–2500 nm) under different irrigation times was depicted in Fig. 8. We found that the correlation coefficient ranged from −0.83 to 0.87, with maximum positive and negative correlation coefficient values of 0.86 (618 nm) and −0.83 (769 nm), respectively. By selecting the maximum absolute value of the correlation coefficient at the local peaks and troughs, we established

Fig. 5 Contour maps of the coefficients of determination (R^2) between LWC and RVI, NDVI and DVI values based on canopy spectra

Zhang *et al. Plant Methods*   (2021) 17:34

Page 9 of 14

Fig.6 Coefficient of determination between 13 spectral indices and LWC

Fig.7 Relationships between leaf water content (LWC) and spectral index DVI (R1185, R1307) and model verification
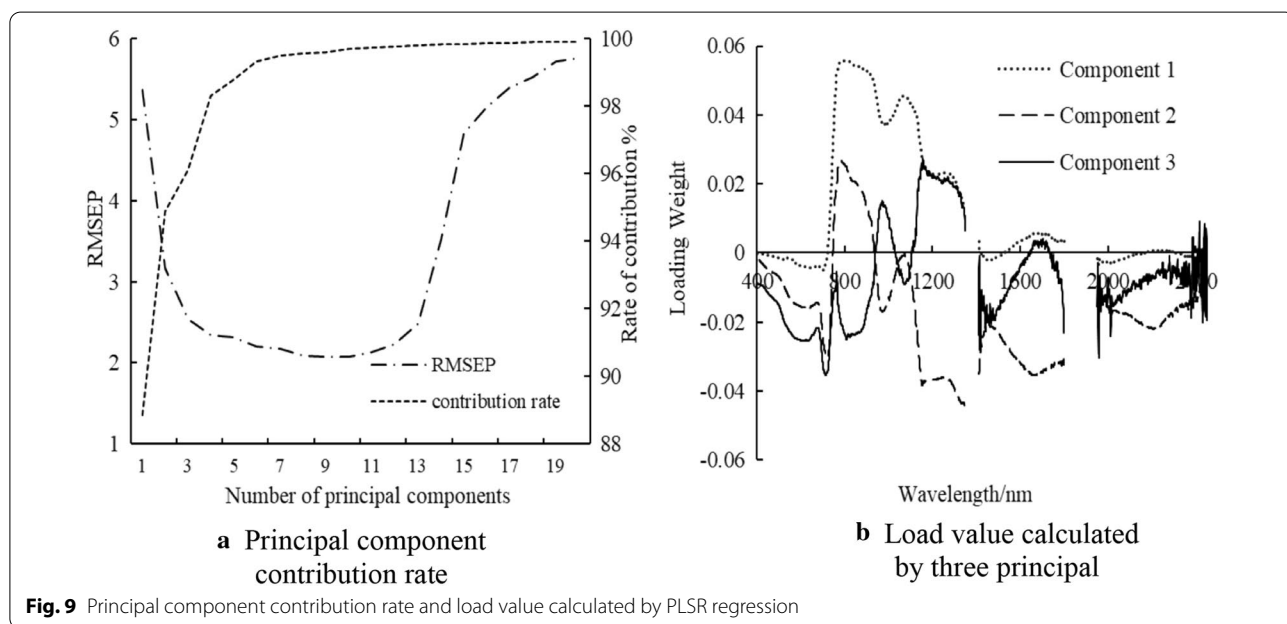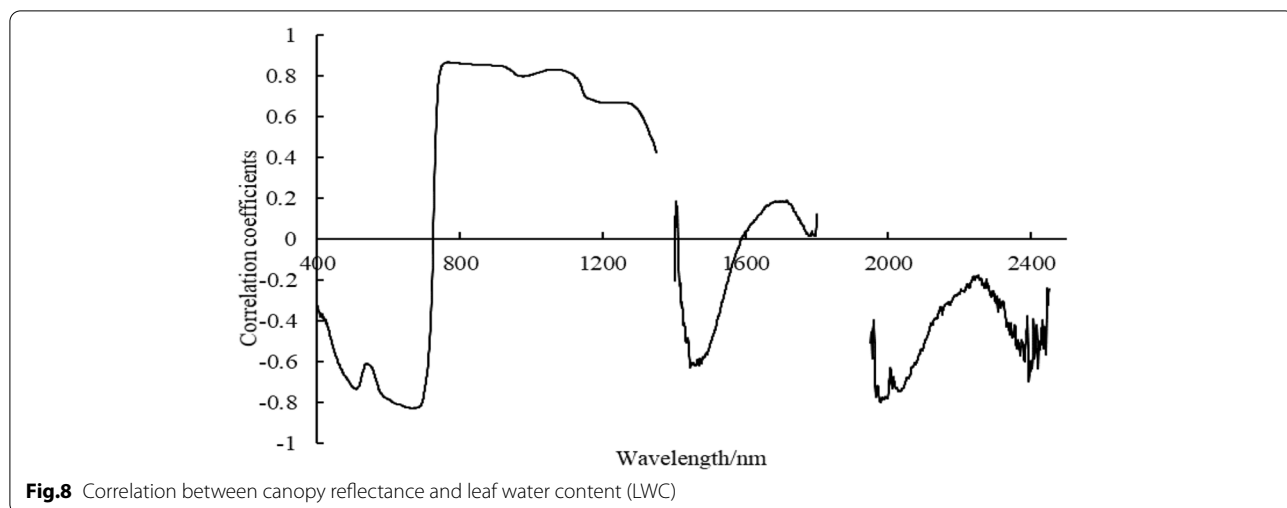
the optimal wavelengths, which were 505, 551, 681, 747–831, 989, 1101, 1158, 1445, 1716, 1782, 1978, 2000, 2007, 2038, 2242, and 2394 nm.

Based on the contribution rate of the PLSR model and RMSEP, the number of principal components was determined. When the principal component was 3, RMSEP was 2.34% and explained 96.12% of the variance (88.86%, 6.03%, and 1.23% for PC1, PC2, and PC3,

correspondingly), which was presented in Fig. 9a. Therefore, the best characteristic band was determined by the peaks and troughs of the loading weight values of the three principal components (Fig. 9b). According to the data if the above analysis, the optimal bands determined were 588, 663, 674, 680, 700, 763, 777, 783, 808, 816, 970, 977, 984, 1070, 1072, 1156, 1205, 1246, 1264, 1402, 1445, 1456, 1660, 1678, 1957, 1702, 2221, and 2252 nm.

Fig.8 Correlation between canopy reflectance and leaf water content (LWC)

a Principal component contribution rate

b Load value calculated by three principal

Fig. 9 Principal component contribution rate and load value calculated by PLSR regression

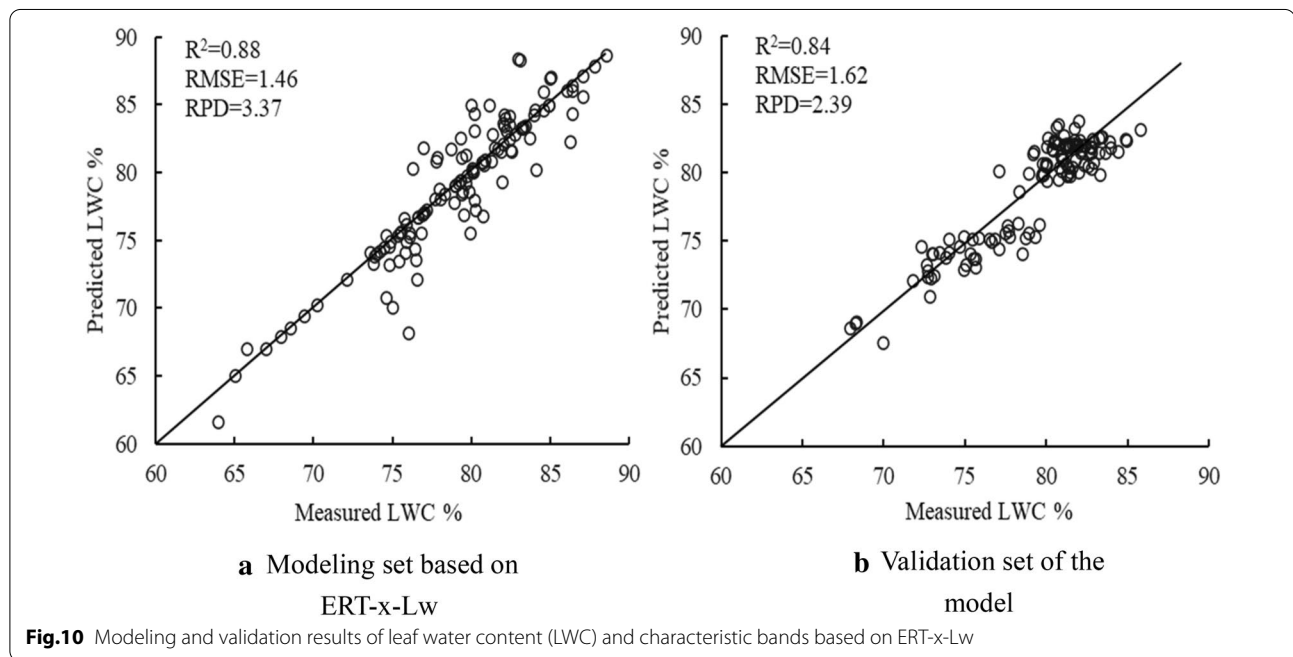## Comparison of LWC inversion models constructed by different methods

The full band and the characteristic band selected by correlation coefficient (CA) and x-Loading weight (x-Lw) were used as independent variables respectively, and LWC was taken as the dependent variable. After the PLSR, KNN, RF and ERT models were established.

The full and characteristic bands that were selected by the two methods as input and the models constructed by different modeling methods are presented in Table 4. The accuracy of the model was not significantly lower than that of the full-band model, with feature bands extracted from CA and x-Lw as independent variables, but the input variables were reduced, which improved the efficiency of the model. For 100 characteristic bands selected by CA method was used as the input of the model, and the performance of the model was as follows: PLSR-CA > ERF-CA > RF-CA > KNN-CA. For the 28 characteristic bands selected by x-Lw method, the performance of the model was as follows: ERF-x-Lw > PLSR-x-Lw > RF-x-Lw > KNN-x-Lw. Compared with the CA method, the number of the independent variable was lower by 98.63% in the x-Lw method, which significantly improved the modeling efficiency.

Table 4 Regression analysis of the characteristic bands and LWC by different modeling methods

Modeling method	Feature band screening method	Number of modeled bands	Modeling			Validation		
			R^2	RMSE	RPD	R^2	RMSE	RPD
PLSR	Full band	1901	0.87	2.11	2.73	0.82	1.78	2.09
	CA	100	0.86	2.10	2.52	0.84	1.61	2.43
	x-LW	28	0.86	2.61	1.96	0.84	1.73	2.04
RF	Full band	1901	0.88	2.80	1.26	0.83	1.69	2.31
	CA	100	0.90	2.49	1.63	0.81	1.75	2.15
	x-LW	28	0.88	1.57	3.10	0.80	1.86	1.92
ERF	Full band	1901	0.87	1.36	2.18	0.85	1.52	2.30
	CA	100	0.86	1.95	2.48	0.82	1.76	2.34
	x-LW	28	0.88	1.46	3.37	0.84	1.62	2.39
KNN	Full band	1901	0.82	2.10	1.61	0.83	1.61	2.20
	CA	100	0.85	2.00	2.25	0.80	1.79	2.16
	x-LW	28	0.84	2.04	1.80	0.80	1.74	1.83

Fig.10 Modeling and validation results of leaf water content (LWC) and characteristic bands based on ERT-x-Lw

Among the 12 models constructed, the ERT-x-Lw model had the higher R^2 value and lowest RMSE. In the modeling set, R2, RMSE, and RPD were 0.88, 1.46, and 3.37, respectively, whereas in the validation set, R2, RMSE, and RPD were 0.84, 1.62, and 2.39, correspondingly (Fig. 10).

Discussion

The growth and development of wheat can be directly affected by water deficit. Leaf water content is an important indicator of wheat growth that can be monitored by hyperspectral technology. The hyperspectral reflectance technique, which detects the spectral reflectance of the canopy from visible- (VIS) to shortwave-infrared (SWIR), has been shown to have a great potential to detect even the slight variations and modifications in biophysical and biochemical characteristics of crop canopy [32, 33]. Furthermore, canopy reflectance in the near-infrared (NIR,700–1300 nm) and the shortwave-infrared (SWIR, 1300–2500 nm) regions are strongly influenced by several internal leaf structural and water content of the canopy. Therefore, it is successfully used to track the changes in plant parameters related to plant water status and monitor the changes of biomass and leaf area indirectly

Zhang *et al. Plant Methods*   (2021) 17:34

Page 12 of 14

caused by wheat moisture [34]. In this study, the spectral reflectance of wheat canopy initially increased but then decreased with the advance of plant growth. At the jointing stage, the canopy reflectance of wheat increased after the irrigation was performed, compared with that without irrigation. Further, the reflectance decreased significantly in the treatment of 10 days after the irrigation at the grain-filling stage. The reason for these results may be that a growth of wheat biomass and leaf area at the jointing stage, especially the timely irrigation at that stage, was able to increase the water absorption rate of plants, elevating the leaf water content and thus further accelerating the growth of wheat plants, which ultimately contributed to improving the canopy reflectance. However, at the later stage of grain-filling, the leaf area and leaf water content gradually decreased with plant senescence, which led to a decline in canopy reflectance [35, 36].

In addition, the different irrigation times exerted a significant effect on canopy spectral reflectance. No significant difference was observed among three treatments before and after the jointing stage. In the near-infrared region (750–1350 nm), the canopy reflectance increased significantly with the increase of the irrigation times after the jointing stage, which was due to the rise in the wheat plant height, chlorophyll content, and net photosynthetic rate. It is noteworthy that the canopy reflectance of w2 was significantly higher than those of w1 and w0. This result indicates that the leaf senescence and photosynthesis time after flowering can be delayed by irrigation at the grain-filling stage. However, without irrigation, wheat plants grow shorter; the leaves turn yellow and wither ahead of time; the lower leaf water content causes cell structure changes, which eventually leads to wheat yield reduction [36].

The change of canopy reflectance is caused by the change of LWC. In this study, LWC showed no significant differences among the treatments on Mar 15 (Fig. 3) and canopy reflectance also had no changes (Fig. 4a). At 10 days after the irrigation at the jointing stage, the LWC of the non-irrigated treatment (w0) was significantly lower than those of the other treatments (Fig. 3), and the canopy reflectance was the lowest in the near-infrared band (Fig. 4b). After the irrigation at the grain-filling stage, LWC was marked by significant differences among the treatments (w2 > w1 > w0; Fig. 3), and canopy reflectance also changed (w2 > w1 > w0; Fig. 4c). A previous study established water-sensitive bands at 820, 970, 1200, and 1450 nm [37]. In this study, we found that canopy reflectance in these bands was different among the treatments, and decreased with the decline of LWC. Therefore, our model was constructed by the characteristic response band of leaf water content, which can be used to diagnose and retrieve leaf

water content. Many studies have developed spectral indices for estimating of the leaf water index. However, due to the diversity in the experimental conditions in different studies, various bands have been selected in the spectral index. For example, Liang et al. [26] used the first derivative ratio index at 730 and 955 nm to predict wheat leaf water content, achieving a value of the modeling prediction coefficient R^2 of 0.74. In another investigation, Jiang et al. [38] selected the bands at 1300 and 1200 nm for the development of a ratio spectral index for the prediction of the water content in wheat leaves, with R^2 of 0.63. On the other hand. The bands at 1391 and 1830 nm were used by other researchers to predict the wheat leaf water status [9]. DVI (R1185, R1307) developed in this study showed high accuracy; moreover, it is similar to wavebands selected by Jiang et al. [38] and Wu et al. [39]. Furthermore, the two bands selected in this study were in the water-sensitive near-infrared region [40]. Our model yielded a modeling R^2 value of 0.85, and a prediction R^2 value of 0.78 and was thus superior to the water spectral indices developed in previous studies.

To improve the modeling accuracy, machine learning and other methods have also been applied to model and analyze the water content of wheat. Several researchers in the past [41], based on the grey correlation analysis method, selected a spectral index with a high correlation to be used for leaf water content analysis. These spectral indices were used as independent variables in PLSR and Back Propagation (BP) neural network models to predict wheat leaf water content, with R^2 values of 0.72 and 0.80, respectively [41, 42]. In the present study, the correlation coefficient (CA) and x-loading weight (x-Lw) methods were employed to select the characteristic bands. Compared with the CA method, the x-Lw method reduces the number of bands by 72%, which may be due to the concentration of sensitive bands extracted by the CA method and the smaller adjacent interval [43]. Twenty-eight characteristic bands related to leaf water content were selected by x-Lw. Among them, the wavelengths of 663, 674, 680, and 700 nm are located in the "red edge" region, which can indirectly diagnose the water status of wheat due to the high reflection of the crop leaf structure [44]. On the other hand, 1156, 1205, 1246, and 1264 nm are related to the leaf and canopy cell structure [12]. Additionally, 1402, 1445, 1456, and 1957 nm are associated with the water absorption band [45]. This is basically consistent with the water-related bands selected in a previous experiment [12]. Among the four modeling methods, the one using x-Lw-ERT provided the best prediction for leaf water content retrieval. The values of the coefficient of determination (R^2) during the calibration and validation were 0.88 and 0.84, correspondingly, and

Zhang *et al. Plant Methods*   (2021) 17:34

Page 13 of 14

RMSE were 1.46 and 1.62, respectively which were higher than those of the PLSR, RF, and KNN models. These outcomes were probably due to that the better generalization ability and more stable performance of ERT [46]. However, KNN breaks the continuous characteristics of the band because it learns and predicts according to the distance features between different samples [47]. Our findings suggest that the extreme random tree (ERT) may be a reliable modeling method to improve the modeling accuracy of machine learning and other methods used to model and analyze the moisture content of wheat.

Conclusion

We studied the effect of different irrigation times on the wheat canopy reflectance spectrum based on a field experiment with different irrigation treatments in two consecutive years. Five different models were compared, the results of which showed that the irrigation at the jointing + grain-filling stage increased the leaf water content, leaf area, and biomass; moreover, plant senescence was delayed and canopy reflectance elevated. The new model constructed by DVI (R1185, R1307) can be used to estimate the LWC of winter wheat. The accuracy of the extreme random tree model based on the x-loading weight method had the best performance among the other compared models. Therefore, both models can be used to estimate the water content of wheat leaves effectively.

Availability of data and materials
The processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
This manuscript has not been published elsewhere and is not under consideration by another journal.

Competing interests
This manuscript has no conflicts of interest to declare.

Author details
[1] Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University, #63 Nongye Road, Zhengzhou 450002, Henan, People's Republic of China. [2] College of Agronomy, Henan Agricultural University, #63 Nongye Road, Zhengzhou, Henan 450002, People's Republic of China. [3] Science College of Information and Management, Henan Agricultural University, #63 Nongye Road, Zhengzhou 450002, People's Republic of China. [4] Adelphi University, # One South Avenue, Garden City, NY 11530-0701, USA. [5] Luoyang of Agriculture and Forestry, #1 Nongke Road, Luoyang 471000, Henan, People's Republic of China.

References
1. Zhang CW, Qi YQ, Dai MH, Zhang YC, Shen YJ. Effects of multi-factor on water use efficiency as identifid by the SEM method in irrigated wheat farmlands in the North China Plain. Chin J Eco Agric. 2020;28(6):876–86. https://doi.org/10.13930/j.cnki.cjea.190924.
2. Rodríguez-Pérez JR, Ordóñez C, González-Fernández AB, Sanz-Ablanedo E, Valenciano JB, Marcelo V. Leaf water content estimation by functional linear regression of field spectroscopy data. Biosyst Eng. 2018;165:36–46. https://doi.org/10.1016/j.biosystemseng.2017.08.017.
3. Pockman WT, Sperry JS. Vulnerability to xylem cavitation and the distribution of Sonoran Desert vegetation. Am J Bot. 2000;87(9):1287–99. https://doi.org/10.2307/2656722.
4. Wang J, Zhao CJ, et al. Effect of soil water content on the wheat leaf water content and the physiological function. J Triticeae Crops. 2001;21:42.
5. Yoshio I, Shinsuke M, Michio S. Non-destructive estimation of water status of intact crop leaves based on spectral reflectance measurements. Jpn J Crop Sci. 1993;62(3):462–9. https://doi.org/10.1626/jcs.62.462.
6. Zhao SH, Wang Q, Yao YJ, Du S. Estimating and validating wheat leaf water content with three MODIS spectral indexes: a case study in Ning xia Plain, China. J Agric Sci Technol. 2016;18(2):387–98.
7. Rapaport T, Hochberg U, Cochavi A, Karnieli A, Rachmilevitch S. The potential of the spectral 'water balance index' (WABI) for crop irrigation scheduling. New Phytol. 2017;216(3):741–57. https://doi.org/10.1111/nph.14718.
8. Zhang C, Liu JG, Shang JL, Cai HJ. Capability of crop water content for revealing variability of winter wheat grain yield and soil moisture under limited irrigation. Sci Total Environ. 2018;631–2:677–87. https://doi.org/10.1016/j.scitotenv.2018.03.00.
9. Das B, Sahoo RN, Pargal S, Verma R, Chinnusamy V, Sehgal VK, et al. Comparison of different uni- and multi-varia techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. Biosyst Eng. 2017;160:69–83. https://doi.org/10.1016/j.biosystemseng.201705.007.
10. Chen Y, Tan Y, Lorenzo B, Lu L, Guan R. Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. Remote Sens. 2017;9(8):782. https://doi.org/10.3390/rs9080782.
11. Sun Q, Sun L, Wang M, Gu XH. Spectral response analysis of canopy water content of winter wheat under different irrigation conditions. Scientia Agricultura Sinica. 2019;52:2425–35.
12. Krishna G, Sahoo RN, Singh P, Bajpai V, Patra H, Kumar S, et al. Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing. Agric Water Manag. 2019;213:231–44. https://doi.org/10.1016/j.agwat.2018.08.029.
13. Yue JB, Yang GJ, Feng HK. Comparative of remote sensing estimation models of winter wheat biomass based on random forest algorithm. Chin Soc Agric Eng. 2016;32(18):175–82.
14. Asner GP, Martin RE, Tupayachi R, Emerson R, Martinez P, Sinca F, et al. Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests. Ecol Appl. 2011;21(1):85–98. https://doi.org/10.1890/09-1999.1.
15. Cui RX, Liu YD, Fu JD. Estimation of winter wheat leaf nitrogen accumulation using machine learning algorithm and visible spectral. Spectrosc

Zhang *et al. Plant Methods*   (2021) 17:34

Page 14 of 14

Spectr Anal. 2016;36(6):1837–42. https://doi.org/10.3964/j.issn.1000-0593(2016)06-1837-06.

16. Ding XL, Zhao LX, Zhou TT, Li YB, Huang XM, Zhao YL. Research on wheat leaf water content based on machine vision. Cluster Comput. 2019;22(4):9199–208. https://doi.org/10.1007/s10586-018-2112-4.

17. Ni FP, Zhu XW, Gu F, Hu YH. Nondestructive detection of apple crispness via optical fiber spectroscopy based on effective wavelengths. Food Sci Nutr. 2019;7(11):3654–63. https://doi.org/10.1002/fsn3.1222.

18. Elvidge CD, Lyon RJP. Estimation of the vegetation contribution to the 1·65/2·22 μm ratio in airborne thematic-mapper imagery of the Virginia Range. Nevada Int J Remote Sens. 1985;6(1):75–88. https://doi.org/10.1080/01431168508948425.

19. Gao BC. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ. 1996;58(3):257–66. https://doi.org/10.1016/S0034-4257(96)00067-3.

20. Hunt ER, Rock BN. Detection of changes in leaf water content using near and middle-infrared reflectance. Remote Sens Environ. 1989;30(1):43–54. https://doi.org/10.1016/0034-4257(89)90046-1.

21. Eitel JUH, Gessler PE, Smith AMS, Robberecht R. Suitability of existing and novel spectral indices to remotely detect water stress in Populus spp. Forest Ecol Manag. 2006;229(1–3):170–82. https://doi.org/10.1016/j.foreco.2006.03.027.

22. Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M. (Eds.), Third Earth Resources Techn ology Satellite–1 Syposium. Volume I: Technical Presentations. NASA SP-351, NASA, Washing ton, D.C, pp. 979; p.309–317.

23. Penuelas J, Pinol J, Ogaya R, Filella I. Estimation of plant water concentration by the reflectance water index WI (R900/R970). Int J Remote Sens. 1997;18:2869–75. https://doi.org/10.1080/014311697217396.

24. Zarco-Tejada PJ, Ustin SL. Modeling canopy water content for carbon estimates from MODIS data at land EOS validation sites. Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium. 2001; p.342–344. https://doi.org/10.1109/IGARSS.2001.976152

25. Hardisky MA, Klemas V, Smart RM. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. Remote Sens. 1983;48(1):77–84.

26. Liang L, Zhang LP, Lin HY. Estimating canopy leaf water content in wheat based on derivative spectral. Scientia Agricultura Sinica. 2013;46(1):18–29. https://doi.org/10.3864/j.issn.0578-1752.2013.01.003.

27. Rischbeck P, Elsayed S, Mistele B, Barmeier G, Heil K, Schmidhalter U. Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. Eur J Agron. 2016;78:44–59. https://doi.org/10.1016/j.eja.2016.04.013.

28. Rao CR, Wu Y. Linear model selection by cross-validation. J Stat Plan Inference. 2003;128(1):231–40. https://doi.org/10.1016/j.jspi.2003.10.004.

29. Rodriguez-Galiano V, Mendes MP, Garcia-Soldado MJ, Chica-Olmo M, Ribeiro L. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). Sci Total Environ. 2014;476–7:189–206. https://doi.org/10.1016/j.scitotenv.2014.01.001.

30. Shang K, Yao Y, Li Y, Yang J, Guo X. Fusion of five satellite-derived products using extremely randomized trees to estimate terrestrial latent heat flux over Europe[J]. Remote Sens. 2020;12(4):687. https://doi.org/10.3390/rs12040687.

31. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf theory. 2006;13:21–7. https://doi.org/10.1109/TIT.1967.1053964.

32. El-Hendawy SE, Al-Suhaibani NA, Elsayed S, Hassan WM. Potential of the existing and novel spectral reflectance Potential of the existing and novel spectral reflectance indices for estimating the leaf water status and grain yield of spring wheat exposed to different irrigation rates. Agric Water Manag. 2019;217:356–73. https://doi.org/10.1016/j.agwat.2019.03.006.

33. Vescovo L, Wohlfahrt G, Balzarolo M, Pilloni S, Sottocornola M, Rodeghiero M, et al. New spectral vegetation indices based on the near-infrared shoulder wavelengths for remote detection of grassland phytomass. Int J Remote Sens. 2012;33(7):2178–95. https://doi.org/10.1080/01431161.2011.607195.

34. Wang X, Zhao C, Guo N, Li Y, Kai Y. Determining the canopy water stress for spring wheat using canopy hyperspectral reflectance data in Loess Plateau semiarid regions. Spectrosc Lett. 2015;48(7):492–8. https://doi.org/10.1080/00387010.2014.909495.

35. Yang CB, Feng MC, Sun H, Wang C, Yang WD, Xie YK, et al. Hyperspectral monitoring of aboveground dry biomass of winter wheat under different irrigation treatments. Chinese Journal of Ecology. 2019; 38(6): 1767–73. http://www.cje.net.cn/CN/Y2019/V38/I6/1767

36. Guo R, Zhou J, Yang F, Li F, Xia X, Liu Q. Growth metabolism of wheat under drought stress at the jointing-booting stage. Chin J Plant Ecol. 2016;40(12):1319–27. https://doi.org/10.17521/cjpe.2016.0107.

37. Wang J, Xu R, Yang S. Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940 nm regions. Environ Monit Assess. 2009;157(1–4):459–69. https://doi.org/10.1007/s10661-008-0548-3.

38. Jiang JB, Huang WJ, Chen YH. Using canopy hyperspectral ratio index to retrieve relative water content of wheat under yellow rust stress. Spectrosc Spectr Anal. 2010;30(7):1939–43. https://doi.org/10.3964/j.issn.1000-0593(2010)07-1939-05.

39. Wu CY, Zheng N, Tang Q, Huang WJ. Predicting vegetation water content in wheat using normalized difference water indices derived from ground measurements. J Plant Res. 2009;122(3):317–26. https://doi.org/10.1007/s10265-009-0215-y.

40. Ranjan R, Sahoo RN, Chopra UK, Pramanik M, Singh AK, Pradhan S. Assessment of water status in wheat (*Triticum aestivum* L.) using ground based hyperspectral reflectance. Pro Natl Acad Scit India B Bio Sci. 2017;87(2):377–88. https://doi.org/10.1007/s40011-015-0618-6.

41. Jin XL, Xu XG, Song XY, Li ZH, Wang JH. Estimation of leaf water content in winter wheat using grey relational analysis-partial least squares modeling with hyperspectral data. Agron J. 2013;105(5):1385–92. https://doi.org/10.2134/agronj2013.0088.

42. Hasan U, Sawut M, Kasim N, Taxipulati N, Wang JZ, Ablat I. Hyperspectral estimation model of leaf water content in spring wheat based on grey relational analys. Spectrosc Spectr Anal. 2018;38(12):3905–11.

43. Zhang J, Tian HQ, Zhao ZY, Zhang L, Zhang J, Li F. Moisture content dectection in silage maize raw material based on hyperspectrum and improved discrete particle swarm. Chin Soc Agric Eng. 2014;35(1):285–93.

44. Ullah S, Andrew K, Ramoelo A, Groen TA, Ali A. Retrieval of leaf water content spanning the visible to thermalinf rared spectra. ISPRS J Photogramm Remote Sens. 2014;93(7):56–64. https://doi.org/10.1016/j.isprsjprs.2014.04.005.

45. Wang J, Xu RS, Yang SL. Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940 nm regions. Environ Monit Assess. 2009;157(1–4):459–69.

46. Olson RS, Cava WL, Zairah M, Akshay V, Moore JH. Data-driven advice for applying machine learning to bioinformatics problems. Pac Symp Biocomput. 2018;23:192–203.

47. Wu XY, Wang SH, Zhang YD. Survey on theory and application of k-Nearest-Neighbors algorithm. Comput Eng Appl. 2017;53(21):1–7.

Publisher's Note