Hyperspectral Estimation of Canopy Leaf Biomass Phenotype per Ground Area Using a Continuous Wavelet Analysis in Wheat

Xia Yao[1], Haiyang Si[1], Tao Cheng[1], Min Jia[1], Qi Chen[2], YongChao Tian[1], Yan Zhu[1]*, Weixing Cao[1], Chaoyan Chen[1], Jiayu Cai[1] and Rongrong Gao[1]

[1] National Engineering and Technology Center for Information Agriculture, Key Laboratory for Crop System Analysis and Decision Making, Ministry of Agriculture, Jiangsu Key Laboratory for Information Agriculture, Nanjing Agricultural University, Nanjing, China, [2] Department of Geography and Environment, University of Hawai'i at Mānoa, Honolulu, HI, United States

To extend agricultural productivity by knowledge-based breeding and tailoring varieties to adapt to specific environmental conditions, it is imperative to improve our ability to acquire the dynamic changes of the crop's phenotype under field conditions. Canopy leaf biomass (CLB) per ground area is one of the key crop phenotypic parameters in plant breeding. The most promising technique for effectively monitoring CLB is the hyperspectral vegetation index (VI). However, VI-based empirical models are limited by their poor stability and extrapolation difficulties when used to assess complex dynamic environments with different varieties, growth stages, and sites. It has been proven difficult to calibrate and validate some VI-based models. To address this problem, eight field experiments using eight wheat varieties were conducted during the period of 2003–2011 at four sites, and continuous wavelet transform (CWT) was applied to estimate CLB from large number of field experimental data. The analysis of 108 wavelet functions from all 15 wavelet families revealed that the best wavelet features for CLB in terms of wavelength (W) and scale (S) were observed in the near-infrared region and at high scales (7 and 8). The best wavelet-based model was derived from the Daubechies family (db), and was named db7 (W_{1197} nm, S_8). The new model was more accurate (R_v^2 = 0.67 and RRMSE = 27.26%) than a model obtained using the best existing VI (R_v^2 = 0.54 and RRMSE = 34.71%). Furthermore, the stable performance of the optimal db7 wavelet feature was confirmed by its limited variation among the different varieties, growth stages, and sites, which confirmed the high stability of the CWT to estimate CLB with hyperspectral data. This study highlighted the potential of precision phenotyping to assess the dynamic genetics of complex traits, especially those not amenable to traditional phenotyping.

Keywords: phenotypic parameter, canopy leaf biomass, continuous wavelet transform, optimal wavelet features, hyperspectral reflectance, wheat

INTRODUCTION

A key component to maintain or even increase agricultural production is therefore the development of genotyping and phenotyping technologies. Recently developed genomic approaches promise to further increase progress by breeding, while our ability to characterize the phenome of a plant has changed little. That means genomics has been advancing very rapidly, however, traditional plant phenotyping lags far behind current genotyping technique (Houle et al., 2010). This phenotyping bottleneck is of particular severity because many traits of biological and agricultural importance developed under a complex dynamic environment (Busemeyer et al., 2013). Many scientists devoted to relieve this bottleneck. They successfully developed the novel tool to represent the traditional phenotyping with low cost and high efficiency in greenhouse, for example, the high-throughput rice phenotyping facility (HRPF) (Yang et al., 2014). However, these results from controlled environment are far remove from the situation plants will experience in the field, and field conditions are notoriously hetero generous and the inability to control environmental factors makes results difficult to interpret and therefore, are difficulty to extrapolate to the filed (Araus and Cairns, 2014). Later, some novel field phenotyping systems with multi-sensor were developed for extracting the crop high-throughput phenotype properties (Bai et al., 2016; Pandey et al., 2017).

Biomass is one of the important crop phenotype traits, which are central to crop productivity. The commonly used hyperspectral vegetation index (VI) approach has been widely used to monitor crop canopy leaf biomass (CLB) at different scales due to its simplicity (Dong et al., 2003; Hese et al., 2005; Le Maire et al., 2008; Yan et al., 2013), but the non-invasive monitoring of biomass using these VIs has so far yielded only moderate prediction accuracies. For example, the VI $[(\rho_{NIR}/\rho_{Green}) - 1]$, which is based on specific hyperspectral bands, can predict the green leaf biomass in corn on the ground (Gitelson et al., 2003), while the spectral reflectance (SR) index (R_{900}, R_{680}) can predict aboveground biomass in wheat (Serrano et al., 2000). Le Maire et al. (2008) used the normalized difference vegetation index (NDVI) (R_{2160}, R_{1540}) to monitor canopy foliar biomass in forests. Hansen and Schjoerring (2003) used a VI (R_{708}, R_{565}) to predict green biomass. These classical VIs, with a simple formulation, are very convenient in practical applications. However, hyperspectral reflectance spectroscopy has not commonly been used in plant breeding now, a major limitation to the utility of hyperspectral data is variability in environmental conditions during measurement and the spectrum feature extraction of the crop phenotype traits (Furbank and Tester, 2011). Despite the high accuracy, those VI-based models have a weak stability due to the limited number of wavebands they used. The wavebands in published hyperspectral VIs are often selected through the mathematical optimization of millions of waveband combinations that have non-casual relationships with the absorption of dry matter (Curran, 1989; Qi et al., 1995; Kokaly and Clark, 1999; Luo et al., 2013). While, the underlying mechanism of the selected waveband features is difficult to explain from the statistical results based on a limited number of samples or a poor representation of growing conditions (Noh et al., 2006; Perez-Marin et al., 2007; Atzberger et al., 2010; Yi et al., 2010). An effective way to construct a highly sensitive VI or predictive model would be to establish a comprehensive database that could represent the large variability in treatment conditions (Le Maire et al., 2008). Therefore, it is often difficult to extrapolate CLB models over different growing conditions including crop varieties, years, and ecological sites. Consequently, it is necessary to adopt a new and effective analysis method to select sensitive features from more comprehensive samples for developing more stable CLB models.

A wavelet transform (WT), as a new emerging signal processing method, has been recently used to characterize the spectral information in the crop classification, estimation of forest leaf area index (LAI) and **canopy closure (CC) structural parameters, and the determination of leaf mass per unit area, plant leaf water content, and chlorophyll content (Koger et al., 2003; Pu and Gong, 2004; Cheng et al., 2010, 2011, 2014). Some previous studies (Koger et al., 2003; Pu and Gong, 2004) have focused on a discrete wavelet transform (DWT); however, the continuous wavelet transform (CWT) method could be accurately used to extract spectral features (Cheng et al., 2010, 2011, 2014). To date, the potential to undertake a wavelet analysis of hyperspectral reflectance with CWT for the crop biomass model has not been well documented. In addition, little is known regarding the selection of appropriate spectral features using the CWT method for crop biomass. Most existing studies have selected the wavelet function and wavelet features based on waveform similarity (Blackburn and Ferwerda, 2008; Bigerelle et al., 2013) or have focused directly on the most commonly used Mexican Hat wavelet function (Muraki, 1995; Torrence and Compo, 1998; Cheng et al., 2010, 2011, 2014). Whether the Mexican Hat is the best function for extracting hyperspectral information for crop biomass remains unknown.

Here, 108 wavelet functions from 15 wavelet families were investigated to select the sensitive wavelet features for CLB, which were used to develop robust models with hyperspectral data. These data were collected under different nitrogen rates, planting densities, and varieties of winter wheat in eight field experiments over 8 years at four sites. Such a comprehensive dataset has the potential to produce a reliable model for the spectroscopic estimation of CLB. To evaluate the performance of the new model, we compared the accuracy and stability of the prediction with the existing VI models under various growing conditions with different partitioning strategies.

MATERIALS AND METHODS

Experimental Design

Eight field experiments were conducted for eight consecutive years in this study. The experiment involved various nitrogen levels and density measurements, and included eight varieties of wheat grown at four sites. A randomized complete-block design was used with three replications per plot. For all treatments, sufficient $Ca(H_2PO4)_2$ and KCl were applied (150 kg ha^{-1}) prior to seeding. Crop management followed local standard

TABLE 1 | Summary of the conditions used in the eight wheat growth experiments.

Exp.	Site	Sowing time	Nitrogen rate (kg ha⁻¹)	Variety	Sampling dates	No. of samples	Function
1	JAAS	November 2, 2003	0, 75, 150, 225, 300	Ningmai 9 Huaimai 20 Yangmai 10 Xumai 26	April 8, April 20, May 4, May 17	127	Validation
2	JAAS	October 31, 2004	0, 75, 150, 225, 300	Ningmai 9 Yumai 34 Yangmai 10	March 19, April 13, April 23, May 6, May 19, May 25	219	Calibration
3	NAFB	November 3, 2005	0, 90, 180, 270	Ningmai 9 Yumai 34	March 30, April 11, April 20, April 29, May 19, May 24, June 2	136	Validation
4	JPF, NJAU	November 1, 2007	0, 90, 180, 270	Ningmai 9	March 11, March 25, April 18, April 25, May 6, May 20	126	Calibration
5	YCY	November 6, 2009	150, 225	Yangmai 16 Aikang 58	March 11, March 19, March 28, April 19, April 28, May 3, May 11, May 20	285	Calibration
6	YCY	November 5, 2009	0, 90, 180, 270	Yangmai 16	March 10, March 19, April 9, April 15	203	Validation
7	YCY	November 10, 2009	0, 75, 150, 225, 300	Ningmai 13 Yangmai 16	March 12, March 18, March 28, April 8, April 16, April 30	240	Calibration
8	YCY	November 9, 2010	150, 225	Yangmai 16	February 23, March 4, March 10, March 23, March 30, April 10, April 19, April 25, May 1, May 13, May 18	166	Calibration

The number after the variety indicates the time of breeding for the variety. JAAS, Jiangsu Academy of Agricultural Sciences (N32.03°, E118.87°); NAFB, Nanjing Agricultural and Forestry Bureau (N32.05°, E118.79°); JPF, NJAU, Jiangpu farm, Nanjing Agricultural University (N32.03°, E118.03°); and YCY, Yizheng county in Yangzhou (N32.32°, E119.30°).

TABLE 2 | Basic canopy leaf biomass (CLB) statistics for the eight experimental data sets (kg/m²).

Exp.	Min.	Max.	Mean	Std. Dev.	No. of samples
5	0.0096	0.240	0.088	0.0450	285
6	0.0154	0.190	0.098	0.0387	203
8	0.0184	0.301	0.105	0.0565	166
7	0.0185	0.298	0.107	0.0617	240
3	0.0278	0.290	0.109	0.0530	136
4	0.0349	0.218	0.113	0.0507	126
1	0.0537	0.330	0.164	0.0650	127
2	0.0507	0.327	0.183	0.0566	219

practices for wheat production. The detailed experimental design is described in **Table 1**.

Measurements

Canopy Leaf Biomass

After each measurement of canopy SR, an area of 0.25 m² (two rows × 0.5 m long) of wheat plants from each plot was collected for the determination of leaf dry biomass per unit ground area at the canopy scale (CLB, g DW m⁻²). For each sample, all green leaves were separated from the stems, oven-dried at 70°C to a constant weight, and then weighed.

Among the eight experiments, the average CLB values ranged from 0.0096 to 0.330 kg/m² with the lowest mean CLB being in EXP.5 and the highest in EXP.2. All standard deviations were less than 0.0650 kg/m², and the variance of the CLB in intergroup experiments was less than 0.05 kg/m² (**Table 2**).

Canopy Hyperspectral Reflectance

All canopy hyperspectral measurements were made using an ASD FieldSpec Pro spectrometer (Analytical Spectral Devices, Boulder, CO, United States). This spectrometer is fitted with 25° field of view fiber optics operating in the 350–2500 nm spectral region 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. The measurements were conducted at a height of 1.0 m above the vertical canopy (the height of wheat was 75–90 cm at maturity) and with a 0.44 m view diameter under clear sky conditions between 10:00 and 14:00 (Beijing local time).

Measurements of reflectance values were acquired at 10 sampling sites in each plot, with each sample observation averaging 20 scans at the optimized integration time. The resulting spectrum file contained a continuous SR at 1 nm steps over the band region of 350–2500 nm. A white panel SR value was taken before and after the vegetation measurement, with two scans obtained each time. In each experiment, data were obtained at several major growth stages, as detailed in the description of the experimental method. The spectral regions 1350–1410, 1790–1950, and 2471–2500 nm were excluded from the spectral analysis due to the strong absorption of water in the atmosphere. Because the CWT method requires a continuous spectrum, the reflectance values in the above three regions were set to zero in this study to avoid interference from noise.

Continuous Wavelet Decomposition of Canopy Hyperspectral Spectra

Continuous Wavelet Transform (CWT)

Wavelets are mathematical functions that are used to dissect data into different frequency components, with each component having a resolution appropriate to its scale (Grinsted et al., 2004). It is a gradual multiscale refinement of the signal (function) through an expansion and shift operation (Mallat, 1989). A wavelet transform can be a discrete wavelet transform (DWT) or continuous wavelets transform (CWT). DWT can reduce the redundant information within a transformation, but it may miss useful signal information, and the superiority of continuous over discontinuous decomposition is possibly due to

TABLE 3 | List of 15 wavelet families encompassing 108 wavelet functions used in this study.

Name of the wavelet family	Short name	Wavelet function in the family	Number of wavelet function	Reference
1. Haar	Haar	Haar	1	Daubechies, 1992
2. Daubechies	Db	db2, db3, …, dbN, …, db20	19	
3. Symlets	Sym	sym2, sym3, …, symN, …, sym17	16	
4. Coiflets	Coif	coif1, coif2, …, coifN, …, coif5	5	
5. Biorthogonal	Bior	bior Nr.Nd (Nr = 1, Nd = 1, 3, 5; Nr = 2, Nd = 2, 4,6,8; Nr = 3, Nd = 1, 3, 5, 7, 9; Nr = 4, Nd = 4; Nr = 5, Nd = 5; Nr = 6, Nd = 8)	15	
6. Reverse Bior	Rbio	rbio Nr.Nd (Nr = 1, Nd = 1, 3, 5; Nr = 2, Nd = 2, 4,6,8; Nr = 3, Nd = 1, 3, 5, 7, 9; Nr = 4, Nd = 4; Nr = 5, Nd = 5; Nr = 6, Nd = 8)	15	
7. Meyer	Meyer	Meyr	1	
8. Mexican hat	Mexh	Mexh	1	
9. Morlet	Morl	Morl	1	
10. Complex Gauusian	Cgau	cgau1, cgau2, …, cgauN, …, cgau8	8	
11. Complex Shannon	Shan	shan Fb-Fc (Fb = 1, Fc = 0.1, 0.5, 1, 1.5; Fb = 2, Fc = 3)	5	Teolis, 1998
12. Complex Frequency B-Spline	Fbsp	fbsp M-Fb-Fc (M = 1, Fb = 1, Fc = 0.5, 1, 1.5; M = 2, Fb = 1, Fc = 0.1, 0.5, 1)	5	
13. Complex Morlet	Cmor	cmor Fb-Fc (Fb = 1, Fc = 0.4, 0.5, 1, 1.5; Fb = 2, Fc = 0.1, 0.5)	5	
14. Dmeyer	Dmey	Dmey	1	Addison, 2002
15. Gaussian	Gaus	gaus1, gaus2, …, gausN, …, gaus8	8	Bernardino and Santos-Victor, 2005

the greater amount of spectral detail (variation with wavelength) (Blackburn and Ferwerda, 2008). Hence, we used the CWT method in this study. Equation (1) is the functional formula of the wavelet coefficient of the CWT:

$$C(a, b; f(t), \varphi(t)) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{a}} \varphi^* \left(\frac{t - b}{a} \right) dt \quad (1)$$

In this study, C is the wavelet coefficient after transformation that expresses the similarity between the original spectra and the wavelet function under a specific scaling and translation; a is the scaling factor (scale), we selected a power ($a = 3, 4, 5..., 8$, which represents $2^3, 2^4, 2^5..., 2^8$); b is the shifting factor (displacement); $f(t)$ represents the SR of each wavelength; $\varphi(t)$ is the wavelet basis function (wavelet mother function), depending on the parameters of ^{a} and ^{b}, that contains 108 wavelet functions from 15 wavelet families (**Table 3**); and t represents each wavelength ($t = 350, 351, 352..., 2,500$ nm). All of the wavelet coefficients were used to estimate the CLB.

Determination of the Optimal Wavelet Function, Wavelength, and Scale

Figure 2 displays the procedure for determining the optimal wavelet function, the best wavelength, and scale. First, the wavelet function was selected, and the CWT then proceeded to obtain the wavelet coefficient. A linear regression model was then established between the transformed wavelet coefficient and the biomass. The top 1% of coefficient of determination (R_c^2) values were determined, The model was validated with the dependent data, and we then defined the sensitive wavelength and scale for the top 1% of R_c^2 and R_v^2 values (R_c^2 and R_v^2 are the coefficient of determination for calibration and validation, respectively), and

the relative root mean square error (RRMSE). We repeated this procedure for all 108 wavelet functions from the 15 families. After systematically comparing the performance of the calibration and validation, the best wavelet mother function was determined using a box plot. A box plot was used in this study to select the optimal wavelet function, in which the outliers can describe the optimal and the poor wavelet functions (Grubbs, 1969; Pierce and Chick, 2013).

Calibration and Validation of the Model

Calibration data from experiments 2, 4, 5, 7, and 8, including information obtained from the various varieties, growth stages, and sites, were selected to construct the CLB model, while data from the remaining three experiments (1, 3, and 6) were used to test the constructed model equation. This partitioning strategy ensured a good representation of the entire data set with calibration samples and an approximate ratio of 2:1 for the number of calibration ($n = 1036$) and validation ($n = 466$) samples.

In addition to R_c^2 and R_v^2, the relative root mean square error (RRMSE) and stand error (SE) were used to evaluate the fit between the predicted and observed data along with a 1:1 plotting of the two sets of values. The RRMSE was calculated using the following equation (Yao et al., 2010):

$$RRMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (P_i - O_i)^2 \times \frac{100\%}{\overline{O_i}}} \quad (2)$$

where, P_i is the predicted biomass value of the model, O_i is the biomass value of an observation, n is the sample number,
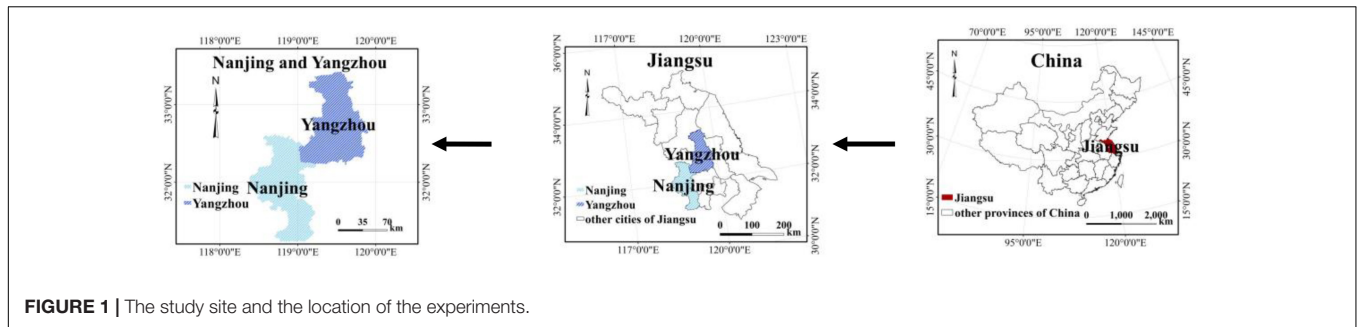
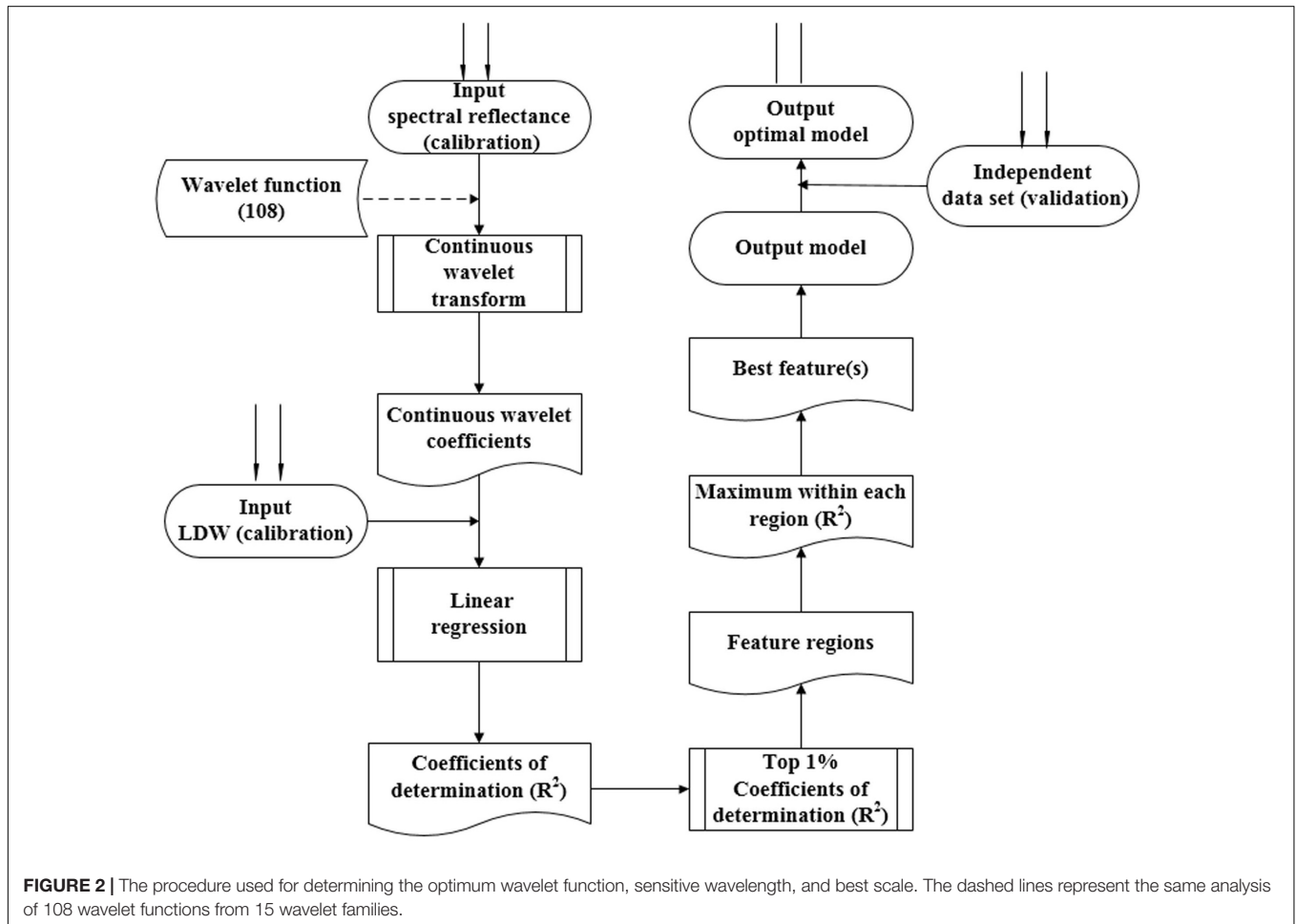FIGURE 1 | The study site and the location of the experiments.

FIGURE 2 | The procedure used for determining the optimum wavelet function, sensitive wavelength, and best scale. The dashed lines represent the same analysis of 108 wavelet functions from 15 wavelet families.

and $\overline{O_i}$ is the mean of the validation biomass data. All of these procedures were completed with self-programmed software based on MATLAB 8.1 (The MathWorks, 2013) and SPSS 20.0 software.

RESULTS

Selection of the Optimal Wavelet Function

To determine the best wavelet features [wavelength (W) and scale (S)] for all 108 wavelet functions, the occurrence score for every feature (W and S) was counted and is shown in **Figure 3**. The maximum score was 33, while the minimum score was nearly 0, which indicated that not all the wavelet functions had the same features (W and S), and different wavelet functions may have special features for each optimum model. The best features were mostly frequently observed at the scale of 7 and 8, and the sensitive wavelength was found to be located in the near-infrared region (780–1,350 nm). Therefore, when estimating CLB based on CWT, a near-infrared wavelength (780–1,350 nm) and at scale of 7 or 8 would be most effective.
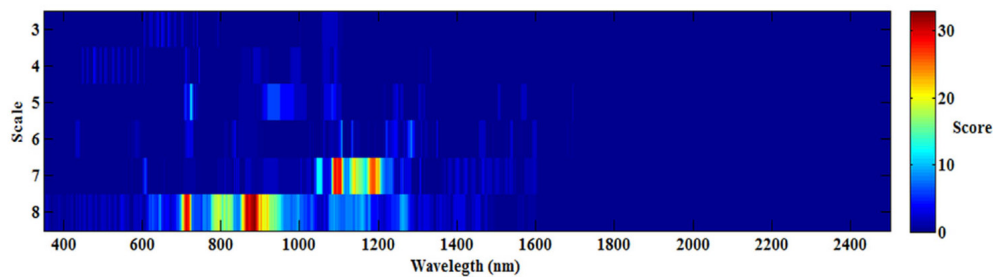
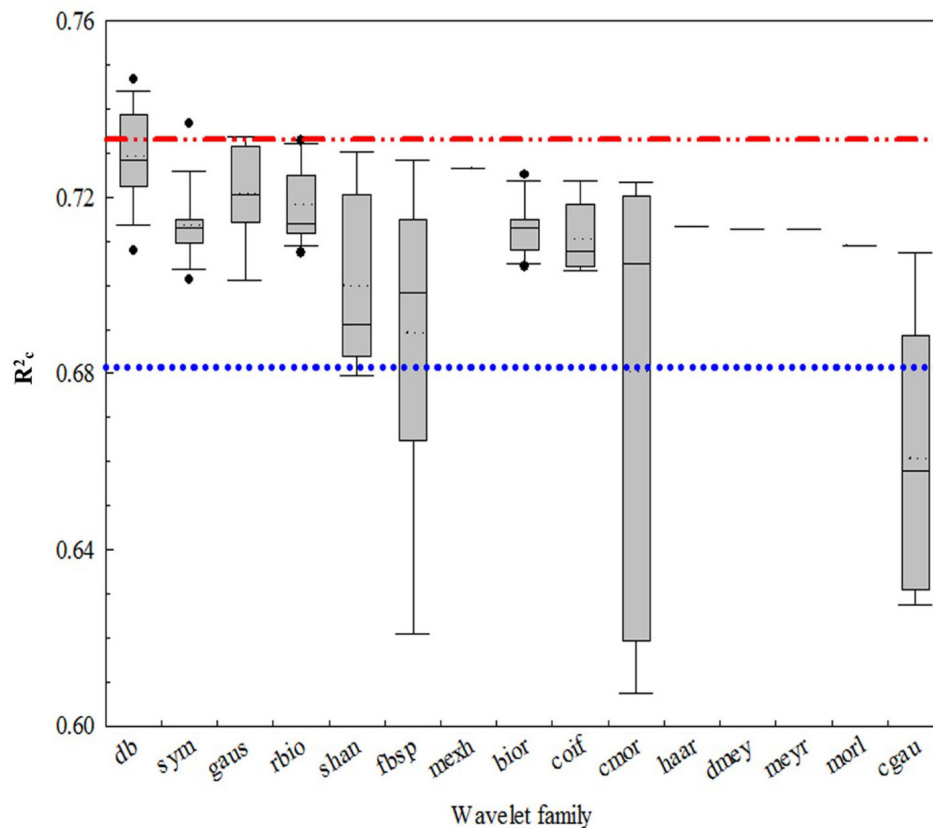**FIGURE 3 |** Occurrence score of all the wavelet functions under various wavelengths and scales.

FIGURE 4 | The coefficient of determination of calibration (R_c^2) values for the canopy leaf biomass (CLB) models for 15 wavelet families selected using a box plot. The blue and red lines are the interval lines of the general performance of the wavelet function. The region above the red line is the optimum wavelet function and the region below the blue line is the area of poor wavelet function.

Figure 4 shows that all of the coefficients of determination (R_c^2) for the CLB models from the 15 wavelet families were over 0.62 with the Daubechies (db) families being the highest (R_c^2 from 0.707 to 0.747). Among the 108 wavelet functions, the best functions were db7, db16, db17, db8, db6, db20, sym3, db3, gaus3, and db4, and the worst were shan1-1, fbsp1-1-1.5, fbsp2-1-1, cmor1-1, cmor-1.5, cgau4, cgau5, cgau6, cgau7, and cgau8. The wavelet function with the highest R^2 value was db7 ($R_c^2 = 0.75$ and SE = 0.032 kg/m^2), while the commonly used mexh was 0.73 and 0.034 kg/m^2, respectively.

Determination of the Optimal Wavelet Features and Models

All the data for db7 were processed with CWT. **Figure 5A** shows the correlation scalogram for the linear regression between the CLB and db wavelet coefficient at the scale from 3 to 8 and at wavelengths from 350 to 2500 nm. The top 1–5% of R_c^2 values were extracted and are shown in **Figure 5B**. Five featured regions (1,099–1122 and 1,194–1,216 nm at scale 7; 724–738, 878–905, and 1,173–1,210 nm at scale 8) were identified using the calibration data, which were identified at higher scales
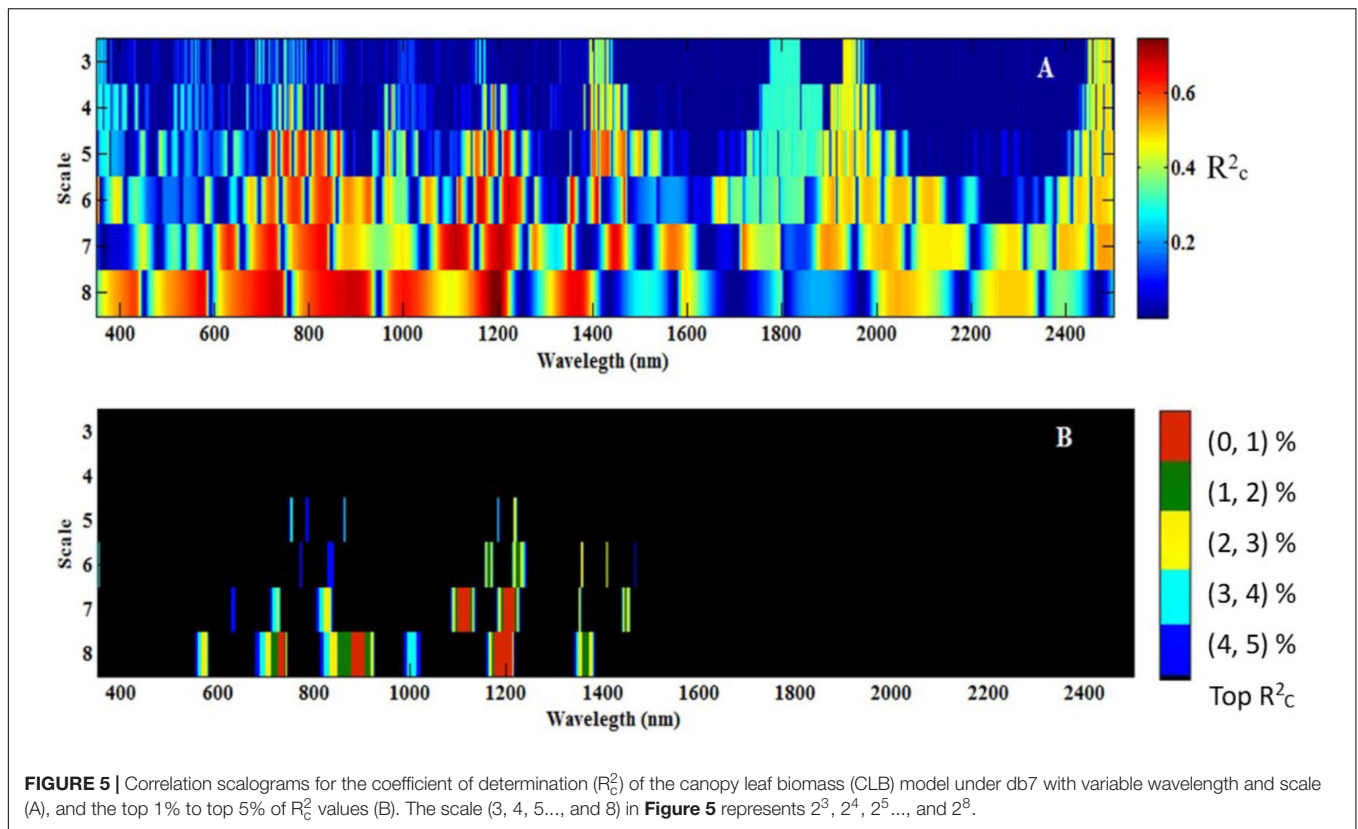
FIGURE 5 | Correlation scalograms for the coefficient of determination (R^2_c) of the canopy leaf biomass (CLB) model under db7 with variable wavelength and scale (A), and the top 1% to top 5% of R^2_c values (B). The scale (3, 4, 5..., and 8) in **Figure 5** represents 2^3, 2^4, 2^5..., and 2^8.

TABLE 4 | Assessment of the five best db7 wavelet features in the estimation of canopy leaf biomass (CLB).

Wavelet Feature	Equation	Calibration (N = 1036)		Validation (N = 466)	
		R^2_c	SE (kg/m^2)	R^2_v	RRMSE (%)
db7 (W_{1197}, S_8)	Y = 0.428x+0.070	0.75	0.032	0.67	27.26
db7 (W_{1206}, S_7)	Y = 0.375x−0.002	0.69	0.036	0.68	23.86
db7 (W_{1111}, S_7)	Y = −0.362x+0.033	0.69	0.036	0.68	23.13
db7 (W_{732}, S_8)	Y = 0.292x+0.009	0.69	0.036	0.73	20.79
db7 (W_{894}, S_8)	Y = −0.199x+0.017	0.68	0.036	0.72	20.91

(7 or 8) and in the near-infrared region, except for the range of 724–738 nm (**Figure 5B**, top 1%).

In this study, the wavelet feature was determined according to the R^2_c value and the absorption principle. Therefore, five wavelet features [(W_{732}, S_8), (W_{894}, S_8), (W_{1111}, S_7), (W_{1197}, S_8), and (W_{1206}, S_7)] were selected with db7 (W_{1197}, S_8) being the optimum due to it having the highest R^2_c and db7 (W_{894}, S_8) being the worst (**Table 4**).

Furthermore, we used the control variable method to qualitatively analyze which wavebands affected the wavelet feature for db7 (W_{1197}, S_8) and the commonly used mexh (W_{1412}, S_8). The convolution algorithm was used to transform the original spectrum into the wavelet power on a specific wavelet function. Therefore, the wavelet power would be influenced by the wavelength, shape, and scale of a certain wavelet function. As consequence, the wavelet feature could be impacted by the neighboring region of the characteristic wavelength. We found

that the features of db7 (W_{1197}, S_8) were mainly affected by the wavelengths of 995, 1187, and 1322 nm (**Figure 6A**). The features of mexh (W_{1412}, S_8) were mainly affected by the wavelengths of 941, 1325, and 1538 nm (**Figure 6B**).

Considering the R^2_c, SE, and the sensitive wavelength, the db family with a wavelength of 1197 nm and at scale of 8, i.e., db7 (W_{1197}, S_8), was determined to be the optimal wavelet function and feature for constructing the CLB model, and calibration and validation of db7 (W_{1197}, S_8) are shown in **Figure 7**.

Comparison of the CLB Models With Previous Models

To determine whether the new model was comparable to previously reported CLB models for wheat, the data collected in this present study were applied to compare the performance of the model for db7 (W_{1197}, S_8) with its performance for mexh (W_{1412}, S_8), NDVI$_{Bleaf}$, and RVI$_{GBM}$ (**Table 5**).
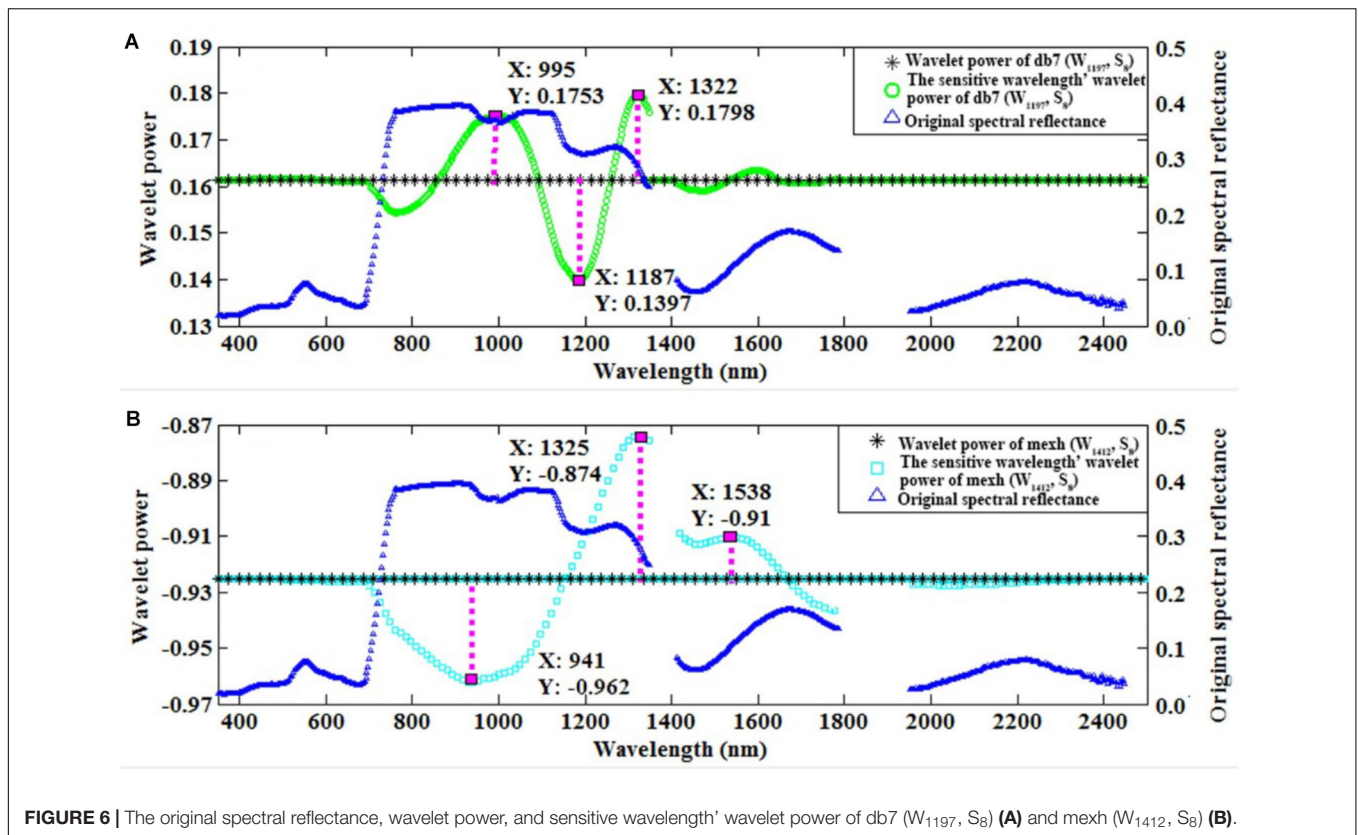
FIGURE 6 | The original spectral reflectance, wavelet power, and sensitive wavelength' wavelet power of db7 (W_{1197}, S_8) **(A)** and mexh (W_{1412}, S_8) **(B)**.
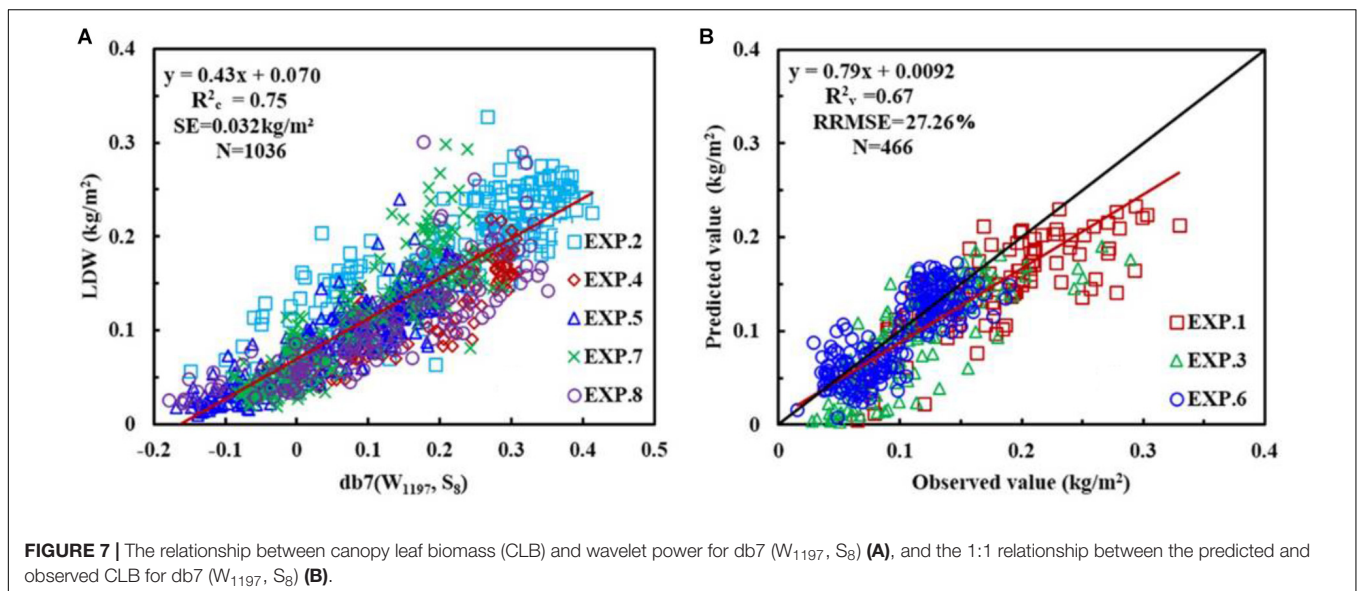

FIGURE 7 | The relationship between canopy leaf biomass (CLB) and wavelet power for db7 (W_{1197}, S_8) **(A)**, and the 1:1 relationship between the predicted and observed CLB for db7 (W_{1197}, S_8) **(B)**.

The results showed that the best model was db7 (W_{1197}, S_8), which had a high accuracy and low predictive performance ($R_c^2 = 0.75$, $R_v^2 = 0.67$ and RRMSE = 27.26%), that was slightly higher than that of the commonly used mexh function (W_{1412}, S_8; $R_c^2 = 0.73$, $R_v^2 = 0.68$ and RRMSE = 23.63%). It also had a better performance than the existing indices $NDVI_{Bleaf}$ ($R_c^2 = 0.62$, $R_v^2 = 0.54$ and RRMSE = 34.71%) and VI_{GBM} ($R_c^2 = 0.50$, $R_v^2 = 0.36$ and RRMSE = 34.38%) (**Table 5**).

DISCUSSION

Stability and Extrapolation of the New CLB Model

In this study, 1502 comprehensive samples were used to compare the stability and extrapolation of a newly developed CLB model with an existing VI model (**Table 6**). The abundance of

TABLE 5 | Comparison of the model performance for db7 (W_{1197}, S_8) with its performance for the commonly used mexh (W_{1412}, S_8), NDVI$_{Bleaf}$, and RVI$_{GBM}$.

Spectral feature		Calibration (N = 1036)		Validation (N = 466)		Reference
		R_c^2	SE (kg/m^2)	R_v^2	RRMSE (%)	
Wavelet feature						
db7	(W_{1197}, S_8)	0.75	0.032	0.67	27.26	This paper
Mexh	(W_{1412}, S_8)	0.73	0.034	0.68	23.63	This paper
Spectral index						
NDVI$_{Bleaf}$	(R_{2160}, R_{1540})	0.62	0.040	0.54	34.71	Le Maire et al., 2008
RVI$_{GBM}$	(R_{708}, R_{565})	0.50	0.045	0.36	34.38	Qi et al., 1995

TABLE 6 | Comparison of the stability and extrapolation potential for models based on db7 and NDVI$_{Bleaf}$ when categorizing samples using the growth stage, site*, variety, and year.

Grouping variable	Sub-group	Validation				NDVIBleaf
		Sample number	Db7 NDVI$_{Bleaf}$			
			R_v^2	RRMSE (%)	R_v^2	RRMSE (%)
Anthesis	After	215	0.57	30.42	0.44	38.24
	Before	251	0.76	23.51	0.63	31.46
	Mean		0.66	26.97	0.53	34.85
Site	Site 1	127	0.68	17.84	0.51	21.80
	Site 2	136	0.64	34.77	0.45	45.64
	Site 4	203	0.67	25.35	0.70	27.50
	Mean		0.67	25.99	0.55	31.64
Variety	Huaimai 20	41	0.60	21.81	0.52	22.70
	Ningmai 9	108	0.74	25.43	0.54	35.66
	Xumai 26	28	0.70	14.40	0.53	21.87
	Yangmai 10	18	0.80	13.51	0.53	13.19
	Yumai 34	68	0.67	36.40	0.50	48.24
	Yangmai 16	203	0.67	25.35	0.70	27.50
	Mean		0.70	22.82	0.55	28.19
Year	2003–2004	127	0.68	17.84	0.51	21.80
	2005–2006	136	0.64	34.77	0.45	45.64
	2009–2010	203	0.67	25.35	0.70	27.50
	Mean		0.67	25.99	0.55	31.64
Total	Mean		0.67	25.44	0.55	31.58

*Site 1: JAAS; Site 2: NAFB; Site 4: YCY.

experimental data enabled an authoritative assessment of the CLB model performance to be made. We categorized the samples into four sub-groups (growth stages, sites, varieties, and years). In the validation of the new model, the R_v^2 value was higher than that of the existing VIs, while the RRMSE was lower than that of the VI-derived models. This indicates that the new model based on db7 (W_{1197}, S_8) was very stable and could be effectively extrapolated across a diverse range of growth stages, sites, varieties, and years. The results were consistent with those of Cheng et al. (2014), who also found that the transferability of the wavelet-based predictive model to the entire measured database was either better than or comparable to a VI-derived model.
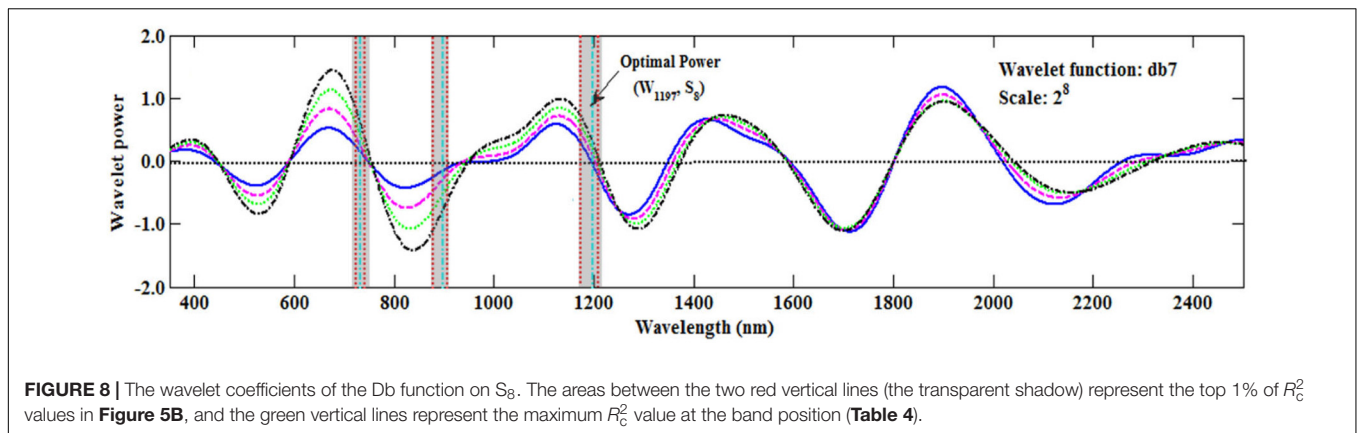
In addition, we noticed that the performance of the new model before anthesis was much better than in the stage after anthesis. The reason for this may be that the canopy cover in the later growth stage (after anthesis) was affected by panicle anthesis or grain development, which may increase the background noise. The model had a similar accuracy among sites and years and therefore could be extrapolated to different sites or years. However, there was less stability among the varieties and we therefore speculated that the plant type or structure among the different varieties influenced the reflectance. It should be considered how to reduce the impact of this issue in the future studies.

The Reason for the Higher Stability and Extrapolation Potential
Sensitive Wavelengths for Monitoring CLB
Previous studies have constructed the following VIs: NDVI (R_{2160}, R_{1540}) (Le Maire et al., 2008), [(ρ_{NIR} / ρ_{Green})-1] (Gitelson et al., 2003), SR (R_{900}, R_{680}) (Serrano et al., 2000), and VI$_{GBM}$

FIGURE 8 | The wavelet coefficients of the Db function on S_8. The areas between the two red vertical lines (the transparent shadow) represent the top 1% of R_c^2 values in **Figure 5B**, and the green vertical lines represent the maximum R_c^2 value at the band position (**Table 4**).

(R_{708}, R_{565}) (Qi et al., 1995) to predict crop biomass. However, the sensitive wavebands of 565, 680, 708, 900, 1540, and 2160 nm are not the core wavelengths of absorption, with the exception of 900 nm (the absorption band of protein) and 1540 nm (the absorption band of starch and cellulose) (Curran, 1989). In this study, the results show that db7 (W_{1197}, S_8) was mainly affected by the wavelengths of 995, 1187, and 1322 nm. The wavelength of 995 nm induces bending of the O-H bond and is the absorption peak of starch (Curran, 1989). In addition, 1187 and 1197 nm are close to 1200 nm, which is the absorption wavelength of water, cellulose, starch, and lignin (Curran, 1989; Qi et al., 1995; Bernardino and Santos-Victor, 2005; Cheng et al., 2014). Each of 1187, 1197, and 1322 nm are located in the near-infrared region, which is sensitive to the LAI (Pu and Gong, 2000). Therefore, db7 (W_{1197}, S_8) could provide information about LAI and LB. Fortunately, CLB is equal to the product of LAI and LB. Therefore, it is sensible to include the sensitive regions of both LAI and LB when selecting the sensitive region of CLB.

Most studies have used the sensitive wavelength at which the wavelet coefficient is at a maximum (peak) or minimum (valley) (Koger et al., 2003). However, in this study the sensitive wavelength region at the zero value of the wavelet coefficient was found to be better than that of the peak or valley (**Figure 8**), which was similar to the result reported by Cheng et al. (2010). This would provide a new method to determine the sensitive wavelength. Therefore, in the future we should pay more attention to those wavelengths in which the wavelet coefficients are close to zero.

Optimal Wavelet Function and Scales

Previous studies have commonly used the function of mexh, db, haar, and bior to extract spectral features for mapping forest LAI and CC (Pu and Gong, 2004), detecting insect damage (Cheng et al., 2010), and plant leaf water content (Cheng et al., 2011). In addition, they implied that predictive accuracy varied according to the choice of wavelet function in some cases; hence, a judicious choice of wavelet function may be necessary. Our study confirmed this viewpoint. **Figure 4** shows that the choice of mother wavelet could greatly affect the efficacy of the wavelet-based feature when changing the mother wavelet. Blackburn and Ferwerda (2008) selected the optimal wavelet

function, and found that bior 1.3 and rbior 5.5 at scale 6 were the best wavelet features when estimating chlorophyll concentration using a CWT and DWT. However, they only focused on 53 individual wavelet functions at the scale of 0–8. In this study, the best wavelet function and features for CLB were determined based on the 108 individual wavelet functions at the scale of 2^3–2^8.

Koger et al. (2003) set an accuracy threshold to select the optimal mother wavelet function, and found that only Haar, db5, db10, bior 2.2, bior 2.4, bior 2.6, bior 2.8, bior 6.8, sym2, and sym7 could qualify the accuracy, with the very simple Haar mother wavelet being the best. However, in our study we determined that the db7 families produced the best performance, which differed from previous results. Therefore, the mother wavelet function should first be applied when the CWA method is used, and the application of commonly used mother wavelet functions should not be taken for granted.

The analysis in this study incorporated many wavelet functions. There were some similar performances between the wavelets within each family, which may account for the lack of a significant statistical difference. This has commonly been the case in previous research. Therefore, it may be possible to develop a new wavelet function specifically for this application, and there are precedents for this approach (e.g., the MATLAB wavelet toolbox) (Blackburn and Ferwerda, 2008). In addition, it may be useful to decompose spectra using a group of wavelet functions that perform well overall, deriving a series of predictive regression models and obtaining an average value of the estimated parameters.

CONCLUSION

A comprehensive analysis of 108 individual wavelet functions from all 15 wavelet families revealed that the best wavelet features for the CLB were mostly located in the near-infrared region and at high scale (7 or 8). The best wavelet-based model was derived from the db family and was named db7 (W_{1197} nm, S_8). It was affected by the neighboring wavelengths of 995, 1187, and 1322 nm. The new model had a better accuracy than that obtained using the existing VIs. Furthermore, the more stable performance

with the db7 wavelet feature was further confirmed by the lack of variation in the R_c^2 and R_v^2 values across different varieties, growth stages, sites, and years. The main outcomes of this study were: (1) the use of 108 CWT methods to select the wavelet function for crop biomass from the hyperspectral data, (2) the identification of the wavelengths in the zero 38 region of the wavelet coefficient, (3) the development of a more stable and robust crop biomass model based on CWT, and (4) the resolution of some of the problems associated with the existing methods. The use of the CWT method would provide theoretical and technical support for the monitoring of crop growth parameters.

However, we only tested the one-dimensional wavelet decomposition functions available in the MATLAB package due to the limited amount of original data. In the future, two-dimensional WTs should be studied to assess the suitability of hyperspectral image data for extracting spatial information (such as texture and shape). In addition, it should be determined whether a CWT or DWT is a more accurate biomass estimation method. Moreover, whether the same structure or physical parameter would have a similar wavelet function and features for vegetation should also be checked.

REFERENCES

Addison, P. S. (2002). *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance.* Bristol: Institute of Physics Publishing. doi: 10.1887/0750306920

Araus, J., and Cairns, J. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Atzberger, C., Guérif, M., Baret, F., and Werner, W. (2010). Comparative analysis of three chemo metric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Comput. Electron. Agric.* 73, 165–173. doi: 10.1016/j.compag.2010.05.006

Bai, G., Ge, Y., Hussain, W., Baenziger, P., and Graef, G. (2016). A multi-sensor system for high throughput field phenotyping soybean and wheat breeding. *Comput. Electron. Agric.* 128, 181–192. doi: 10.1016/j.compag.2016.08.021

Bernardino, A., and Santos-Victor, J. (2005). "A real-time gabor primal sketch for visual attention," in *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis.* Estoril, 335–342. doi: 10.1007/11492429_41

Bigerelle, M., Guillemot, G., Khawaja, Z., El Mansori, M., and Antoni, J. (2013). Relevance of wavelet shape selection in a complex signal. *Mech. Syst. Sig. Proc.* 41, 14–33. doi: 10.1016/j.ymssp.2013.07.001

Blackburn, G. A., and Ferwerda, J. G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sens. Environ.* 112, 1614–1632. doi: 10.1016/j.rse.2007.08.005

Busemeyer, L., Ruckelshausen, A., Moller, K., Melchinger, A. E., Alheit, K. V., Maurer, H. P., et al. (2013). Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Nat. Sci. Rep.* 3, 2442–2446. doi: 10.1038/srep02442

Cheng, T., Rivard, B., and Sanchez-Azofeifa, A. (2011). Spectroscopic determination of leaf water content using continuous wavelet analysis. *Remote Sens. Environ.* 115, 659–670. doi: 10.1016/j.rse.2010.11.001

Cheng, T., Rivard, B., Sánchez-Azofeifa, G. A., Feng, J., and Calvo-Polanco, M. (2010). Continuous wavelet analysis for the detection of green attack damage due to mountain pine beetle infestation. *Remote Sens. Environ.* 114, 899–910. doi: 10.1016/j.rse.2009.12.005

Cheng, T., Rivard, B., Sánchez-Azofeifa, A. G., Féret, J., Jacquemoud, S., and Ustin, S. L. (2014). Deriving leaf mass per area LMA from foliar reflectance across a variety of plant species using continuous wavelet analysis. *ISPRS J. Photogramm. Remote Sens.* 87, 28–38. doi: 10.1016/j.isprsjprs.2013.10.009

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sens. Environ.* 30, 271–278. doi: 10.1016/0034-4257(89)90069-2

Daubechies, I. (1992). *Ten Lectures on Wavelets.* Philadelphia, PA: Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611970104

Dong, J., Kaufmann, R. K., Myneni, R. B., Tucker, C. J., Kauppi, P. E., Liski, J., et al. (2003). Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources, and sinks. *Remote Sens. Environ.* 84, 393–410. doi: 10.1016/S0034-4257(02)00130-X

Furbank, R. T., and Tester, M. (2011). Phonemics-technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* 30:1148. doi: 10.1029/2002GL016450

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlin. Process. Geophys.* 11, 561–566. doi: 10.5194/npg-11-561-2004

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21. doi: 10.1080/00401706.1969.10490657

Hansen, P. M., and Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. doi: 10.1016/S0034-4257(03)00131-7

Hese, S., Lucht, W., Schmullius, C., Barnsley, M., Dubayah, R., Knorr, D., et al. (2005). Global biomass mapping for an improved understanding of the CO2 balance-the Earth observation mission Carbon-3D. *Remote Sens. Environ.* 94, 94–104. doi: 10.1016/j.rse.2004.09.006

Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genetics.* 11, 855–866. doi: 10.1038/nrg2897

Koger, C. H., Bruce, L. M., Shaw, D. R., and Reddy, K. N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morning glory (*Ipomoea lacunosa*) in soybean (Glycine max). *Remote Sens. Environ.* 86, 108–119. doi: 10.1016/S0034-4257(03)00071-3

Kokaly, R. F., and Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67, 267–287. doi: 10.1016/S0034-4257(98)00084-4

Le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailler, J. Y., Bréda, N., et al. (2008). Calibration and validation of hyperspectral indices for the

estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* 112, 3846–3864. doi: 10.1016/j.rse.2008.06.005

Luo, J., Huang, W., Yuan, L., Zhao, C., Du, S., Zhang, J., et al. (2013). Evaluation of spectral indices and continuous wavelet analysis to quantify aphid infestation in wheat. *Precis. Agric.* 14, 151–161. doi: 10.1007/s11119-012-9283-4

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE PAMI* 11, 674–693. doi: 10.1109/34.192463

Muraki, S. (1995). Multiscale volume representation by a DoG wavelet. *IEEE Trans. Vis. Comput. Grap.* 1, 109–116. doi: 10.1109/2945.468408

Noh, H., Zhang, Q., Shin, B., Han, S., and Feng, L. (2006). A neural network model of maize crop nitrogen stress assessment for a multi-spectral imaging sensor. *Biosys. Eng.* 94, 477–485. doi: 10.1016/j.biosystemseng.2006.04.009

Pandey, P., Ge, Y., Stoerger, V., and Schnable, J. C. (2017). High throughput *in vivo* analysis of plant leaf chemical properties using hyperspectral imaging. *Front. Plant Sci.* 8:1348. doi: 10.3389/fpls.2017.01348

Perez-Marin, D., Garrido-Varo, A., and Guerrero, J. E. (2007). Non-linear regression methods in NIRS quantitative analysis. *Talanta* 72, 28–42. doi: 10.1016/j.talanta.2006.10.036

Pierce, R., and Chick, H. (2013). Workplace statistical literacy for teachers: interpreting box plots. *Math. Educ. Res. J.* 25, 189–205. doi: 10.1007/s13394-012-0046-3

Pu, R., and Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sens. Environ.* 91, 212–224. doi: 10.1016/j.rse.2004.03.006

Pu, R. L., and Gong, P. (2000). *Hyperspectral Remote Sensing and Applications.* Beijing: Higher Education Press, 10–12.

Qi, J., Moran, M. S., Cabot, F., and Dedieu, G. (1995). Normalization of sun/view angle effects using spectral albedo-based vegetation indices. *Remote Sens. Environ.* 523, 207–217. doi: 10.1016/0034-4257(95)00034-X

Serrano, L., Filella, I., and Penuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Sci.* 40, 723–731. doi: 10.2135/cropsci2000.403723x

Teolis, A. (1998). *Computational Signal Processing with Wavelets.* Rockville, MA: Birkhauser Boston.

Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78. doi: 10.1175/1520-0477(1998)079<0061: APGTWA>2.0.CO;2

Yan, F., Wu, B., and Wang, Y. (2013). Estimating aboveground biomass in Mu us sandy land using Landsat spectral derived vegetation indices over the past 30 years. *J. Arid Land Res.* 5, 521–530. doi: 10.1007/s40333-013-0180-0

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combing high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087

Yao, X., Zhu, Y., Tian, Y. C., Feng, W., and Cao, W. X. (2010). Exploring hyperspectral bands and estimation indices for leaf nitrogen accumulation in wheat. *Int. J. Appl. Earth Obs. Geoinf.* 12, 89–100. doi: 10.1016/j.jag.2009.11.008

Yi, Q., Huang, J., Wang, F., and Wang, X. (2010). Evaluating the performance of PC-ANN for the estimation of rice nitrogen concentration from canopy hyperspectral reflectance. *Int. J. Remote Sens.* 31, 931–940. doi: 10.1080/01431160902912061