



Retrieval of leaf chlorophyll content in Gannan navel orange based on fusing hyperspectral vegetation indices using machine learning algorithms

Suyun Lian¹  Lixin Guan^{1*}  Zhongzheng Peng²  Gui Zeng¹  Mengshan Li¹  Yin Xu¹ 

¹Intelligent Control Engineering & Technology Research Center, Gannan Normal University, 341000, Ganzhou, China. E-mail: lxguan@gnnu.edu.cn.

*Corresponding author.

²Nanjing University of Information Science and Technology, Nanjing, China.

ABSTRACT: Estimating leaf chlorophyll contents through leaf reflectance spectra is efficient and nondestructive. The literature base regarding optical indices (particularly chlorophyll indices) is wide ranging and extensive. However, it is without much consensus regarding robust indices for Gannan navel orange. To address this problem, this study investigated the performance of 19 published indices using RDS (raw data spectrum), FDS (first derivative data spectrum) and SDS (second derivative data spectrum) for the estimation of chlorophyll content in navel orange leaves. The single spectral index and combination of multiple spectral indices were compared for their accuracy in predicting chlorophyll a content (Chl_a), chlorophyll b content (Chl_b) and total chlorophyll content (Chl_{tot}) content in navel orange leaves by using partial least square regression (PLSR), adaboost regression (AR), random forest regression (RFR), decision tree regression (DTR) and support vector machine regression (SVMR) models. Through the comparison of the above data in three datasets, the optimal modeling data set is RDS data, followed by FDS data, and the worst is SDS data. In modeling with multiple spectral indices, good results were obtained for Chl_a (NDVI750, NDVI800), Chl_b (Datt, DD, Gitelson 2) and Chl_{tot} (Datt, DD, Gitelson2) by corresponding index combinations. Overall, we can find that the AR is also the best regression method judging by prediction performance from the results of single spectral index models and multiple spectral indices models. In comparison, result of multiple spectral indices models is better than single spectral index models in predicting Chl_a and Chl_{tot} content using FDS data and SDS data, respectively.

Key words: chlorophyll content, spectral index, regression, Gannan navel orange.

Recuperação do teor de clorofila de folhas em laranja do umbigo Gannan com base na fusão de índices de vegetação hiperespectral usando algoritmos de inteligência artificial

RESUMO: Estimar os teores de clorofila foliar através de espectros de refletância foliar é eficiente e não destrutivo, a base da literatura sobre índices ópticos (principalmente índices de clorofila) é ampla e extensa. No entanto, não há muito consenso sobre índices robustos para a laranja de Gannan. O estudo investigou o desempenho de 19 índices publicados usando RDS (espectro de dados brutos), FDS (espectro de dados de primeira derivada) e SDS (espectro de dados de segunda derivada) para a estimativa do teor de clorofila em folhas de laranja de umbigo. Os índices espectrais foram comparados quanto à sua precisão na previsão do teor de clorofila a ($Chla$), teor de clorofila b ($Chlb$) e teor de clorofila total ($Chltot$) em folhas de laranja de umbigo usando regressão dos mínimos quadrados parcial (PLSR), regressão adaboost (AR), modelos de regressão de floresta aleatória (RFR), regressão de árvore de decisão (DTR) e regressão de máquina de vetor de suporte (SVMR). Através da comparação dos dados acima em três conjuntos de dados, o conjunto de dados de modelagem ideal são os dados RDS, seguidos pelos dados FDS, e o pior são os dados SDS. Na modelagem com vários índices espectrais, bons resultados foram obtidos para $Chla$ (NDVI750, NDVI800), $Chlb$ (Datt, DD, Gitelson 2) e $Chltot$ (Datt, DD, Gitelson2) por combinações de índices correspondentes. No geral, podemos descobrir que o AR também é o melhor método de regressão a julgar pelo desempenho de previsão dos resultados de modelos de índice espectral único e modelos de índices espectrais múltiplos. Em comparação, o resultado de modelos de índices espectrais múltiplos é melhor do que os modelos de índices espectrais únicos na previsão do conteúdo de $Chla$ e $Chltot$ usando dados FDS e dados SDS, respectivamente.

Palavras-chave: teor de clorofila, índice espectral, regressão, Gannan laranja umbigo.

INTRODUCTION

Gannan navel orange is a national geographical indication product of China. It is of high quality, rich in essential nutrients and enjoys the reputation of Chinese famous fruit. As a major indicator of nutrients, chlorophyll content is involved

in various biochemical and physiological processes that are vital for navel orange production. Real-time and non-destructive assessment of navel orange chlorophyll content is important for evaluating crop productivity and improving the precise management of N(LI et al., 2016). Although, traditional measurement of chlorophyll content via wet chemistry methods

is precise but cost intensive, time-consuming and a destructive for measuring leaves (LIU et al., 2018). Therefore, traditional wet chemistry destructive method is obviously undesirable for assessing chlorophyll content in navel oranges. For this reason, hyperspectral remote-sensing technology can be applied to estimate the spatio-temporal variations in the physical and chemical parameters of vegetation, including the chlorophyll content, at a relatively low cost compared to field measurements (MUTANGA et al., 2004; ZHANG et al., 2008; HE, 2013; PU et al., 2014; PENG et al., 2018; HOEPPNER et al., 2020; JI et al., 2020; LI et al., 2020).

Hyperspectral remote-sensing methods for estimating vegetation chlorophyll content rely on observed spectral features via an empirical relationship that links the variables of interest to the sensitive bands, spectral indices, or spectral transform values (MEER, 2001; DORIGO et al., 2007; L. LIANG et al., 2012; LIANG et al., 2013). However, hyperspectral reflection is always affected by the biophysical characteristics of vegetation, canopy shape and structure, atmospheric absorption and scattering, and soil backgrounds (THENKABAIL et al., 2000; A et al., 2010). Therefore, to capture more effective information from the spectra, derivative spectra was applied to minimize the influences of background interference and spectral noise. Several studies showed that the first derivative reflectance (FDR) was sensitive to crop chlorophyll (CHO & SKIDMORE, 2006). Feng et al. (FENG et al., 2008) reported that the chlorophyll content was highly related to the ratio of the red-edge integral areas and the blue-edge integral areas. Liu et al. (LIU et al., 2002) also demonstrated that the first derivative spectrum (FDS) at 740–760 nm was highly correlated with chlorophyll content in rice.

In the field of hyperspectral, given the spatial resolution, neighbor intensities are highly correlated so a spectral signature contains a lot of redundant information. The main issue consists of extracting the discriminative information to reduce the set of relevant bands (LE BRIS et al., 2015). Spectral indices are an important method for extracting information from remotely sensed data because indices reduce, but do not eliminate, effects of soil, topography, and view angle (RDJ & ARH, 1991; HATFIELD & PRUEGER, 2013; HATFIELD et al., 2015). Identically, spectral indices are also an important method for analyzing imaging spectrometer data (GITELSON et al., 2006). Several studies have successfully estimated chlorophyll content in leaves based on their own study dataset using different

vegetation indices. Extensive use has been made of visible ratios (BISUN & DATT, 1998), visible/NIR ratios (HABOUDANE et al., 2002; GITELSON et al., 2006), red edge reflectance-ratio indice (SIMS & GAMON, 2002), spectral and derivative red edge indices (MILLER et al., 1990). Combinations of the red edge and infrared region built a spectral index for measuring chlorophyll contents (ZARCO-TEJADA et al., 2003; MAIN et al., 2011). In addition, BROGE & LEBLANC (BROGE & LEBLANC, 2001) developed the triangular vegetation index (TVI) based on the area of a triangle with vertices at green, red and NIR wavelengths, which is sensitive to both chlorophyll content and LAI. However, most VIs have been focused on retrieving chlorophyll contents in specific species (MAIN et al., 2011; MALENOVSKY et al., 2013), which have a limitation and un-stability for common use in other species.

The research objects of this study are: (1) to find an indexes suitable for navel orange leaves because index may not be suitable for all species datasets (LE MAIRE et al., 2008); (2) to the best of our knowledge, derivative of the indexes has not yet been extensively applied in regression analysis, so we explored the capability of the derivative of the index in retrieval chlorophyll content; (3) to establish the relationship between the spectra data and chlorophyll content by using five different regression algorithms that are commonly used for data analysis in the field of remote sensing and comparative analysis the performances of these five regression methods.

MATERIALS AND METHODS

Data acquisition and processing

The study area was performed in a navel orange planting site stand located in National Navel Orange Engineering Research Center, Gannan Normal University, GanZhou, China. Sixty leaf samples were collected by the lopper from the branches of the canopy with ten trees on June 23, June 29, July 03, and July 04, 2018, between 9:00 and 11:00 or 14:00 and 16:00, respectively. Leaf samples were all in a normal range of maturity and health. A laboratory hyperspectral imaging system (GaiaField-V10E, Sichuan Dualix Spectral Image Technology Co. Ltd, Sichuan, China) was used to acquire hyperspectral images of the leaf samples in reflectance mode. Imaging spectrograph (ImSpectorV10E) covers the spectral range from 400 to 1000nm with 360 wavebands (wavelength interval 1.67 nm). Three-point average spectral reflectance was used as the spectral value of each navel orange leaf sample.

The datasets of 60 leaf samples were randomly split into two parts: a calibration set of 45 samples and a validation set of 15 samples. First derivatives (FDS) and second derivatives (SDS) of the apparent absorbance spectra were calculated. The chlorophyll a content (Chl_a), chlorophyll b content (Chl_b) and total chlorophyll content (Chl_{tot}) of the leaf samples was measured by spectrophotometry. The absorbance at 665 nm (A_{665}) and 649 nm (A_{649}) was subsequently measured by V-5100 spectrophotometer. Chlorophyll content was calculated according to the reference using the correction equations (1), (2) and (3).

$$Chl_a = (13.95A_{665} - 6.80A_{649}) \times \frac{V}{W \times 1000} \quad (1)$$

$$Chl_b = (24.96A_{649} - 7.32A) \times \frac{V}{W \times 1000} \quad (2)$$

$$Chl_{tot} = Chl_a + Chl_b \quad (3)$$

where V is the volume of the extraction solution (ml), and W is the weight of the leaf sample (g). The results of Chl_a , Chl_b and Chl_{tot} were expressed in mg/g.

Modeling methods

Some commonly used machine learning algorithms, such as partial least square regression (PLSR), adaboost regression (AR), random forest regression (RFR), decision tree regression (DTR) and support vector machine regression (SVMR) can cope with the strong nonlinearity of the functional dependence between spectral variables and biophysical or biochemical parameters. In this study, PLSR, AR, RFR, DTR and SVMR were applied to learn the relationship between the vegetation indices (VIs) and chlorophyll content by fitting a flexible model directly from the spectrum datasets. K-fold cross validation is used to avoid model overfitting, and the value of K is 5. The grid search algorithm is used to find optimal model hyper parameters for 5 algorithms.

For AR parameters, the "base_estimator" is DecisionTree Regressor, the "loss" is linear, the "n_estimators" is 50 and the "learning_rate" is 1; for RFR, the "n_estimators" is 81, the "max_features" is 6; for DTR, the "criterion" is mse, the "splitter" is best, the "min_samples_leaf" is 3 and the "max_depth" is 30; for SVMR, the C and gamma are 9 and 0.1 respectively.

The coefficient of determination (R^2) and root mean square error (RMSE) are used as an evaluation indicators. Assuming that the relationship follows a normal distribution, R^2 ranges from 0 to 1. The higher the R^2 value is, the stronger the predicted relationship is. The lower the RMSE is, the better performance of the model is. Which calculate as follows:

$$R^2 = \frac{[\sum_{i=1}^n (y_{ai} - \bar{Y}_a) \cdot \sum_{i=1}^n (y_{mi} - \bar{Y}_m)]^2}{\sum_{i=1}^n (y_{ai} - \bar{Y}_a)^2 \cdot \sum_{i=1}^n (y_{mi} - \bar{Y}_m)^2} \quad (4)$$

Where \bar{Y}_a is the average predicted value and \bar{Y}_m is the average measured value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{ai} - y_{mi})^2}{n}} \quad (5)$$

Where y_{ai} is the output value; y_{mi} is the measured data; and n is the sample number.

RESULTS AND DISCUSSION

Spectral reflectance and leaf chlorophyll content

The apparent reflectance of spectra was subjected to first and second derivation in agreement (SHI et al., 2013) in order to increase signal-to-noise ratios at the different wavelengths. Higher signal-to-noise ratios obtained by derivation seem to have been able to highlight the parts of the spectral where important chemical information is located. They are expected to subsequently improve the regression models for chlorophyll content determination. The content of Chl_a measured by spectrophotometry ranged from 0.45mg/g to 2.73mg/g, Chl_b range from 0.08mg/g to 0.75mg/g and Chl_{tot} range from 0.53mg/g to 3.48mg/g.

Vegetation indices used in this study

A number of hyperspectral indices have been established to estimate chlorophyll content at both the leaf and canopy scales. In this study, using the leaf reflectance data, we calculated 19 published chlorophyll indices (Table1). In this paper, five superior indices used by LIU et al. (LIU et al., 2018), top five in the 20 spectral indices used by Main et al. (MAIN et al., 2011) and nine indices used by DIAN et al. (DIAN et al., 2015) were selected as spectral indices group.

Analysis of modeling results

In this experiment, the raw data spectrum (RDS), first derivative data spectrum (FDS) and second derivative data spectrum (SDS) were used to calculate the index, and then observed their performance on chlorophyll estimation. In addition, five different regression methods are applied for modeling, the main purpose is to get the best suitable regression method and indexes for navel orange chlorophyll content regression. The consistency and robustness of the various VIs in estimating leaf chlorophyll content was assessed in two different

Table 1 - Vegetation indices used in the study.

-----Spectral index-----		Formula	Source
I	1 BGI	R^{1450}/R^{1550}	ZARCO-TEJADA et al. 2003
	2 PRI_5	$R^{1550} * R^{1700} * (R^{1700} - R^{1550}) / R^{1670}$	LIU et al. 2011
	3 TVI	$[120 * (R^{1750} - R^{1550}) - 200 * (R^{1670} - R^{1550})] / 2$	BROGE AND LEBLANC 2001
	4 CARI	$(R^{1700} / R^{1670}) * [(R^{1670} * a + R^{1670} * b) / \text{SQRT}(a^2 + 1)]$ $a = (R^{1700} - R^{1550}) / 150; b = R^{1550} - 550 * a$	KIM et al. 1994
	5 CRI	$100 / R^{1510} - 100 / R^{1550}$	GITELSON et al. 2002
II	6 Datt	$(R^{1850} - R^{1710}) / (R^{1850} - R^{1680})$	DATT 1999
	7 DD	$(R^{1749} - R^{1720}) - (R^{1701} - R^{1672})$	LE MAIRE et al. 2004
	8 Gitelson 2	$(R^{1750} - R^{1800} / R^{1695} - R^{1740}) - 1$	GITELSON et al. 2003
	9 Vogelmann	R^{1740} / R^{1720}	VOGELMANN et al. 1993
	10 Vogelmann3	D_{715}^2 / D_{705}^2	VOGELMANN et al. 1993
III	11 NDVI ₇₅₀	$(R^{1750} - R^{1705}) / (R^{1750} + R^{1705})$	GITELSON & MERZLYAK 1994
	12 NDVI ₈₀₀	$(R^{1800} - R^{1680}) / (R^{1800} + R^{1680})$	RICHARDSON et al. 2002
	13 SR ₈₀₀	R^{1800} / R^{1680}	SIMS & GAMON 2002
	14 SR ₇₅₀	R^{1750} / R^{1705}	SIMS & GAMON 2002
	15 mND	$(R^{1750} - R^{1705}) / (R^{1750} + R^{1705} - 2R^{1445})$	SIMS & GAMON 2002
	16 mSR	$(R^{1750} - R^{1445}) / (R^{1705} - R^{1445})$	SIMS & GAMON 2002
	17 MCARI	$[(R^{1700} - R^{1670}) - 0.2(R^{1700} - R^{1550})] R^{1700} / R^{1670}$	HABOUDANE et al. 2002
	18 TCARI	$3[(R^{1700} - R^{1670}) - 0.2(R^{1700} - R^{1550}) R^{1700} / R^{1670}]$	HABOUDANE et al. 2002
	19 OSAVI	$(1 + 0.16)(R^{1800} - R^{1670}) / (R^{1800} + R^{1670} + 0.16)$	HABOUDANE et al. 2002

¹represents reflectance at wavelength X nm.

²represents the derivative of the reflectance spectrum at wavelength X nm.

ways, namely, using single index and using the combined of multiple spectral indexes.

Single spectral index modeling

The table 2 lists 45 indicators that perform best in specific data sets and inversion variables, for example, in PLSR algorithm, RDS data set and Chl_a variable, the index with the best performance among all indicators is TCARI with serial number 18 in table 1.

The horizontal line in the table is to highlight the data set with the best inversion effect in the same algorithm. Among the five machine algorithms, except for RFR algorithm, RDS data set got the best regression effect, followed by FDS, and finally SDS. However, when modeling with SDS data, the best estimate of Chl_b ($R^2=0.931$ and $RMSE=0.045$) in navel orange leaves is obtained. Using RDS data as the input vector of the model, AR shows great potential in chlorophyll content estimating for Chl_a ($R^2=0.955$ and $RMSE=0.139$) and Chl_{tot} ($R^2=0.966$ and $RMSE=0.150$), yet the estimate of Chl_b is a bit worse, but is still better than other modeling methods. In addition, R^2 value from Chl_a

and Chl_{tot} is better than Chl_b , the reason may be that these selected indices are not particularly sensitive to Chl_b . As can be seen from table 2, the regression effect of AR algorithm is the best, regardless of different data sets or different regression variables.

In order to find the spectral indices suitable for searching for Chl_a , Chl_b and Chl_{tot} in RDS data, FDS data and SDS data, we select the best index according to the frequency of index occurrence, and the index with the most occurrence is identified as the best index. Compared with other spectral indices, during all the regression methods, Vogelmann, which appeared 10 times in the RDS data were considered as the best spectral indices in the RDS data. Similarly, when modeling with FDS data and SDS data, DD and Vogelmann3 are the best modeling spectral indices respectively. To choose a spectral index that is excellent for the estimation of Chl_a , Chl_b and Chl_{tot} , Vogelmann3, DD and Vogelmann would be recommend respectively. Lastly, an index that is good for various modeling methods and various modeling data sets can be selected from, which is Vogelmann.

Table 2 - The performance of the vegetation indices to predict chlorophyll content (mg/g) according to the single index.

Method		-----Chl _a -----			-----Chl _b -----			-----Chl _{tot} -----		
		VIS_N	R ²	RMSE	VIS_N	R ²	RMSE	VIS_N	R ²	RMSE
PLSR	RDS	NO.18	0.88	0.22	NO.18	0.79	0.11	NO.18	0.87	0.30
	FDS	NO.10	0.85	0.25	NO.9	0.75	0.11	NO.10	0.84	0.33
	SDS	NO.7	0.82	0.27	NO.7	0.73	0.12	NO.7	0.81	0.36
AR	RDS	NO.9	0.92	0.14	NO.9	0.93	0.05	NO.9	0.97	0.15
	FDS	NO.9	0.95	0.14	NO.7	0.92	0.05	NO.9	0.95	0.19
	SDS	NO.10	0.94	0.16	NO.14	0.93	0.05	NO.16	0.95	0.18
RFR	RDS	NO.9	0.9	0.18	NO.9	0.80	0.08	NO.9	0.92	0.23
	FDS	NO.7	0.94	0.15	NO.7	0.88	0.06	NO.7	0.95	0.18
	SDS	NO.15	0.92	0.18	NO.15	0.86	0.07	NO.10	0.92	0.22
DTR	RDS	NO.17	0.91	0.18	NO.17	0.85	0.08	NO.17	0.91	0.24
	FDS	NO.7	0.88	0.21	NO.7	0.80	0.08	NO.7	0.89	0.27
	SDS	NO.10	0.88	0.21	NO.10	0.80	0.08	NO.9	0.88	0.27
SVMR	RDS	NO.6	0.94	0.16	NO.18	0.89	0.06	NO.18	0.95	0.19
	FDS	NO.10	0.93	0.17	NO.16	0.88	0.06	NO.16	0.94	0.19
	SDS	NO.10	0.93	0.17	NO.6	0.82	0.08	NO.6	0.90	0.26

*VIS_N represents the serial number of vegetation index in table1.

Multiple spectral index modeling

Combinations of different indices were used to estimate chlorophyll content, show as table 3.

The AR models built by the FDS data, RDS data and SDS data provided relatively robust results for predicting Chl_a (R² =0.966 and RMSE=0.121), Chl_b (R² =0.921 and RMSE=0.048) and Chl_{tot} (R² =0.969 and RMSE=0.144), compared with the other models. Through the comparison of the above data in these three datasets, there is not much difference between RDS and FDS in the regression effects, both of which are relatively excellent, and the worst is SDS data. Consistent with the conclusion of single index, AR is the algorithm with the best performance among the five algorithms. In modeling with multiple spectral indices, good results were obtained for Chl_a (NDVI750, NDVI800), Chl_b (Datt, DD, Gitelson2) and Chl_{tot} (Datt, DD, Gitelson2) by corresponding index combinations.

Comparative analysis of the optimal models

The prediction results for Chl_a, Chl_b and Chl_{tot} in both single spectral index and multiple spectral indices using the RDS data, FDS data and SDS data are shown in table 2 and table 3. It can be founded that the AR is also the best regression method judging by prediction performance from the results of single spectral index models and multiple

spectral indices models. In comparison, result of multiple spectral indices models is better than the single spectral index. In contrast, single spectral index models got better prediction results for Chl_b using the SDS data. In general, modeling with multiple spectral indices is suitable for the estimation of Chl_a and Chl_{tot}, and using single spectral index is suitable for the retrieval of Chl_b.

DISCUSSION

This study investigated the performance of 19 published indices using RDS data, FDS data and SDS data for the estimation of chlorophyll content in navel orange leaves. The aim was to understand which of the myriad of published VIs would be consistent or robust enough when applied to Gannan navel orange leaves. The indices varied greatly in terms of their original focus, and intended targets, but they were tested none-the-less and produced interesting results. We applied the indices to would provide a more than adequate examination of their abilities, due to the variety of leaf structures, leaf surfaces, moisture contents and chlorophyll contents present. Among the navel orange leaves, the Chl_{tot} content is the highest, followed by the Chl_a, and the least content is Chl_b. A common observation in the study was that AR appeared regularly in the top of the rankings for each

Table 3 - The performance of the vegetation indices to predict chlorophyll content (mg/g) according to the multiple index.

Method		-----Chl _a -----			-----Chl _b -----			-----Chl _{tot} -----		
		VIS_N	R ²	RMSE	VIS_N	R ²	RMSE	VIS_N	R ²	RMSE
PLSR	RDS	6/7	0.92	0.19	6/7 /8/9	0.86	0.08	6/7	0.91	0.26
	FDS	6/7 /8/9	0.92	0.19	6/7/8	0.88	0.11	6/7 /8/9	0.92	0.25
	SDS	2/3 /4/1	0.83	0.24	2/3 /4/1/5	0.73	0.10	2/3	0.82	0.33
AR	RDS	6/7	0.95	0.14	6/7 /8	0.92	0.05	6/7 /8	0.97	0.14
	FDS	11/12	0.97	0.12	11/12	0.91	0.05	11/12 /13	0.96	0.15
	SDS	6/7 /8/9	0.96	0.13	6/7 /8/9/10	0.88	0.06	6/7 /8/9	0.94	0.20
RFR	RDS	6/7/8	0.94	0.16	6/7/8 /9/10	0.88	0.06	6/7/8 /9/10	0.94	0.20
	FDS	6/7/8/ 9/10	0.93	0.17	6/7/ 8/9	0.89	0.06	6/7/8/ 9/10	0.94	0.20
	SDS	6/7	0.920	0.18	6/7	0.84	0.07	6/7	0.92	0.23
DTR	RDS	11/12/13 /14/15/1 6 /17/18	0.900	0.19	11/12/13/ /14/15/ 16/17	0.82	0.08	11/12/13/1 4/15/ 16/17/18	0.89	0.26
	FDS	2/3/4	0.87	0.23	2/3/4/1	0.80	0.28	2/3/5	0.88	0.29
	SDS	6/7/8 /9/10	0.85	0.24	6/7/ 8/9	0.71	0.11	6/7 /8/9	0.86	0.30
SVMR	RDS	6/7	0.87	0.23	6/7/8/9	0.81	0.08	6/7	0.90	0.26
	FDS	11/12/13 /14/15/1 6	0.93	0.17	11/12/13 /14/15/16/	0.89	0.06	11/12/13 /14/15/16	0.92	0.23
	SDS	6/7	0.89	0.21	6/7/ 8/9	0.85	0.07	6/7	0.89	0.27

*VIS_N represents the serial number of vegetation index in table 1.

of the scenarios. The phenological state of the leaves, as well as inherent differences between species, results in datasets with variable moisture contents, leaf surfaces, and leaf internal structures. The indices would have had different responses to these moisture and structure variations, which in turn could have influenced their ability to predict for chlorophyll content. It can be assumed that the best performing indices probably show a decreased sensitivity to varying leaf structures or moisture contents and can be considered more robust than indices that only did well for crop species. Results also have similarities to those reported in MAIN et al. (MAIN et al., 2011), where the data from various species to test the performance of 70 published chlorophyll indices was used. Some of the same indices that performed well

in the MAIN et al. (MAIN et al., 2011) study also perform well in this study (e.g. Vogelmann 3 index, Datt index and Vogelmann indices).

The question regarding whether there is one single index or multiple indices to use in order to estimate Gannan navel orange chlorophyll content has been answered in this paper. A number of robust and consistent indices for Chl_a, Chl_b and Chl_{tot} are proposed; and could therefore, be seen as priority indices to be tested in any follow up work. For instance the Vogelmann index by VOGELMANN et al. (VOGELMANN et al., 1993) consistently performed well in all modeling methods, and was the best performer for the low Chl_b. Further research is recommended regarding whether these indices be able to measure Gannan navel orange canopy

spectral. Another further study is to combine the actual spectrum with the radial transfer models (RTMs) to make the model simulation data set make up the shortcomings of the actual measurement data set.

CONCLUSION

Firstly, the regression of Chl_a , Chl_b and Chl_{tot} has achieved good results, thus it's available to use hyperspectral vegetation indices estimate the Chl_a , Chl_b and Chl_{tot} contents in navel orange leaves. Secondly, in 19 published spectral indices, modeling by the single index or multiple indices, a suitable index was founded for navel orange leaves. Finally, when RDS data, FDS data and SDS data are used as input vectors of the models, among the five modeling methods the best modeling method is AR, which can establish a quantitative monitoring model of chlorophyll content and control the nutritional status of navel orange fruit trees in real time.

ACKNOWLEDGEMENTS

This research was supported in part by the National Natural Science Foundation of China under Grant 61741103 and 42061067, in part by the Guangdong Natural Science Foundation of China under Grant 2017A030313343, in part by the Shenzhen Subject Layout Project of China under Grant 20180172, and in part by the Shenzhen Science and Technology Projection of China under Grant JCYJ20160520173822387.

DECLARATION OF CONFLICT OF INTEREST

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

AUTHORS' CONTRIBUTIONS

All authors contributed equally for the conception and writing of the manuscript. All authors critically revised the manuscript and approved of the final version.

REFERENCES

- A, C. H. J., et al. Estimating Leaf Chlorophyll Content Using Red Edge Parameters - ScienceDirect. *Pedosphere*, v.20, n.5, p.633-644. 2010. Available from: <[https://doi.org/10.1016/S1002-0160\(10\)60053-7](https://doi.org/10.1016/S1002-0160(10)60053-7)>. Accessed: Aug. 20, 2021. doi: 10.1016/S1002-0160(10)60053-7.
- BISUN; DATT. Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves. *Remote Sensing of Environment*. 1998. Available from: <[https://doi.org/10.1016/S0034-4257\(98\)00046-7](https://doi.org/10.1016/S0034-4257(98)00046-7)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(98)00046-7.
- BROGE, N. H.; E. LEBLANC. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, v.76, n.2, p.156-172. 2001. Available from: <[https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(00)00197-8.
- CHO, M.; A. SKIDMORE. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, v.101, p.181-193. 2006. Available from: <<https://doi.org/10.1016/j.rse.2005.12.011>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.rse.2005.12.011.
- DIAN, Y., et al. Influence of Spectral Bandwidth and Position on Chlorophyll Content Retrieval at Leaf and Canopy Levels. *Journal of the Indian Society of Remote Sensing*, v.44, n.4, p.1-11. 2015. Available from: <<https://doi.org/10.1007/s12524-015-0537-2>>. Accessed: Aug. 20, 2021. doi: 10.1007/s12524-015-0537-2.
- DORIGO, W. A., et al. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation & Geoinformation*, v.9, n.2, p.165-193. 2007. Available from: <<https://doi.org/10.1016/j.jag.2006.05.003>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.jag.2006.05.003.
- FENG, W., et al. Monitoring leaf pigment status with hyperspectral remote sensing in wheat. *Australian Journal of Agricultural Research - AUST J AGR RES*, v.59. 2008. Available from: <<https://doi.org/10.1071/ar07282>>. Accessed: Aug. 20, 2021. doi: 10.1071/ar07282.
- GITELSON, A., et al. Three-Band Model for Noninvasive Estimation of Chlorophyll Carotenoids and Anthocyanin Contents in Higher Plant Leaves. *Geophysical Research Letters - GEOPHYS RES LETT*, v.33. 2006. Available from: <<https://doi.org/10.1029/2006gl026457>>. Accessed: Aug. 20, 2021. doi: 10.1029/2006gl026457.
- HABOUDANE, D., et al. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, v.81, n.2, p.416-426. 2002. Available from: <[https://doi.org/10.1016/S0034-4257\(02\)00018-4](https://doi.org/10.1016/S0034-4257(02)00018-4)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(02)00018-4.
- HATFIELD, J. L., et al. Application of Spectral Remote Sensing for Agronomic Decisions. *Agronomy Journal*, v.100, n.3, p.117-131. 2015. Available from: <<https://doi.org/10.2134/agronj2006.0370c>>. Accessed: Aug. 20, 2021. doi: 10.2134/agronj2006.0370c.
- HATFIELD, J. L.; J. H. PRUEGER. Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sensing*, v.2, n.2, p.562-578. 2013. Available from: <<https://doi.org/10.3390/rs2020562>>. Accessed: Aug. 20, 2021. doi: 10.3390/rs2020562.
- HE, K. K. W. A. Y. Estimating grassland chlorophyll content using remote sensing data at leaf, canopy, and landscape scales. *Canadian Journal of Remote Sensing: Journal canadien de télédétection*. 2013. Available from: <<https://doi.org/10.5589/m13-021>>. Accessed: Aug. 20, 2021. doi: 10.5589/m13-021.

- HOEPPNER, J. M., et al. Mapping Canopy Chlorophyll Content in a Temperate Forest Using Airborne Hyperspectral Data. **Remote Sensing**, v.12, n.21. 2020. Available from: <<https://doi.org/10.3390/rs12213573>>. Accessed. doi: 10.3390/rs12213573.
- Ji, T., et al. Hyperspectral-Based Estimation on the Chlorophyll Content of Turfgrass. **Spectroscopy and Spectral Analysis**, v.40, n.8, p.2571-2577. 2020. Available from: <[https://doi.org/10.3964/j.issn.1000-0593\(2020\)08-2571-07](https://doi.org/10.3964/j.issn.1000-0593(2020)08-2571-07)>. Accessed. doi: 10.3964/j.issn.1000-0593(2020)08-2571-07.
- L. LIANG, et al. Chlorophyll content inversion with hyperspectral technology for wheat canopy based on support vector regression algorithm. **Transactions of the Chinese Society of Agricultural Engineering**, v.28, n.20, p.162-171. 2012. Available from: <<https://doi.org/10.3969/j.issn.1002-6819.2012.20.022>>. Accessed: Aug. 20, 2021. doi: 10.3969/j.issn.1002-6819.2012.20.022.
- LE BRIS, A., et al. EXTRACTION OF OPTIMAL SPECTRAL BANDS USING HIERARCHICAL BAND MERGING OUT OF HYPERSPECTRAL DATA. **ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v.XL-3/W3, p.459-465. 2015. Available from: <<https://doi.org/10.5194/isprsarchives-XL-3-W3-459-2015>>. Accessed: Aug. 20, 2021. doi: 10.5194/isprsarchives-XL-3-W3-459-2015.
- LE MAIRE, G., et al. 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 Sensing of Environment**, p.3846-3864. 2008. Available from: <<https://doi.org/10.1016/j.rse.2008.06.005>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.rse.2008.06.005.
- LI, K., et al. Hyperspectral Estimation Method of Chlorophyll Content in MOSO Bamboo under Pests Stress. **Spectroscopy and Spectral Analysis**, v.40, n.8, p.2578-2583. 2020. Available from: <[https://doi.org/10.3964/j.issn.1000-0593\(2020\)08-2578-06](https://doi.org/10.3964/j.issn.1000-0593(2020)08-2578-06)>. Accessed. doi: 10.3964/j.issn.1000-0593(2020)08-2578-06.
- LI, L., et al. Evaluating chlorophyll density in winter oilseed rape (*Brassica napus* L.) using canopy hyperspectral red-edge parameters. **Computers and Electronics in Agriculture**, v.126, p.21-31. 2016. Available from: <<https://doi.org/10.1016/j.compag.2016.05.008>>. Accessed. doi: 10.1016/j.compag.2016.05.008.
- LIANG, L., et al. Estimating Canopy Leaf Water Content in Wheat Based on Derivative Spectra. **Scientia Agricultura Sinica**. 2013. Available from: <<https://doi.org/10.3864/j.issn.0578-1752.2013.01.003>>. Accessed: Aug. 20, 2021. doi: 10.3864/j.issn.0578-1752.2013.01.003.
- LIU, P., et al. Estimating leaf chlorophyll contents by combining multiple spectral indices with an artificial neural network. **Earth Science Informatics**, v.11, n.1, p.147-156. 2018. Available from: <<https://doi.org/10.1007/s12145-017-0319-1>>. Accessed. doi: 10.1007/s12145-017-0319-1.
- LIU, W., et al. Relationship between rice LAI, CHD, and hyperspectral data. **Proceedings of SPIE - The International Society for Optical Engineering**, v.4897. 2002. Available from: <<https://doi.org/10.1117/12.466710>>. Accessed: Aug. 20, 2021. doi: 10.1117/12.466710.
- MAIN, R., et al. An investigation into robust spectral indices for leaf chlorophyll estimation. **Isprs Journal of Photogrammetry & Remote Sensing**, v.66, n.6, p.751-761. 2011. Available from: <<https://doi.org/10.1016/j.isprsjprs.2011.08.001>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.isprsjprs.2011.08.001.
- MALENOVSKY, Z., et al. Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. **Remote Sensing of Environment**, v.131, n.8, p.85-102. 2013. Available from: <<https://doi.org/10.1016/j.rse.2012.12.015>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.rse.2012.12.015.
- MEER, W. B., K. SCHOLTE, A. SKIDMORE, H. JONG, J. G. P. W. CLEVERS, ET AL. Spatial scale variations in vegetation indices and above-ground biomass estimates: Implications for MERIS. **International Journal of Remote Sensing**, v.22, n.17, p.3381-3396. 2001. Available from: <<https://doi.org/10.1080/01431160152609227>>. Accessed: Aug. 20, 2021. doi: 10.1080/01431160152609227.
- MILLER, J. R., et al. Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model. **International Journal of Remote Sensing**, v.11, n.10, p.1755-1773. 1990. Available from: <<https://doi.org/10.1080/01431169008955128>>. Accessed: Aug. 20, 2021. doi: 10.1080/01431169008955128.
- MUTANGA, O., et al. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. **Remote Sensing of Environment**, v.89, n.3, p.393-408. 2004. Available from: <<https://doi.org/10.1016/j.rse.2003.11.001>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.rse.2003.11.001.
- PENG, Y., et al. Best hyperspectral indices for assessing leaf chlorophyll content in a degraded temperate vegetation. **Ecology and Evolution**, v.8, n.14, p.7068-7078. 2018. Available from: <<https://doi.org/10.1002/ece3.4229>>. Accessed. doi: 10.1002/ece3.4229.
- PU, R., et al. Mapping and assessing seagrass bed changes in Central Florida's west coast using multitemporal Landsat TM imagery. **Estuarine Coastal & Shelf Science**, v.149, p.68-79. 2014. Available from: <<https://doi.org/10.1016/j.ecss.2014.07.014>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.ecss.2014.07.014.
- RDJ, A.; B. ARH. Interpreting vegetation indices. **Preventive Veterinary Medicine**, v.11, n.3-4, p.185-200. 1991. Available from: <[https://doi.org/10.1016/S0167-5877\(05\)80004-2](https://doi.org/10.1016/S0167-5877(05)80004-2)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0167-5877(05)80004-2.
- SHI, T., et al. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. **Plant & Soil**, v.366, n.1-2, p.363-375. 2013. Available from: <<https://doi.org/10.1007/s11104-012-1436-8>>. Accessed: Aug. 20, 2021. doi: 10.1007/s11104-012-1436-8.
- SIMS, D. A.; J. A. GAMON. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. **Remote Sensing of Environment**, v.81, n.2-3, p.337-354. 2002. Available from: <[https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(02)00010-X.
- THENKABAIL, P. S., et al. Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. **Remote Sensing of Environment**, v.71, n.2, p.158-182. 2000. Available from: <[https://doi.org/10.1016/S0034-4257\(99\)00067-X](https://doi.org/10.1016/S0034-4257(99)00067-X)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(99)00067-X.

VOGELMANN, J. E., et al. Red edge spectral measurements from sugar maple leaves. **International Journal of Remote Sensing**, v.14, n.8, p.1563-1575. 1993. Available from: <<https://doi.org/10.1080/01431169308953986>>. Accessed: Aug. 20, 2021. doi: 10.1080/01431169308953986.

ZARCO-TEJADA, P. J., et al. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. **Remote Sensing of Environment**,

v.84, n.2, p.283-294. 2003. Available from: <[https://doi.org/10.1016/S0034-4257\(02\)00113-X](https://doi.org/10.1016/S0034-4257(02)00113-X)>. Accessed: Aug. 20, 2021. doi: 10.1016/S0034-4257(02)00113-X.

ZHANG, Y., et al. Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. **Remote Sensing of Environment**, v.112, n.7, p.3234-3247. 2008. Available from: <<https://doi.org/10.1016/j.rse.2008.04.005>>. Accessed: Aug. 20, 2021. doi: 10.1016/j.rse.2008.04.005.