Model for estimation of total nitrogen content in sandalwood leaves based on nonlinear mixed effects and dummy variables using multispectral images

Zhulin Chen, Xuefeng Wang [*]

Research Institute of Forestry Resource Information Techniques, Chinese Academy of Forestry, Beijing, China

A B S T R A C T

Fertilizer overuse is a common phenomenon in global agroforestry production, and this overuse causes ecological destruction. The ability to accurately estimate the nutrient content of plant leaves in real-time would be a wonderful solution to reduce the degree of environmental damage. In recent years, remote sensing technology has been widely used in the diagnosis of crop nutrition in many countries. Most studies focus on optimal band selection or create new vegetation indices, but these studies have ignored the random impact of natural environmental factors on the estimated results. This paper proposed an estimation model of total nitrogen content (TNC) in sandalwood leaves that takes sampling season and site conditions as the dummy variable and random effect, respectively. Three forestry farms with different locations and site conditions were selected as study areas to enhance the universality of this model. Multispectral images of leaves were obtained using a low-cost five-band camera (RedEdge3, MicaSense, USA), and the experimental results indicate the following: (1) the growth of the tree height, crown width and stem effectively increased under the medium gradient level (N2), whereas a high gradient level (N3) significantly promoted all aspects except tree height; (2) the mean and variance of some image texture features of the G, RE and NIR band were significantly correlated with TNC at the 0.05 and 0.01 levels, and the texture mean value index (TMVI) proposed in this paper can improve the correlation with TNC; and (3) the results obtained using the nonlinear mixed-effects model with dummy variables improved the fitting degree and estimation accuracy compared with results of SVR and BPNN. This study demonstrates the advantages of using the nonlinear mixed-effects model with dummy variables to obtain a more reliable estimation model for the nutritional diagnosis of rare tree species.

1. Introduction

Sandalwood is a semi-parasitic evergreen tree that is mainly distributed in Southeast Asia, Australia and the Pacific [1]. The wide range of applications of this tree in fragrance production and in pharmaceutical, medicinal and other commercial products makes sandalwood valuable worldwide. However, due to overharvesting and the destruction of its ecological environment, wild sandalwood resources have sharply decreased globally, resulting in a marked decline in exports. Since sandalwood is not native to China, substantial labour and material resources have been invested to study artificial methods of cultivating sandalwood to meet market needs. After decades of research and promotion, people have mastered the basic methods of cultivating sandalwood [2]. Researchers have found that sandalwood cultivation is highly sensitive to water and nutrition resources; these aspects are impacted by poor forest farm management, which can seriously affect sandalwood growth and heartwood formation. The ability to monitor nutritional status in real time is the key to improving the growth rate and quality of sandalwood.

The optimization of nitrogen utilization is a major factor in improving plant yield and quality. Today, many countries apply nitrogen fertilizer to obtain high levels of food production. However, the nitrogen utilization rate decreases when the amount of fertilizer exceeds demand, and the surplus nitrogen will be lost through runoff, leaching, ammonia volatilization and denitrification, which lead to groundwater nitrate pollution, water eutrophication and the greenhouse effect [3,4]. The traditional method of monitoring crop nitrogen content is to collect plant leaves in the field and perform chemical analysis in the laboratory. Although the results are reliable, this method cannot meet the need for rapid and non-destructive monitoring. Therefore, an effective technology to improve the nitrogen utilization efficiency would be to monitor the nitrogen status of plants in real time in an effort to balance the applied nitrogen supply and the plant's nitrogen utilization [5].

Plant nutrition diagnosis technology has rapidly developed over the past 20 years. The main diagnosis methods are the fertilizer window, leaf colour chart, chlorophyll meter, hand-held spectrometer, hyperspectral remote sensing and digital image processing technology. However, there are some disadvantages associated with these different methods. For example, the fertilizer window cannot determine specific amounts of topdressing, and laboratory chemical analysis is also required [6]. Leaf colour charts cannot explain the reason for chlorosis and are also affected by planting density and crop varieties [7]. SPAD is widely used in different types of research, but its ability to measure leaf area is limited [8]. When plant nitrogen content is close to or higher than the optimal value, the SPAD value cannot accurately characterize the chlorophyll content [9]. Hence, hand-held spectrometers, hyperspectral remote sensing and digital image processing technologies are currently more reliable and popular in plant nutrition diagnosis.

In the early stages of non-destructive plant nutrition diagnosis, digital image processing was widely used because of its convenience and low cost. Due to its flexibility and feasibility, it is only necessary to select suitable carriers for research on different scales. Some results have shown that the changes in plant colour were significantly correlated with chlorophyll content, and the chlorophyll content was strongly affected by the total nitrogen content [10]. José et al. [11] used single-colour channels and double-colour channels from the RGB, HSV, and L*a* b* colour spaces to predict SPAD values in maize. This analysis shows that the correlation of SPAD values and colour data with the concentration of nitrogen is possible. Furthermore, the study shows that when single-colour channel correlations are performed, the channels G, b* and V provide the best modelling accuracy. Miguel et al. [12] developed a method in which principal component analysis is applied to digital images to calculate a greenness index using RGB components of the colour image that yield an estimate of the amount of nitrogen in the plant. The results show that the capacity of this index to predict the nitrogen deficiencies that affect the barley yield was equal to or better than that of the SPAD measurements. Confalonieri et al. [13] estimated leaf and plant nitrogen content using an 18% grey-dark-green colour index (DGCI) method. Compared to the DGCI and the corrected DGCI, the new method is considerably more stable with regard to both trueness and precision. Baresel et al. [14] used a consumer digital camera, SPAD and reflectance spectrometry for the non-destructive detection of the leaf chlorophyll content and the nitrogen nutrition status in wheat. With spectral measurements, the biomass and leaf nitrogen content could not be clearly differentiated; chlorophyll measurements do not reflect the biomass, whereas the described procedure of image analysis permits the consideration of biomass and leaf nitrogen content.

Although image processing technology has many merits, the band of digital images remains limited to visible light and large bandwidths, which results in insensitivity to subtle changes of leaf colour. Therefore, hand-held spectrometer technologies and hyperspectral remote sensing, with their wide band range and rich number of bands, are favoured by researchers. The vegetation indices calculated from the red to near-infrared bands, which can be obtained from wide bands or narrow bands, show a high correlation with the plant nitrogen stress status [15]. Studies that have estimated biophysical and chemical characteristics have suggested that the saturation phenomenon easily occurs in the biophysical parameter modelling of the index obtained by the wide band [16,17] and that the narrow-band index is more sensitive to changes in plant nitrogen content [18]. While hyperspectral sensors describe the plant canopy reflectance in greater detail than described by multispectral sensors, they also suffer from issues with data redundancy and spectral autocorrelation, termed "high dimensional disaster" or the "Hughes" phenomenon. Therefore, to optimize the advantages of narrow bands, methods and approaches must be provided that can overcome data redundancy [19]. Thorp et al. [20] selected 25 relevant bands from 701 narrow bands; the estimation accuracy of durum wheat characteristics, which included growth, nitrogen status, and grain yield, was higher when using the less redundant spectral data. Yu et al. [21] proposed a

detection and quantification method for nitrogen concentration in oilseed rape leaf using a deep-learning-based regression model composed of stacked auto-encoders (SAEs) and a fully connected neural network (FNN). The SAE-FNN model achieved a reasonable performance for the nitrogen concentration. Corti et al. [22] assessed the capability of hyperspectral line-scan imaging (400–1000 nm) to estimate crop variables in a greenhouse under combined water and nitrogen stress using multivariate data analysis and two data compression methods: canopy average spectra and hyperspectrogram extraction. Image hyperspectrogram compression, without spatial information loss, produced more encouraging results when canopy structure in crop variables was considered than those results produced by the average canopy spectra method.

In addition to data types, many factors affect the accuracy of the non-destructive diagnosis of nitrogen, such as growth stages and crop varieties. Khanal et al. [23] obtained visual images of a corn canopy from manned aircraft and then established the relationship between vegetation indices and yield. The results showed that image correlation to corn yield increased as the corn growth stage progressed but weakened towards the end of the growing season. Because of the different cultivated varieties, Zhao et al. [24] determined that the most sensitive bands for the nitrogen nutrition index in summer maize was located at 710 nm and 512 nm, whereas Goel et al. [25] reported that the change of nitrogen stress was more obvious at 498 nm and 671 nm. Plant water content is an important factor that affects the estimation of nitrogen content. Therefore, reducing the influence of water stress on colour is key to improving the prediction accuracy. Kusnierek et al. [26] developed a method that can distinguish between nitrogen and water status in spring wheat. The results showed that the first component of processed spectral data was related to the water regime, whereas the second component was related to nitrogen fertilizer.

Most studies focus on improving the correlation between vegetation indices and nitrogen content, but they neglect the importance of the modelling algorithm [27]. At present, the most commonly used modelling methods are multivariate statistics and machine learning algorithms. A comparison of those two methods shows that the nonlinear mapping ability of the machine learning algorithm is stronger than that of multivariate statistics, but the machine learning algorithm is inferior in interpreting the results. However, those two methods considered only the fixed effect of parameters while ignoring the random effects on the estimated results.

Nitrogen content estimation by remote sensing is based on the changes in the reflectance in different bands. However, in addition to the nitrogen content in plants, many other factors can affect the reflectance value, such as sampling season and site condition. In this context, the objects of the current study are as follows: (i) to determine the effects of different nitrogen levels on tree height, crown width and stem of sandalwood; (ii) to discuss the correlation between image features and total nitrogen content in sandalwood leaves, and to select or create best features to represent the nitrogen content; and (iii) to propose an estimation model using nonlinear mixed effects and dummy variables that take other influencing factors into consideration.

2. Materials and methods

Differences exist in spectral reflectance curves between healthy plants and those under nitrogen stress, allowing vegetation indices to effectively represent the total nitrogen content [28–30]. In addition, several texture features were calculated and applied in variable selection. After the variable set was determined, five kinds of different formulas, including linear combination, reciprocal combination, power function, logarithmic function and exponential function, were tested, and the formula with the best fitting degree was chosen as the basic model. Then, the sampling season and site condition were added as the dummy variable and random effect, respectively, to complete the modelling process. For ease of understanding, the overall diagram of the total nitrogen content estimation

in sandalwood leaves is displayed in Fig. 1.

2.1. Study area and sample preparation

The data used in this paper were collected from three different forestry farms that are the main cultivation base of sandalwood in Hainan Province (Fig. 2). As illustrated in Table 1, differences exist in many aspects that usually cause random effects in modelling. Ding'an and Longlou are in the same flat terrain, whereas the soil in Ding'an is much more fertile than the soil in Longlou. Tunchang County seat also has fertile soil, but the altitude is higher because this forestry farm is close to Wuzhi Mountain. The sunshine duration is also different in each of the sample plots, which may affect the rate of photosynthesis.

Three-year-old sandalwood seedlings with the same growth status were selected for use in this experiment. The planting distance was set to avoid competing for sunshine and space between each seedling (Fig. 3). Urea $(CO(NH_2)_2)$ was selected as the nitrogen source in this research. After allowing the seedlings a two-week rejuvenation period, the nitrogen stress experiment, which consisted of four levels - 0 g/seedling (CK), 15 g/seedling (N1), 30 g/seedling (N2), and 80 g/seedling (N3), was conducted, with ten seedlings in each level.

The fertilization and sampling date are shown in Table 2. When sandalwood leaves were picked, the fresh weight was at least 5 g per sample. After being sampled, the collected leaves were placed in a bag for fresh keeping and quickly transported to the laboratory for multispectral image acquisition and chemical composition determination.

2.2. Multispectral image acquisition and correction

A low-cost multispectral camera (RedEdge 3, MicaSense, USA), which can detect the B, G, R, RE and NIR bands, was used to obtain multispectral images of the leaves placed into a camera obscura (Fig. 4). The central wavelength of each band was 475 nm, 560 nm, 668 nm, 717 nm, and 840 nm, and the bandwidth was 20 nm, 20 nm, 10 nm, 40 nm and 10 nm for B, G, R, RE and NIR, respectively. The light source used in the camera obscura was an LED single light lamp consisting of 10 bulbs whose centre wavelengths corresponded to the camera. Before leaf image acquisition, a correction operation was conducted using a blackboard and whiteboard. The spectral reflectance of each band was calculated as follows:

$$I_C = \frac{O - B}{W - B} \times 100\% \tag{1}$$

where I_C is corrected reflectance, and O, B and W are the mean colour values of the objects, blackboard and whiteboard, respectively.

Image acquisition devices and the experimental procedure are shown in Fig. 5. The camera was placed parallel to and 60 cm higher than the white background paper, which was placed under the leaf samples. Images were obtained 3 times, continuously, and the features were calculated using an average value.

2.3. Measurement of total nitrogen concentration

The leaves were placed in an oven to dry at 105 °C for half an hour after the multispectral images were taken and then cooled at 70 °C until a constant weight was recorded. The dry leaves were ground through a 20-mesh screen and analysed for total nitrogen content (TNC) via micro-Dumas combustion by the Soil Testing and Plant Analysis Laboratory at the Chinese Academy of Tropical Agricultural Sciences.

2.4. Independent variable extraction

The foreground of the leaf images were extracted using PhotoShop 5.0. In this study, the texture and vegetation index characteristics of each group of leaves were calculated, and the total nitrogen content in the leaves was used to represent the whole plant. The image features included the following: 32 types of vegetation indices (VIs); texture feature mean values; and texture feature variance, including the homogeneity, contrast, dissimilarity, entropy, second moment and correlations of the 5 bands (Table 3).

2.5. Modelling methods

2.5.1. Machine learning

2.5.1.1. SVR. SVR is the application of a support vector machine in regression estimation, which is used to choose the appropriate kernel function to transform the sample from the low-dimensional space to the high-dimensional space so that the nonlinear problems in the low-dimensional space can be solved linearly in the high-dimensional space, while guaranteeing favourable generalization ability [48,49]. Given a set of datapoints $G = (x_i, d_i)_i^n$ (x_i is the input vector, d_i is the desired value, and n is the total number of data patterns), SVR can be described as a function $f(x)$ using the following equation [50]:
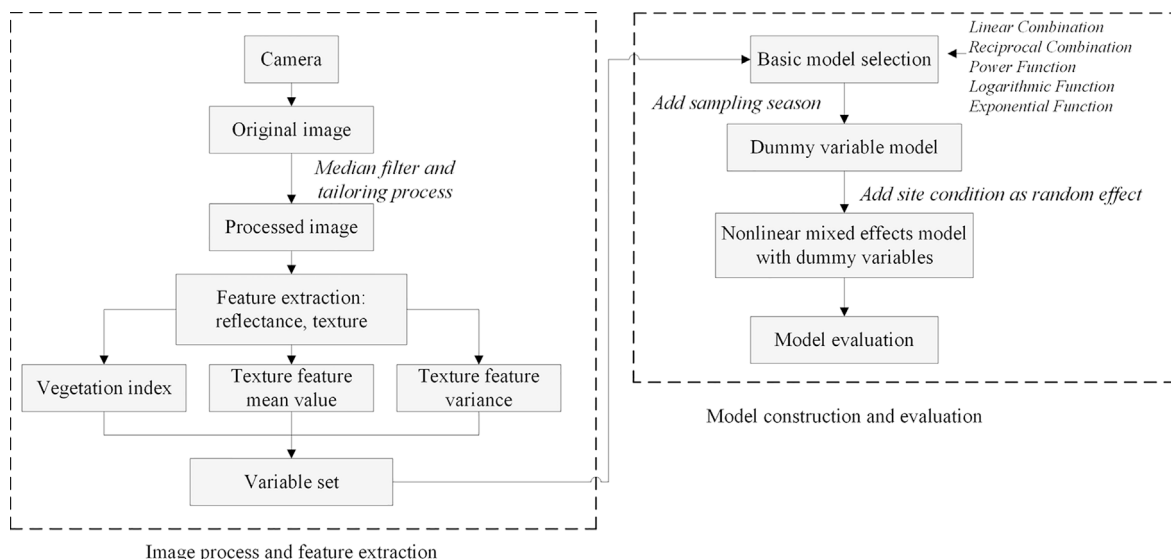

Fig. 1. Overall diagram of the total nitrogen content estimation in sandalwood leaves.
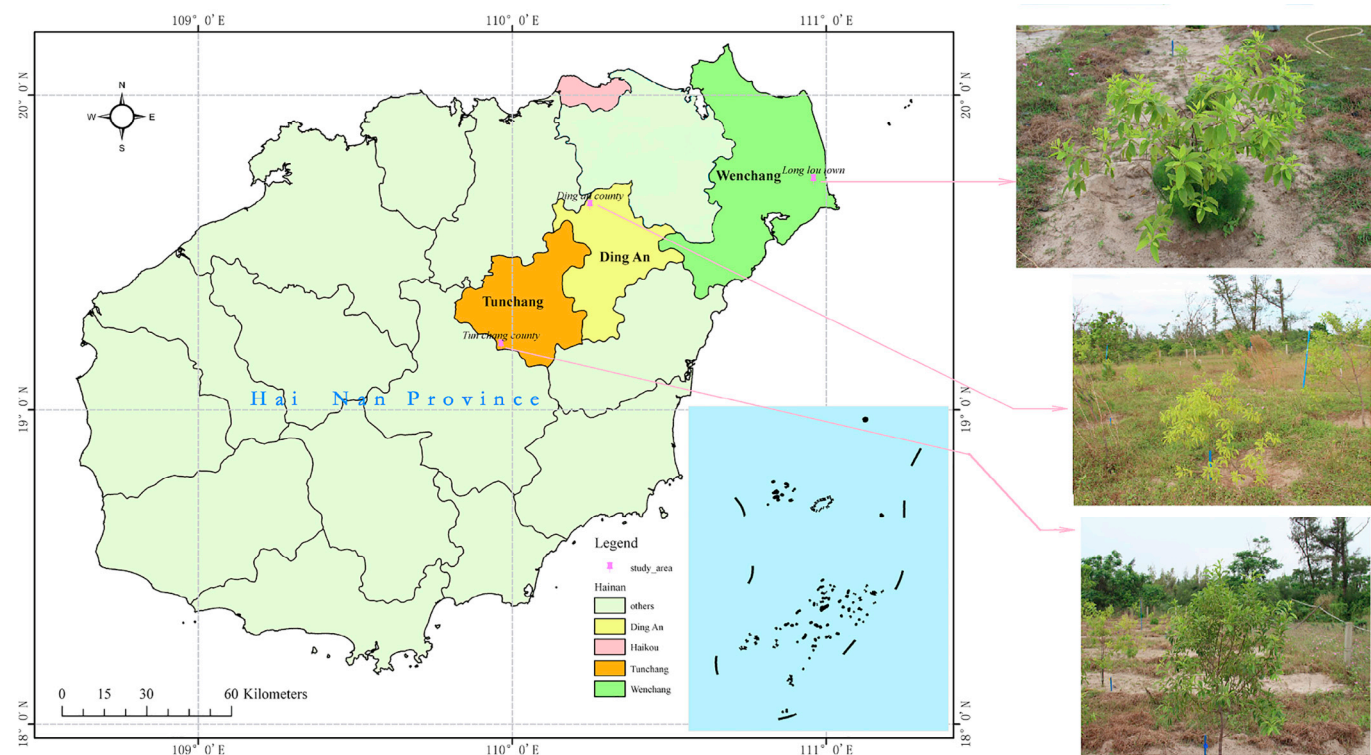
Fig. 2. Distribution of sandalwood samples in northern Hainan Province.

Table 1
Geographical conditions in different sample plots.

Sample plot	Soil type	Altitude	Geographical location	Sunshine duration
Ding'an county seat	Dark brown earth	158 m	19°39′08″Ñ19°41′56″N 110°14′51″E~110°16′05″E	1880 h
Long lou town	White sandy loam	129 m	19°43′58″Ñ19°44′58″N 110°57′34″E~110°57′50″E	1954 h
Tunchang county seat	Red earth	367 m	19°11′28″Ñ19°13′31″N 109°57′52″E~109°59′18″E	2000 h

Table 2
Fertilization and sampling time in this study.

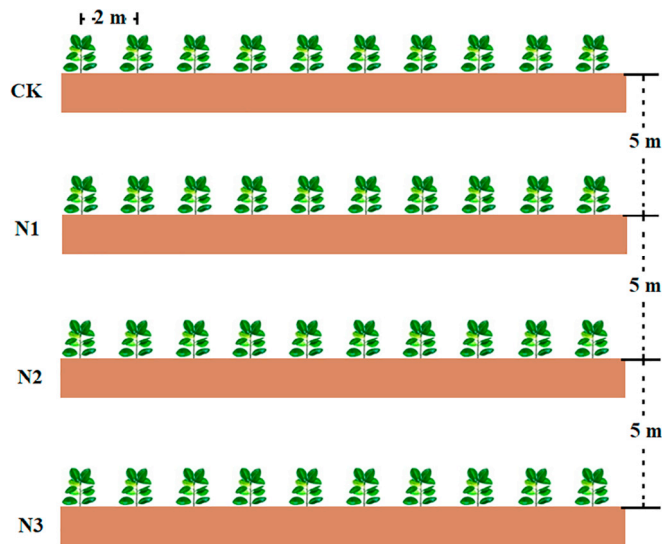| Sample plot	Fertilization date		Sampling date	
	First	Second	First	Second
Ding'an county seat	13/April/2017	13/August/2017	13/June/2017	13/December/2017
Long lou town	16/April/2017	16/August/2017	17/June/2017	16/December/2017
Tunchang county seat	20/April/2017	20/August/2017	22/June/2017	20/December/2017

Fig. 3. Planting method.
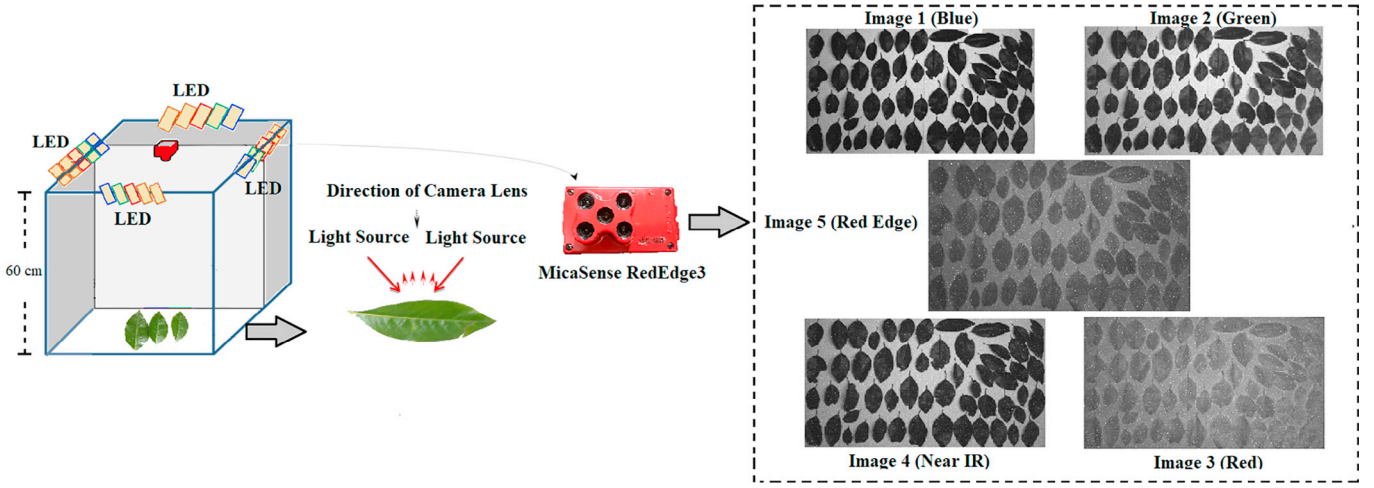

Fig. 4. Camera obscure.

4

Fig. 5. Screen method.

$$f(x) = \omega \times \tau(x) + b \tag{2}$$

where ω is the weight vector, b is the bias, and $\tau(x)$ is the kernel function, which is typically a nonlinear function for transforming nonlinear inputs into a linear mode in a high-dimensional feature space.

When the empirical risk and structure risk are considered together, the SVR model can be constructed to minimize the following quadratic programming problem:

$$\text{Min}: \frac{1}{2}z^T z + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{3}$$

$$\text{subject to} \begin{cases} y_i - z^T x_i - b \leq \varepsilon + \xi_i \\ z^T x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi, \xi_i^* \geq 0 \end{cases} \tag{4}$$

where $i = 1, ..., n$ is the number of training data points; $(\xi + \xi_i^*)$ is the empirical risk; $\frac{1}{2}z^T z$ is the structure risk preventing over-learning and the lack of applied universality; and C is the regularized constant, which determines the trade-off between the empirical risk and regularization terms. This study adopted the general form of the SVR-based regression function, defined as follows [51]:

$$y(x) = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)K(x, x_i) + b \tag{5}$$

where α_i and α_i^* are Lagrange multipliers that satisfy $\alpha_i \times \alpha_i^* = 0$, n is the number of support vectors, and b is the bias. K is the kernel function, which in this study was the radial basis function (RBF) kernel. The RBF is defined as follows:

$$K(x, x_i) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{6}$$

2.5.1.2. BPNN. BPNN is a multilayer feed forward neural network that is characterized by forward signal propagation and back error propagation. According to the error signal of forward propagation, BPNN uses a gradient descent method to learn back propagation with a tutor and minimizes the signal error through repeated forward and backward learning [52]. The basic structure of a BPNN includes three layers that are known as input layers, hidden layers and output layers. The difference between the desired and calculated network output, defined as the goal error of the network, will gradually become smaller until it meets the desired value. The goal error of the network is calculated as follows [53]:

$$E = \sum_{i=1}^{k} E_i = \frac{1}{2k}\sum_{i=1}^{k}(y_i - t_i)^2 \tag{7}$$

where k is the number of training bags, y_i is the actual output value of the ith bag, and t_i is the target value of the ith bag, respectively.

An off-line training algorithm based on a gradient descent approach was used to update network weights that ensure the designed neural convergence to the desired controller output in BPNN. One cycle through all the training pattern is defined as an epoch. Before the optimal network output error is achieved for all training patterns, many epochs are required for the back propagation. The transfer function of the node takes the following form:

$$f(x) = \frac{1}{1 + e^{-x/Q}} \tag{8}$$

where Q is the parameter of the sigmoid function.

2.5.2. Nonlinear mixed-effects model

2.5.2.1. Basic model selection. Before using the nonlinear mixed-effects model, a basic model needs to be determined. In previous studies, the estimation model of total nitrogen content is usually expressed by linear, reciprocal, power function, logarithmic and exponential functions; therefore, in this paper, we used the five models shown in Table 4 and selected the best fitting one as the basic model. In addition, only three independent variables were set in order to prevent model form complexity.

2.5.2.2. Nonlinear mixed model establishment. After the basic model was established, the random effects of location on the total nitrogen content estimation model in sandalwood leaves were further considered. The model was expressed as follows:

$$\begin{cases} Y_{ij} = f\left(\varphi_{ij}, X_{1ij}, X_{2ij}, X_{3ij}\right) + \varepsilon_{ij} \\ \varphi_{ij} = A_{ij}\beta + B_{ij}u_i \\ u_i \sim N(0, \Psi) \\ i = 1, ..., M; j = 1, ..., n_i \end{cases} \tag{9}$$

where Y_{ij}, X_{1ij}, X_{2ij} and X_{3ij} are the TNC and three types of sandalwood leaf image features, labelled as j in sample plot i; φ_{ij} is the parameter vector; ε_{ij} is the intra-group error, which follows a normal distribution; $f(\cdot)$ is a basic model; β is a fixed effect parameter vector; u_i is the random effect generated by sample plot i, with an expected distribution of 0 and variance-covariance matrix of Ψ, which has a normal distribution. A_{ij} and B_{ij} are designed matrixes of β and u_i, respectively; M is the sample

Table 3
Vegetation index equations.

Vegetation Index	Formula	Ref.
Ratio Vegetation Index1,RVI1	NIR/B	[31]
Ratio Vegetation Index2,RVI2	NIR/G	[32]
Ratio Vegetation Index3,RVI3	NIR/R	[31]
Ratio Vegetation Index4,RVI4	NIR/RE	[33]
Difference Vegetation Index1,DVI1	NIR-B	[31]
Difference Vegetation Index2,DVI2	NIR-G	[31]
Difference Vegetation Index3,DVI3	NIR-R	[31]
Difference Vegetation Index4,DVI4	NIR-RE	[33]
Normalized Difference Vegetation Index1,NDVI1	(NIR-B)/(NIR+B)	[32]
Normalized Difference Vegetation Index2,NDVI2	(NIR-G)/(NIR+G)	[34]
Normalized Difference Vegetation Index3,NDVI3	(NIR-R)/(NIR+R)	[32]
Normalized Difference Vegetation Index4,NDVI4	(NIR-RE)/(NIR+RE)	[35]
Renormalized Difference Vegetation Index1,RDVI1	$(NIR-B)/(NIR+B)^{1/2}$	[36]
Renormalized Difference Vegetation Index2,RDVI2	$(NIR-G)/(NIR+G)^{1/2}$	[36]
Renormalized Difference Vegetation Index3,RDVI3	$(NIR-R)/(NIR+R)^{1/2}$	[36]
Renormalized Difference Vegetation Index4,RDVI4	$(NIR-RE)/(NIR+RE)^{1/2}$	[37]
Chlorophyll Index,CI	NIR/G-1	[38]
Wide Dynamic Range Vegetation Index1,WDRVI1	(0.12NIR-B)/(0.12NIR+B)	[39]
Wide Dynamic Range Vegetation Index2,WDRVI2	(0.12NIR-G)/(0.12NIR+G)	[39]
Wide Dynamic Range Vegetation Index3,WDRVI3	(0.12NIR-R)/(0.12NIR+R)	[39]
Wide Dynamic Range Vegetation Index4,WDRVI4	(0.12NIR-RE)/(0.12NIR+RE)	[39]
Modified Simple Ratio,MSR1	$(NIR/B-1)/(NIR/B+1)^{1/2}$	[40, 41]
Modified Simple Ratio,MSR2	$(NIR/G-1)/(NIR/G+1)^{1/2}$	[40, 41]
Modified Simple Ratio,MSR3	$(NIR/R-1)/(NIR/R+1)^{1/2}$	[40]
Modified Simple Ratio,MSR4	$(NIR/RE-1)/(NIR/RE+1)^{1/2}$	[40, 41]
Modified Normalized Difference Vegetation Index1,MNDVI1	(NIR-R+2B)/(NIR+R-2B)	[42]
Modified Normalized Difference Vegetation Index2,MNDVI2	(NIR-R+2G)/(NIR+R-2G)	[42]
Modified Normalized Difference Vegetation Index3,MNDVI3	(NIR-R+2RE)/(NIR+R-2RE)	[42, 43]
Modified Chlorophyll Absorption in Reflectance Index, MCARI	[(NIR-R)-0.2(R-G)] × (NIR/R)	[44]
Triangular Vegetation Index,TVI	0.5[120(NIR-G)-200(R-G)]	[45]
Enhanced Vegetation Index,EVI	2.5(NIR-R)/(1 + NIR+6R-7.5B)	[46]
Triangular Chlorophyll Index,TCI	$1.2(NIR-G)-5(R-G)(NIR/R)^{1/2}$	[47]

Table 4
Basic model of total nitrogen content in sandalwood.

Number	Model	Expression
(1)	Linear Combination	$y = \sum_{i=1}^{3} a_i x_i + a_0$
(2)	Reciprocal Combination	$y = \sum_{i=1}^{3} \frac{a_i}{x_i} + a_0$
(3)	Power Function	$y = b_o \sum_{i=1}^{3} |x_i|^{a_i} + a_0$
(4)	Logarithmic Function	$y = b_0 \ln \left| \sum_{i=1}^{3} a_i x_i \right| + a_0$
(5)	Exponential Function	$y = b_0 e^{\left(\sum_{i=1}^{3} a_i x_i' \right)} + a_0$

Note: $x_i (i = 1 \sim 3)$ are the independent variables and x_i' represents $\frac{x_i - x_{imin}}{x_{imax} - x_{imin}}$.

plot number; n_i is the number of sandalwood trees planted in sample plot i; and ε and u_i are independent of each other.

2.5.2.3. Nonlinear mixed-effects models.
Two types of predictions are usually made when considering nonlinear mixed effects: those that ignore random effects and those that predict random effects. The former does not require prior information about the variables while the latter requires the true value of the total nitrogen content of the leaves. Usually, the estimation of random effects requires the following empirical best linear unbiased prediction theory:

$$u_i = \psi Z_i^T (R_i + Z_i \psi)^{-1} e_i = \psi Z_i^T (R_i + Z_i \psi R_i)^{-1} \left[H_i - f(\beta, u_i^*, D_i) + Z_i u_i^* \right] \tag{10}$$

where $u_i = (u_{i1}, u_{i2}, ..., u_{ik})^T$ is the estimation value of random effects of location type i ($i = 1, 2, ..., M$), and $u_i^* = (u_{i1}^*, u_{i2}^*, ..., u_{ik}^*)^T$ is the empirical best linear unbiased predictive value of $u_i = (u_{i1}, u_{i2}, ..., u_{ik})^T$, which is the random effect. The value of k represents the number of random-effect parameters; β is the estimated value of the fixed-effect parameter β; Ψ is the estimated value of Ψ, which is the variance-covariance matrix of the sample plot random effect u_i; $R_i = diag(R_{i1}, ..., R_{in_i})$ is the estimated value of R, which is the variance-covariance matrix of random error in the group; and Z_i is the estimated value of function $f(\beta, u, D)$, which is related to the random effect partial derivatives.

The specific estimate formula is as follows:

$$Z_i = \frac{\partial f(\beta, u_i, D_i)}{\partial u_i} \Big|_{\beta, u_i^*} \tag{11}$$

The algorithm for estimating the random effect value u_i is based on the walking iterative algorithm proposed by Meng [54].

Step 1: Assuming the value of u_i^* is 0, the initial value of the random effect is calculated as follows:

$$u_i^{(0)} = \Psi Z_i^T (R_i + Z_i \Psi Z_i^T)^{-1} [H_i - f(\beta, 0, D_i)] \tag{12}$$

where

$$Z_i^{(0)} = \frac{\partial f(\beta, u_i, D_i)}{\partial u_i} \Big|_{\beta, u_i^* = 0} \tag{13}$$

Step 2: Then, the value of $u_i^{(k)}$ is calculated according to the following formula:

$$u_i^{(k)} = \Psi \left(Z_i^{(k)} \right)^T \left(R_i + \Psi \left(Z_i^{(k)} \right)^T \right)^{-1} \cdot \left[H_i - f\left(\beta, u_i^{(k-1)}, D_i \right) + Z_i^{(k)} u_i^{(k-1)} \right] \tag{14}$$

where

$$Z_i^{(k)} = \frac{\partial f(\beta, u_i, D_i)}{\partial u_i} \Big|_{\beta, u_i^* = u_i^{(k-1)}}, k = 1, 2, ... \tag{15}$$

Step 3: Next, step 2 is repeated until convergence accuracy is achieved. For example, when $\left| u_i^{(k)} - u_i^{(k-1)} \right| < 1.0 \times 10^{-7}$, the estimated value of the random effects is $u_i = u_i^{(k)}$.

2.6. Model evaluation criteria

The total number of data points obtained in this study is 240; 160 data points were randomly selected for modelling, and the remaining 80 data points were used for testing. To evaluate the fitting degree of the basic models, the coefficient of determination (R^2), the mean error (ME) and the root mean square error of calibration (RMSEC) were selected. The mean absolute error (MAE), mean absolute percentage error (MAPE), and
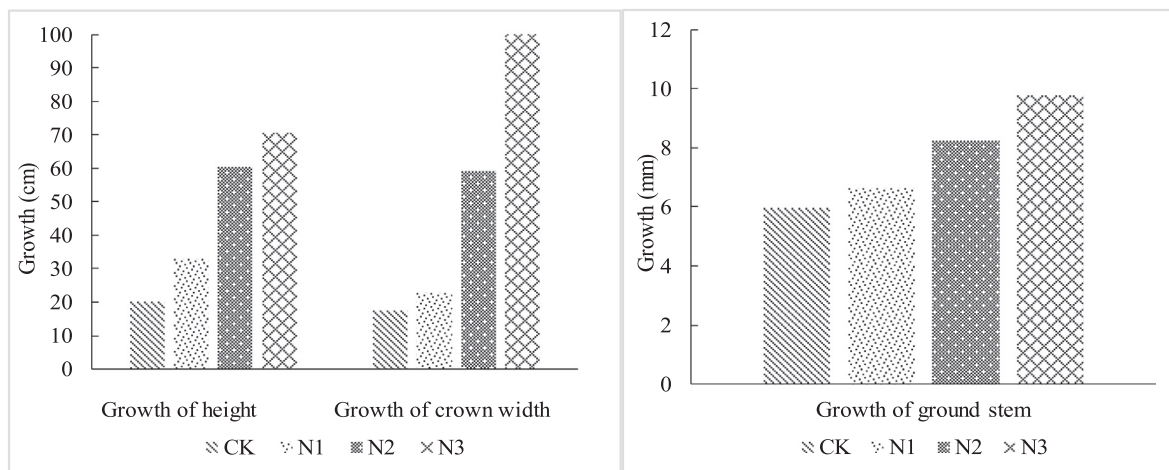
Fig. 6. Growth of height, crown width and ground stem.

root mean square error of prediction (RMSEP) were used to evaluate the estimation accuracy of the models.

Akaike information criterion (AIC) and Bayes information criterion (BIC) were selected to evaluate and compare the fitting accuracy of the nonlinear mixed model together with R^2, and the RMSEC. MAE, MAPE and RMSEP were also used to evaluate the estimation accuracy. All calculations were implemented with the nls function and nlme function of R3.3.2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_i^{'})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (16)$$

$$ME = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^{'}) \qquad (17)$$

$$RMSEC/RMSEP = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^{'})^2 \right]^{1/2} \qquad (18)$$

$$MAE = \sum_{i=1}^{n} |\frac{y_i - y_i^{'}}{n}| \qquad (19)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y_i^{'}}{y_i} \right| \times 100\% \qquad (20)$$

$$AIC = -2 \ln l + 2p \qquad (21)$$

$$BIC = -2 \ln l + \ln n \cdot p \qquad (22)$$

where $y_i, y_i^{'}, \bar{y}_i$ n, p, and l are the observed values, predicted values, mean of the observed values, number of samples, number of parameters in the model, and model maximum likelihood function value, respectively.

3. Results and analysis

3.1. Changes in sandalwood plant growth under different nitrogen levels

The changes in height, crown width and stem growth were calculated for the different levels from the beginning of fertilization to the end of the experiment. In Fig. 6, those three indicators show a rising trend in the different degrees achieved with the increase in nitrogen application. Compared with the control group, the growth of height of N1 to N3 increased by 65%, 301% and 353%. Regarding the crown width, the

values are 29%, 237% and 484%, whereas the growth of the stem only increased by 11%, 38% and 64%, respectively.

3.2. Changes of spectral reflectance under different nitrogen levels

All samples were divided into 4 groups according to the fertilizer gradients. The average TNC of each group was taken to analyse the changes in the image features with the different nitrogen levels. As shown in Fig. 7, the multispectral reflectance of 5 bands is connected to a polygonal map. The spectral reflectance of sandalwood decreases in the blue (475 nm), green (560 nm), red (668 nm) and red edge (717 nm) bands upon increasing TNC. The reason for this phenomenon is that the increased nitrogen levels lead to an increase in chlorophyll content and photosynthesis, and thus, the absorption of visible light increases, which causes a decrease in the reflectance. From the red edge (717 nm) to the near infrared (841 nm), the reflectivity changed from a negative correlation to a positive correlation. This positive correlation is due to the effect of nitrogen levels on the canopy structure and leaf area index. With the increase in total nitrogen content, the leaf area index, water content and biomass also increase, resulting in an increase in near-infrared reflectance [55].
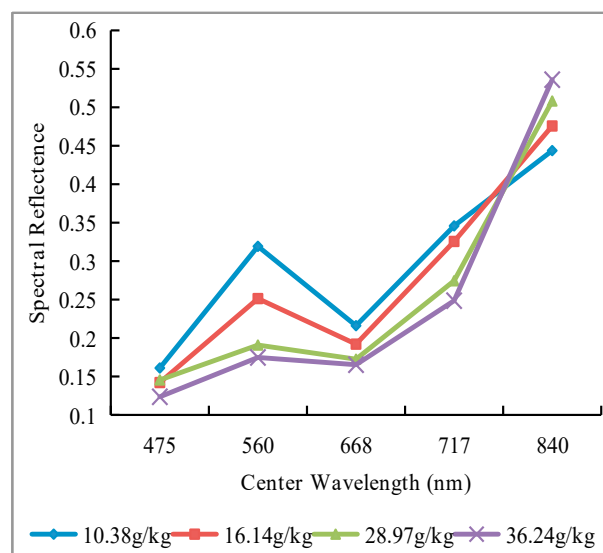

Fig. 7. Spectral reflectance of sandalwood leaves with different TNCs.

3.3. Independent variable selection

The results for the significance test between TNC and the image features are shown in Tables 5–7. In Table 4, the results indicate that significant correlations exist between the TNC and VIs calculated by the G and NIR band and the RE and NIR band, such as RVI2, RVI4, DVI2, and DVI4. This result occurs because the chlorophyll content of leaves are in direct proportion to the total nitrogen content, and the increase in the chlorophyll content always leads to a rise in the photosynthesis intensity and the absorption of visible light, which results in the decrease of spectral reflectance, especially in the G and RE band (Fig. 7). Plants can synthesize proteins effectively with the increase in photosynthesis intensity, which promotes cell division and growth, whereas spectral reflectance in NIR is greatly affected by the leaf structure. Therefore, vegetation indices obtained from the G and NIR bands and from the RE and NIR bands can reflect leaf total nitrogen content well.

Compared to the VIs, the correlations between TNC and texture features are much weaker. In the mean values of the texture features, only five variables are related to TNC at the 0.05 level: contrast; dissimilarity of the G band; and homogeneity, entropy, and second moment of the NIR band. In the texture variance features, the contrast in the G band and entropy of the RE band are related to the TNC at the 0.05 level, whereas the dissimilarity in the G band and correlation in the NIR band are related to the TNC at the 0.01 level. The results illustrate that the G, RE and NIR bands are better than others in reflecting the TNC of sandalwood leaves in terms of VIs and texture features.

3.4. Independent variance selection and transformation

Independent variances must be selected, since the number is limited to three to reduce the complexity of the nonlinear mixed-effects model. Therefore, the selection of the best variables to form each feature is necessary to construct the final model. Although the linear relationship between the VIs and TNC are better than that among the texture features, good correlation exists between those VIs according to the calculation formula. In this paper, RVI2 and CI have the same largest Pearson coefficient. However, according to Table 3, the D-value between those variables is a constant. Therefore, RVI2, which has a simpler formula, has been selected as one of the independent variables.

Five variables related to the TNC at the 0.05 level are selected to form the texture feature mean value. According to Table 6, the homogeneity and the second level of the NIR band show a decreasing trend when the TNC increases while the other three features show the opposite trend. Based on this character, a texture mean value index (with a Pearson coefficient of 0.748) is proposed in this paper, and the formula is calculated as follows:

$$\text{TMVI} = \frac{(Con_{BM} + Diss_{BM} + Entr_{NIR}) - (Homo_{NIR} + SM_{NIR})}{(Con_{BM} + Diss_{BM} + Entr_{NIR}) + (Homo_{NIR} + SM_{NIR})} \quad (23)$$

where Con_{BM} and $Diss_{BM}$ represent the contrast mean value and dissimilarity mean value of the B band, and $Entr_{NIR}$, $Homo_{NIR}$ and SM_{NIR} represent the entropy mean value, homogeneity mean value and second moment mean value, respectively, of the NIR band.

In texture feature variance, the variables related to TNC at the 0.05 and 0.01 levels are selected. However, the texture variance index cannot be formed because all those variables have positive correlation with the TNC. Therefore, the correlation of band NIR with the largest Pearson coefficient in all texture feature variance is chosen as the third independent variable in this study. In conclusion, the RVI2, TMVI and correlation variance of the NIR band (Correlation$_{vNIR}$) corresponds to x_1, x_2 and x_3 in Table 4, respectively.

3.5. Basic model determination

The models shown in Table 3 were tested for fitting degree, and the results are shown in Table 8. All the evaluation indices calculated in model (5) are superior to the other four models; therefore, the nonlinear mixed model that we used in this paper was constructed by the exponential function of variables:

$$y = b_0 e^{\left(a_1 \text{RVI2}' + a_2 \text{TMVI}' + a_3 \text{Correlation}'_v\right)} + a_0 \quad (24)$$

3.6. Dummy variable model

The dummy variable method can divide samples into many kind of categories to build models. Each category has the same model form with different parameter values. In this study, sampling season and site condition are the two main factors that affect the TNC estimation accuracy. Because variables obtained by multispectral images barely distinguish the difference between nutritional deficiency and water imbalance of plants, sampling season is chosen to be the dummy variable to improve the fitting degree and estimation accuracy of the models.

Considering the different sampling seasons, dummy variables were introduced into b_0, a_1, a_2, a_3 and their combinations; the performance of the best five models are shown in Table 9.

As shown in Table 9, the number and position of the dummy variables affect the fitting degree and estimation accuracy of the model. Although R^2 and ME were optimal when all variables were considered, the RMSEC, MAE, MAPE and RMSEP were smaller when the dummy variables were only added into b_0 and a_1. Thus, the model form is shown as follows:

$$y_{ij} = \left(b_0 + b_0' K_{ij}\right) e^{\left\{\left(a_1 + a_1' K_{ij}\right)\text{RVI2}' + a_2 \text{TMVI}' + a_3 \text{Correlation}'_v\right\}} + a_0 + \varepsilon_{ij} \quad i$$
$$= 1, ..., M, \quad j = 1, ..., n_i \quad (25)$$

where K_{ij} is the dummy variable model, which takes a value of 0 when the sampling time is summer and takes a value of 1 when the sampling time is winter.

3.7. Nonlinear mixed-effects model with dummy variables

The site condition of the study areas was introduced into the dummy variable model as random effects to reflect the effects of different natural environments on the nitrogen uptake and leaf colour of the sandalwood. To reduce the complexity of models, only the unstructured matrix was discussed. However, because the number of combinations of random effects is very large, only the 5 best performing models are listed in

Table 5
Significance test between vegetation index and TNC of sandalwood leaves.

Vegetation index	RVI1	RVI2	RVI3	RVI4	DVI1	DVI2	DVI3	DVI4
Pearson coefficient	0.682*	0.845**	0.797**	0.833**	0.682*	0.820**	0.746**	0.821**
Vegetation index	NDVI1	NDVI2	NDVI3	NDVI4	RDVI1	RDVI2	RDVI3	RDVI4
Pearson coefficient	0.656*	0.821**	0.770**	0.832**	0.680*	0.822**	0.762**	0.828**
Vegetation index	WDRVI1	WDRVI2	WDRVI3	WDRVI4	MSR1	MSR2	MSR3	MSR4
Pearson coefficient	0.673*	0.841**	0.789**	0.833**	0.672*	0.837**	0.786**	0.833**
Vegetation index	MNDVI1	MNDVI2	MNDVI3	CI	TCI	TVI	EVI	MCARI
Pearson coefficient	−0.399	−0.003	0.290	0.845**	−0.659*	0.473	0.770**	0.771**

Note:* denotes significance at the 0.05 level; ** denotes significance at the 0.01 level.

Table 6
Significance test between texture feature mean value and TNC of sandalwood leaves.

Band	Homogeneity$_M$	Contrast$_M$	Dissimilarity$_M$	Entropy$_M$	Second Moment$_M$	Correlation$_M$
Band B	−0.228	0.447	0.374	−0.314	0.183	−0.211
Band G	−0.305	0.597*	0.611*	0.252	−0.129	0.149
Band R	−0.208	0.458	0.351	0.251	−0.263	−0.121
Band RE	−0.129	0.178	0.297	−0.328	−0.266	0.244
Band NIR	−0.528*	0.264	0.298	0.571*	−0.561*	0.259

Note:* denotes significance at the 0.05 level; Homogeneity$_M$, Contrast$_M$, Dissimilarity$_M$, Entropy$_M$, Second Moment$_M$, Correlation$_M$ are represent the mean value of homogeneity, contrast, dissimilarity, entropy, second moment, correlation in each band, respectively.

Table 7
Significance test between texture feature variance and TNC of sandalwood leaves.

Band	Homogeneity$_V$	Contrast$_V$	Dissimilarity$_V$	Entropy$_V$	Second Moment$_V$	Correlation$_V$
B	−0.188	0.508	0.432	0.054	0.073	−0.173
G	−0.194	0.619*	0.719**	−0.051	−0.032	0.158
R	−0.144	0.509	0.394	−0.057	−0.057	0.009
RE	0.437	−0.084	0.156	0.601*	−0.061	0.431
NIR	0.236	0.214	0.199	0.064	−0.004	0.764**

Note:* denotes significance at the 0.05 level; ** denotes significance at the 0.01 level; Homogeneity$_V$, Contrast$_V$, Dissimilarity$_V$, Entropy$_V$, Second Moment$_V$, Correlation$_V$ are represent the variance of homogeneity, contrast, dissimilarity, entropy, second moment, correlation in each band, respectively.

Table 8
Parameter values and performance of basic model.

	Parameter value					Calibration set			Validation set		
	a_1	a_2	a_3	a_0	b_0	R^2	ME	RMSEC	MAE	MAPE	RMSEP
(1)	13.261	−174.103	210.626	82.711	–	0.704	−1.342	8.150	5.783	14.522	8.948
(2)	−39.042	141.072	−66.788	60.755	–	0.684	−0.938	9.982	6.393	14.991	10.834
(3)	2.032	−109.423	−2.066	2.960	3.432	0.705	1.329	8.525	5.628	13.764	8.831
(4)	2.253	31.060	−83.826	20.928	2.891	0.378	5.294	15.294	9.824	19.285	17.924
(5)	1.379	−1.152	1.013	9.371	2.759	**0.765**	**−1.153**	**7.506**	**4.328**	**13.597**	**7.996**

Table 9
Performance of the different model alternatives with dummy variables.

Dummy variables	Calibration set			Validation set		
	R^2	ME	RMSEC	MAE	MAPE (%)	RMSEP
b_0, a_1, a_2, a_3	0.868	−0.384	4.049	2.985	11.933	4.482
b_0, a_1	**0.855**	**−0.249**	**3.848**	**2.825**	**10.822**	**4.299**
b_0, a_1, a_2	0.836	0.528	4.844	3.015	11.425	5.247
b_0, a_1, a_3	0.835	0.423	4.633	3.219	11.652	4.914
a_1, a_2, a_3	0.841	−0.215	4.893	3.449	11.034	5.338

Table 10.

The best model was determined after all the evaluation indices had been compared. This consists of the basic model, which takes the form of an exponential function, with the added dummy variables of b_0 and a_1 reflecting the different sampling seasons, and with a_1, a_2, and a_3 added to reflect the random effects caused by site condition. The model expression is as follows:

$$y_{ij} = \left(b_0 + b_0^{'}K_{ij}\right)e^{\left\{\left(a_1 + a_1^{'}K_{ij} + u_{i1}\right)\text{RVI2}^{'} + (a_2 + u_{i2})\text{TMVI}^{'} + (a_3 + u_{i3})\text{Correlation}_{V}^{'}\right\}} + a_0 + \varepsilon_{ij},$$
$$i = 1, ..., M, \quad j = 1, ..., n_i$$

(26)

The error distribution is as follows:

$$u_i = (u_{i1}, u_{i2}, u_{i3})^{\text{T}} \sim N(0, \Psi), \Psi = \begin{pmatrix} 2.4482 & -3.5924 & -0.0337 \\ -3.5924 & 13.2494 & 0.2743 \\ -0.0337 & 0.2743 & 0.0081 \end{pmatrix},$$

$$\varepsilon_{ij} \sim N(0, 1.4286^2)$$

3.8. Model comparison

The SVR and BPNN methods were also used to estimate the TNC, which used input factors for vegetation indices with Pearson coefficients greater than 0.83 and texture features with Pearson coefficients greater than 0.6. Equations numbered 3 to 6 illustrate that the parameters C, ε, and γ must be determined for the SVR training. To obtain the optimal SVR model, the training dataset was randomly divided into 80% and 20% for model building and testing, respectively. The grid search method was also applied to determine the best parameter set of C and γ that could

Table 10
Evaluation indices of nonlinear mixed-effects models with a dummy variable.

Dummy variable	Random effects	Calibration set					Validation set		
		AIC	BIC	R^2	ME	RMSEC	MAE	MAPE(%)	RMSEP
b_0, a_1	b_0, a_1, a_2, a_3	316.295	259.296	0.888	−0.531	2.894	2.852	12.29	2.973
b_0, a_1	b_0, a_1, a_2	320.241	259.252	0.876	−0.235	3.429	3.049	11.93	3.686
b_0, a_1	a_1, a_2, a_3	**310.879**	**252.264**	**0.912**	**−0.312**	**2.789**	**2.425**	**10.05**	**2.974**
b_0, a_1	a_1, a_2, a_3	314.594	258.353	0.905	0.424	2.829	2.644	12.14	3.007

generate the minimum error. The results show that it reached the best fitting degree when *c* and *g* were set to 1.414 and 2.063, respectively. In the optimization of BPNN, the activation function in the hidden layer was set to "tansig", the training function was set to "trainlm", and the transfer function for the output layer was set to "purelin". An empirical formula (equation (27)) was used to determine the number of nodes in the hidden layer [56]. The best accuracy was achieved when the number of nodes was set to 10. Because of its efficient convergence capacity, the Levenberg-Marquardt minimization algorithm was used to calibrate the synaptic coefficients [57].

$$n = \sqrt{n_i + n_0} + a \qquad (27)$$

where *n* is the number of nodes in the hidden layer; n_i is the number of nodes in the input layer; n_0 is the number of nodes in the output layer; and *a* is a constant between 1 and 10.

The estimation results of the three models are shown in Table 12 and Fig. 8. The SVR and BP neural network could explain the relationship between the image features and the TNC of sandalwood leaves while the nonlinear mixed model with dummy variables was superior to those methods, both in terms of goodness of fitting degree and estimation accuracy. From the measured-estimated graph (Fig. 8), the estimated results obtained by SVR are lower than the measured values while the estimated values obtained by the BP neural network show the opposite trend. According to the plots, the values calculated from the proposed method are more uniformly distributed on both sides of the y = x trend line. Table 11.

4. Discussion

4.1. Changes of tree measurement factors

The increase in nitrogen application has a greater impact on tree height and crown width but has only a slight influence on the stem. The changes of those tree-measurement factors were not significant at the N1 level except for the tree height, which indicates that the nutrition was first absorbed by cells in the stem when the nitrogen content was insufficient. When the level of fertilization reached N2, the growth rate of both the height and crown width increased substantially. Under a high nitrogen fertilization level (N3), the crown width still increased sharply while the growth rate of height obviously slowed; the growth of stem also showed an upward trend. Therefore, the results illustrate that insufficient nitrogen fertilizer only improves the growth of the tree height, whereas appropriate fertilization plays a significant role in all three aspects. By contrast, excessive nitrogen fertilizer has a greater effect on the crown and stem than on the height.

4.2. VI and texture feature responses in different stress levels

Previous studies have reported that nitrogen stress experiments have a significant effect on reflectance, which causes great differences in the vegetation index [12,17,18]. Our research confirmed this conclusion by observing that the reflectance in the visible band decreased with the increase in the nitrogen application, whereas the opposite trend was observed in the near infrared band. Generally, the G, RE and NIR bands are considered the most suitable variables for indicating plant health because, when the nitrogen content is sufficient, and the physiological activity of a plant leaf is vigorous, the capacity of chlorophyll to absorb blue light and red light is strong, which leads to an obvious green peak and a steep red edge [58]. The more encouraging result, however, was that some texture features calculated by the G and NIR bands also showed sensitivity to TNC. Although the Pearson coefficients were not as good as with the VIs (Tables 6 and 7), those parameters do have some merits in representing TNC by using a grey-level co-occurrence matrix and not just pigment. A new parameter, the texture mean value index (TMVI), was proposed in this study and improved the Pearson coefficient to 0.748. This parameter considered the positive and negative relationship between those texture features and the TNC, so it obtained better results using the normalized difference method.

4.3. Nonlinear mixed-effects model vs. machine learning

As one of the most popular statistical methods in recent years, the nonlinear mixed-effects model plays an important role in forestry and environmental management. This model considers the random effect caused by different site conditions, which most methods, such as multiple linear regression and machine learning, neglect [26–28]. In this research, the nonlinear mixed-effects model with dummy variables was applied, five model forms were tested, and the results show that the exponential function has the best fit compared with the other forms. This phenomenon illustrated that the pigment and texture feature information better indicate the TNC only at low and middle levels. Furthermore, when excessive nitrogen content is present in the leaves, the changes in colour and texture feature would not better represent the TNC because most of that nutrition would be absorbed and converted into biomass.

The sampling season was considered as the dummy variable because the precipitation differs between summer and winter. Tables 9 and 10 showed that the fitting degree would be better when the dummy variable is added into b_0 and a_1, and when the random effect is added into a_1, a_2, and a_3. This finding explains how the soil moisture content influences the VIs more than it influences the texture features and indicates why the texture responds more slowly when under water stress [59,60]. By contrast, the site condition influences all the variables. Therefore, we can
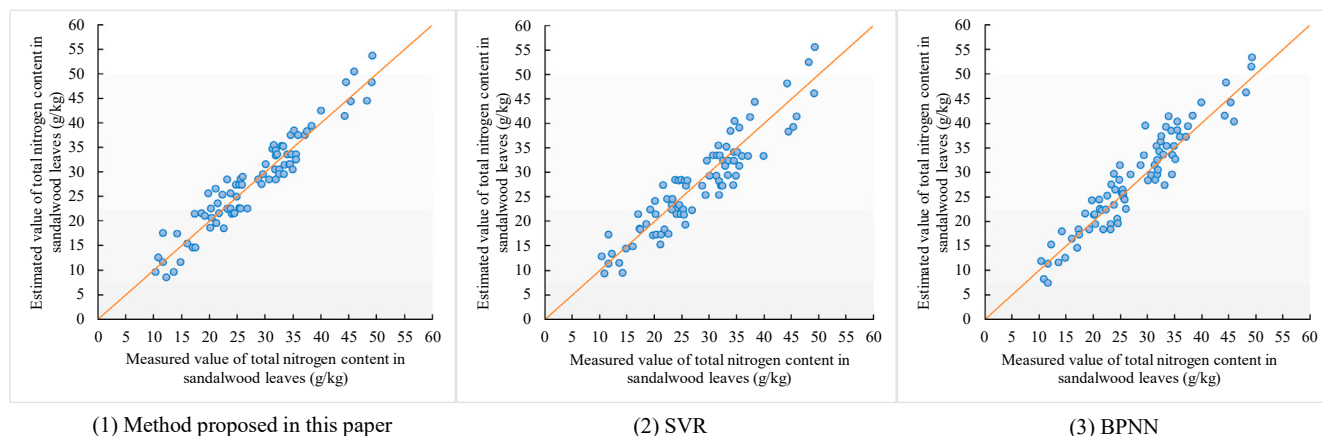

(1) Method proposed in this paper   (2) SVR   (3) BPNN

Fig. 8. Relationship between the measured value and estimated value in the verification set.

Table 11
Parameter values of nonlinear mixed effects model.

Event	Parameter						
	b_0	$b_0^{'}$	a_1	$a_1^{'}$	a_2	a_3	a_0
Value	2.022	1.485	0.949	0.663	−1.266	0.824	7.425

Table 12
Accuracy evaluation in different models.

Methods	Calibration set (n = 160)			Validation set (m = 80)		
	R^2	RMSEC	ME	MA	MAPE(%)	RMSEP
Method proposed in this paper	**0.912**	**2.789**	**−0.312**	**2.425**	**10.05**	**2.974**
SVR	0.844	3.678	0.495	3.186	12.09	3.936
BPNN	0.874	3.308	−0.469	2.689	10.99	3.678

assume that different site conditions will have a greater impact on the TNC.

When this nonlinear mixed model was compared with the SVR and BPNN (Fig. 8), the predicted results validated those conclusions. When the TNC of the sandalwood leaves exceed 40 g/kg, the estimated value would be more unreliable. However, the residual values in this range are much smaller than the results calculated by the SVR and BPNN. Moreover, the estimated values calculated by the BP neural network were lower than the measured values, whereas the results show the opposite trend when using SVR. By using dummy variables and random effect factors to reduce the influence of sampling season and site condition, the results obtained by the proposed method were more uniformly distributed on both sides of the y = x trend line.

5. Conclusion

To monitor the nutritional status of sandalwood in real time, we proposed a method based on the nonlinear mixed-effects model with dummy variables to predict the TNC in sandalwood leaves. The results are as follows:

(1) The VIs calculated by G, RE and NIR bands show great correlation to the TNC. In terms of texture features (including texture mean value and texture variance), nine features calculated by the G, RE and NIR band are related to the TNC at the 0.05 or 0.01 level, which indicate that those bands can represent the TNC in both aspects.

(2) The texture mean value index (TMVI) proposed in this paper using the positive and negative correlations between the different texture features improved the linear correlation with TNC.

(3) By analysing the basic model, the position and structure of the dummy variables, and the random effect factors, the final estimation equation could be obtained; an exponential function is used as the basic equation, the dummy variables reflecting the sampling season were added using the parameters b_0 and a_1, and the random effects of the sample plots with unstructured matrix were added using the parameters a_1, a_2, and a_3.

(4) When the proposed method was compared with SVR and BPNN, the results show that both the calibration and the validation samples are optimal. The R^2, RMSEC and ME of the calibration set using the nonlinear mixed-effects model with dummy variables were 0.912, 2.789 and −0.312, respectively. The MA, MAPE and RMSEP of the validation set were 2.425, 10.05% and 2.974, respectively.

Acknowledgment

References

[1] H.S. Kusuma, M. Mahfud, The extraction of essential oil from sandalwood (Santalum album) by microwave air-hydrodistillation method, J. Mater. Environ. Sci. 7 (5) (2016) 1597–1606.

[2] G.H. Ma, Y.M. He, J.F. Zhang, et al., Study on semi-parasitism of sandalwood seedlings, J. Trop. Subtropical Bot. 13 (3) (2005) 233–238 (In Chinese).

[3] T. Huang, X.T. Ju, H. Yang, Nitrate leaching in a winter wheat-summer maize rotation on a calcareous soil as affected by nitrogen and straw management, Sci. Rep. 7 (2017) 42247.

[4] J. Dai, Z.H. Wang, F.C. Li, et al., Winter wheat grain yield and summer nitrate leaching: long-term effects of nitrogen and phosphorus rates on the Loess Plateau of China, Field Crop. Res. 196 (2016) 180–190.

[5] J.F. Shanahan, N.R. Kitchen, W.R. Raun, et al., Responsive in-season nitrogen management for cereals, Comput. Electron. Agric. 61 (1) (2008) 51–62.

[6] G.P. Asner, R.E. Martin, Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests, Front. Ecol. Environ. 7 (5) (2009) 269–276.

[7] V. Balasubramanian, A.C. Morales, R.T. Cruz, S. Abdulrachman, On-farm adaptation of know ledge-intensive nitrogen management technology for rice systems, Nutrient Cycl. Agroecosyst. 53 (1999) 59–69.

[8] T.M. Blackmer, J.S. Scherpers, Use of a chlorophyll meter to monitor nitrogen status and schedule fertigation for corn [J], J. Prod. Agric. 8 (1) (1995) 56–60.

[9] F.F. Lin, L.F. Qiu, J.S. Deng, Y.Y. Shi, L.S. Chen, K. Wang, Investigation of SPAD meter-based indices for estimating rice nitrogen status, Comput. Electron. Agric. 71 (Supplement 1) (2010) 60–65.

[10] S.M. Westerveld, A.W. Mckeown, M.R. Mcdonald, et al., Chlorophyll and nitrate meters as nitrogen monitoring tools for selected vegetables in southern Ontario, Acta Hortic. (Wagening.) 627 (627) (2003) 259–266.

[11] José F. Reyes, C. Correa, JavierZúñiga, Reliability of different color spaces to estimate nitrogen SPAD values in maize, Comput. Electron. Agric. 143 (2017) 14–22.

[12] P. Miguel, O. Rubén, I. Ignacio, et al., New method to assess barley nitrogen nutrition status based on image colour analysis : comparison with SPAD-502, Comput. Electron. Agric. 65 (2) (2009) 213–218.

[13] R. Confalonieri, L. Paleari, E. Movedi, et al., Improving in vivo plant nitrogen content estimates from digital images: trueness and precisi on of a new approach as compared to other methods and commercial devices, Biosyst. Eng. 135 (2015) 21–30.

[14] J.P. Baresel, P. Rischbeck, Y. Hu, et al., Use of a digital camera as alternative method for non-destructive detection of the leaf chlorophyll content and the nitrogen nutrition status in wheat, Comput. Electron. Agric. 140 (2017) 25–33.

[15] P.M. Hansen, J.K. Schjoerring, 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 (4) (2003) 542–553.

[16] G.P. Asner, R.E. Martin, Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests, Front. Ecol. Environ. 7 (5) (2009) 269–276.

[17] P.S. Thenkabail, R.B. Smith, E.D. Pauw, Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, Remote Sens. Environ. 71 (2) (2000) 158–182.

[18] Eillen M Perry, Joan R Davenport, Spectral and spatial differences in response of vegetation indices to nitrogen treatments on apple, Comput. Electron. Agric. 59 (1–2) (2007) 56–65.

[19] H.P. Liu, H.J. An, B. Wang, et al., Recognition and avoidance of Hughes phenomenon in classification of greening tree species with maximum likelihood, ZRID Zone Res. 33 (2) (2016) 449–454 (In Chinese).

[20] K.R. Thorp, G. Wang, K.F. Bronson, et al., Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield, Comput. Electron. Agric. 136 (2017) 1–12.

[21] X. Yu, H. Lu, Q. Liu, Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf, Chemometr. Intell. Lab. Syst. (2017) 172.

[22] M. Corti, P.M. Gallina, D. Cavalli, et al., Hyperspectral imaging of spinach canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content, Biosyst. Eng. 158 (2017) 38–50.

[23] S. Khanal, J. Fulton, N. Douridas, et al., Integrating aerial images for in-season nitrogen management in a corn field, Comput. Electron. Agric. 148 (2018) 121–131.

[24] B. Zhao, A. Duan, S.T. Ata-Ul-Karim, et al., Exploring new spectral bands and vegetation indices for estimating nitrogen nutrition index of summer maize, Eur. J. Agron. 93 (2018) 113–125.

[25] P.K. Goel, S.O. Prasher, J.A. Landry, et al., Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn, Comput. Electron. Agric. 38 (2) (2003) 99–124.

[26] K. Kusnierek, A. Korsaeth, Simultaneous identification of spring wheat nitrogen and water status using visible and near infrared spectra and Powered Partial Least Squares Regression, Comput. Electron. Agric. 117 (C) (2015) 200–213.

[27] L. Wang, X. Zhou, X. Zhu, et al., Estimation of leaf nitrogen concentration in wheat using the MK-SVR algorithm and satellite remote sensing data, Comput. Electron. Agric. 140 (2017) 327–337.

[28] R. He, H. Li, X. Qiao, et al., Using wavelet analysis of hyperspectral remote-sensing data to estimate canopy chlorophyll content of winter wheat under stripe rust stress, Int. J. Remote Sens. 39 (12) (2018) 4059–4076.

[29] J. Clevers, L. Kooistra, M.V.D. Brande, Using sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop, Remote Sens. 9 (5) (2017) 405.

[30] S.K. Singh, J.H. Houx, M.J.W. Maw, et al., Assessment of growth, leaf N concentration and chlorophyll content of sweet sorghum using canopy reflectance, Field Crop. Res. 209 (2017) 47–57.

[31] C.J. Tucker, Red and photographic infrared linear combinations for monitoring vegetation, Remote Sens. Environ. 8 (1979) 127–150.

[32] C. Buschmann, E. Nagel, In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation, Int. J. Remote Sens. 14 (1993) 711–722.

[33] J. Heiskanen, M. Rautiainen, P. Stenberg, et al., Sensitivity of narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species composition, ISPRS J. Photogrammetry Remote Sens. 78 (2013) 1–14.

[34] A.A. Gitelson, Y.J. Kaufman, M.N. Merzlyak, Use of green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sens. Environ. 58 (1996) 289–298.

[35] G.J. Fitzgerald, D. Rodriguez, L.K. Christensen, et al., Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments, Precis. Agric. 7 (2006) 233–248.

[36] J.L. Roujean, F.M. Breon, Estimating PAR absorbed by vegetation from bidirectional reflectance measurements, Remote Sens. Environ. 51 (1995) 375–384.

[37] Liwei Xing, Xiaojuan Li, Angsheng Li, et al., A comparative study on estimation model for leaf area index of vegetation in marshes in honghe national nature reserve based on hyperspectral and multispectral vegetation indices, Wetland Sci. 11 (3) (2013) 313–319 ([in Chinese]).

[38] A.A. Gitelson, Y. Gritz, M.N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, J. Plant Physiol. 160 (2003) 271–282.

[39] A.A. Gitelson, Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation, J. Plant Physiol. 161 (2004) 165–173.

[40] J.M. Chen, Evaluation of vegetation indices and a modified simple ratio for boreal applications, Can. J. Remote Sens. 22 (1996) 229–242.

[41] D. Haboudane, J.R. Miller, E. Pattey, et al., Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture, Remote Sens. Environ. 90 (3) (2004) 337–352.

[42] W. Wang, X. Yao, X.F. Yao, et al., Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat, Field Crop. Res. 129 (2012) 90–98.

[43] C. Jurgens, The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data, Int. J. Remote Sens. 18 (17) (1997) 3583–3594.

[44] C.S.T. Daughtry, C.L. Walthall, M.S. Kim, et al., Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance, Remote Sens. Environ. 74 (2000) 229–239.

[45] N.H. Broge, 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 Sens. Environ. 76 (2000) 156–172.

[46] A. Huete, K. Didan, T. Miura, et al., Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens. Environ. 83 (2002) 195–213.

[47] D. Haboudane, N. Tremblay, J.R. Miller, et al., Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data, IEEE Trans. Geosci. Remote Sens. 46 (2008) 423–437.

[48] W.Y. Duan, L.M. Huang, Y. Han, et al., A hybrid AR-EMD-SVR model for the short-term prediction of nonlinear and non-stationary ship motion, J. Zhejiang Univ. - Sci. 16 (7) (2015) 562–576.

[49] T. Fei, W. Zhang, J. Liang, et al., Springback prediction and optimization of variable stretch force trajectory in three-dimensional stretch bending process, Chin. J. Mech. Eng. 28 (6) (2015) 1132–1140.

[50] M. Dalponte, L. Bruzzone, D. Gianelle, A system for the estimation of single tree stem diameters and volume using multireturn LIDAR, IEEE Trans. Geosci. Remote Sens. 49 (7) (2011) 2479–2490.

[51] S. Shataee, S. Kalbi, A. Fallah, et al., Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms, Int. J. Remote Sens. 33 (19) (2012) 6254–6280.

[52] E.S.E. Din, Z. Yun, A. Suliman, Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework, Int. J. Remote Sens. 38 (4) (2017) 1023–1042.

[53] J. Chen, W. Quan, T. Cui, et al., Estimation of total suspended matter concentration from MODIS data using a neural network model in the China eastern coastal zone, Estuar. Coast Shelf Sci. 155 (2015) 104–113.

[54] S. Meng, S. Huang, Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function[J], For. Sci. 55 (3) (2009) 238–248.

[55] Z.F. Qin, Q.R. Chang, B.N. Xie, et al., Rice leaf nitrogen content estimation based on hyperspectral imagery of UAV in Yellow river diversion irrigation district, Trans. Chin. Soc. Agric. Eng. 32 (23) (2016) 77–85 ([in Chinses]).

[56] T. Shao, K. Song, J. Du, et al., Retrieval of CDOM and DOC using in situ hyperspectral data: a case study for potable waters in northeast China, J. Indian Soc. Remote Sensing 44 (1) (2015) 1–13.

[57] L.S.H. Ngia, J. Sjoberg, Efficient training of neural nets for nonlinear adaptive filtering using a recursive Levenberg-Marquardt algorithm, IEEE Trans. Signal Process. 48 (7) (2000) 1915–1927.

[58] Y.Y. Lu, Y.L. Bai, D.L. Ma, et al., Nitrogen vertical distribution and status estimation using spectral data in maize, Commun. Soil Sci. Plant Anal. 49 (9) (2018) 1–11.

[59] L. Liang, L.P. Di, T. Huang, et al., Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm, Remote Sens. 10 (12) (2018) 1940.

[60] M. Alessandro, Baraldi Rita, Berton Andrea, et al., Estimation of water stress in grapevines using proximal and remote sensing methods, Remote Sens. 10 (1) (2018) 114.