Estimation of Nitrogen Content in Citrus Leaves Using Stacking Ensemble Learning

Tong Wu¹, Yong Li², Ying Ge², Shunzhong Xi², Mengjie Ren², Xiaohui Yuan¹ and Cuizhen Zhuang³

¹ College of Hydrology and Water Resources, Hohai University, Nanjing, Jiangsu, China
² School of Earth Science and Engineering, Hohai University, Nanjing, Jiangsu, China
³ Xinping Chu’s Agriculture Co., Ltd, Yuxi, Yunnan, China
Email: liyong@hhu.edu.cn

Abstract. Nitrogen is an essential biochemical component of citrus growth, and also an important indication to determine the quality and yield of citrus. The traditional method to obtain Leaf Nitrogen Content (LNC) is costly and time-consuming. Therefore, the rapid and accurate acquisition of nitrogen content information by satellite remote sensing is of great significance for citrus cultivation and production. Empirical models based on vegetation indices (VIs) have been widely used to estimate LNC, while individual model only extracts limited information. The ensemble learning strategy has shown great potential in machine learning, so Landsat8 OLI satellite remote sensing images and ground sample data are used to construct a two-layer Stacking ensemble learning framework for estimating the nitrogen content of citrus leaves in this study. In the proposed model, K-Nearest Neighbors (KNN), Random Forest (RF) and Support Vector Regression (SVR) are utilized as base-models. Linear Regression (LR) is employed as the meta-model. Results show that Stacking model can obtain better estimation results of citrus LNC, thus providing scientific decision support for orchard planting and production.

Keywords. Satellite; remote sensing; models; citrus; nitrogen; stacking; vegetation index.

1. Introduction
Nitrogen is an essential nutrient element in the growth process of fruit trees. It plays an important role in the physiological and biochemical processes of fruit trees [1], and also affects the quality and yield of fruit. In the actual agricultural production, the application of nitrogen fertilizer often depends on experience, and excessive nitrogen fertilizer will lead to cost increases and environmental pollution. Rapid and accurate determination of Leaf Nitrogen Content (LNC) in fruit trees is of great significance for scientific and rational application of fertilizer. Chemical detection requires field sampling and a long time of laboratory analysis. Although the results are accurate, this kind of method is expensive and causes damage to trees. Satellite remote sensing has large area, multi-temporal and periodic observation, so it is widely used in the monitoring and research of agricultural [2, 3]. Many vegetation indices (VIs) have been proposed and VIs-based models have been demonstrated to be effective to estimate LNC. However, most of studies only use individual model. In this study, stacking model was constructed to integrate the advantages of different models and make full use of spectral information to estimate LNC, the orchard nitrogen content spatial distribution map was also generated. Therefore, it provides decision support for planting and scientific management of orchard.
2. Materials

2.1. Ground Field Data
The study area is located in Yuxi City, Yunnan Province, China (23° 19'~24° 53' N, 101° 16 '~103° 09'E). Field campaign for sampling was carried out in the middle of July from 2017-2020. And the RTK-GPS measuring system was used to record the coordinates of every sample point. In order to obtain the representative nitrogen content of citrus leaves, 20 leaves were collected around the four directions and central positions of each fruit tree. The collected leaves were packed into dry kraft paper bags and send to the laboratory for total nitrogen determination. The average nitrogen content of these leaves was taken as the LNC for the corresponding sample point.

2.2. Remote Sensing Image Data
Since the leaves were already mature and the biochemical components were relatively stable at the time of sampling, the satellite images corresponding to the sampling time could be used to estimate the LNC. In this study, a total of 4 Landsat8 OLI images from 2017 to 2020 were downloaded. All the data have been geometric corrected. In order to overcome the influence of atmosphere on the spectral reflectivity, the radiation calibration and atmospheric correction were also carried out on the remote sensing images.

3. Methods and Results

3.1. Spectral Features Analysis
According to planting experience, 30~32 g/kg was the ideal LNC in the planting process. The field LNC values were sorted from high to low, and divided into 6 groups at equal intervals. Average spectral reflectance of each group was calculated, and the relationship between different citrus LNC and spectral reflectance was shown in figure 1.

Due to the change of nutrient elements, the internal physiological metabolic process and biochemical elements of plant leaves will change, which lead to differences in spectral reflectance [4]. Figure 1 showed that the spectral reflectance of citrus canopy leaves was similar to the typical plants in the visible range (460-750nm). Since nitrogen promoted chlorophyll synthesis [5], the spectral reflectance in this band range decreased with the increase of LNC. The reflectance in near infrared band (750-1000nm) was significantly higher than that of other bands due to multiple reflections caused by the structure of fruit canopy. At 1750nm and 2250nm, two short-wave infrared bands, the higher LNC value, the lower reflectivity.

![Figure 1. Spectral response curves of different LNC.](image)

Through the correlation analysis of the band reflectance and citrus LNC, it was found that there was a certain correlation between LNC and the reflectance of the original spectral band. At 440nm, 480nm, 560nm and 650nm, LNC was significantly correlated with spectral reflectance (P <0.01), and the correlation coefficients were 0.622, 0.628, 0.594 and 0.579, respectively. It was consistent with the research results of Liu et al. [4]. At the same time, combined with the conclusion of Min et al. [6], the
sensitive bands of nitrogen content in citrus leaves (around 448 nm and 669 nm) were corresponding to satellite original bands. The nitrogen-related VIs were be constructed to estimate LNC on the basis of correlation analysis, and the results are shown in table 1. All the VIs calculation formulas could be inquired from the Index Database (https://www.indexdatabase.de/info/idb.php).

Table 1. Correlation analysis between VIs and LNC.

| Index | Spectral variables | Formulaa | Correlation coefficientb |
|-------|-------------------|----------|--------------------------|
| 1     | NDVI              | $(R_{nir} - R_r)/(R_{nir} + R_r)$ | 0.587* |
| 2     | RVI               | $R_{nir}/R_r$ | 0.587* |
| 3     | GNDVI             | $(R_{nir} - R_g)/(R_{nir} + R_g)$ | 0.616* |
| 4     | DVI               | $R_{nir}/R_r$ | 0.437* |
| 5     | SAVI              | $1.5(R_{nir} - R_b)/(R_{nir} + R_r + 0.5)$ | 0.554* |
| 6     | SIPI              | $(R_{nir} - R_b)/(R_{nir} + R_b)$ | 0.640* |
| 7     | GNDVI             | $1.16(R_{nir} - R_r)/(R_{nir} + R_r + 0.16)$ | 0.582* |
| 8     | GBNDVI            | $(R_g - R_b)/(R_g + R_b)$ | 0.678* |
| 9     | TCARI             | $3[(R_{nir} - R_r) - 0.2(R_r - R_g)]/(R_{nir} - R_r)$ | 0.621* |
| 10    | NDWI              | $(R_g - R_{nir})/(R_g + R_{nir})$ | 0.616* |
| 11    | TVI               | $(NDVI + 0.5)^{1/2}$ | 0.587* |
| 12    | GRNDVI            | $[(R_{nir} - R_r + R_r)/((R_{nir} + R_r + 0.1)$ | 0.601* |
| 13    | RDVI              | $(R_{nir} - R_r)/(R_{nir} + R_r)^{1/2}$ | 0.557* |

Note: a $R_g$, $R_b$, $R_r$ and $R_{nir}$ represent band reflectance in blue, green, red and near-infrared bands, respectively; b "*" indicates the $P$-value < 0.01.

3.2. Stacking Ensemble Model

Stacking is an ensemble learning strategy which combines various regression models as a whole model. Stacking can effectively improve the robustness and generalization ability of the model. The performance of an individual model affects the final regression effect of Stacking model, so the selection of base-model should fully consider the adequacy and diversity [7]. Machine learning models such as K-Nearest Neighbors (KNN), Random Forest (RF) sand Support Vector Regression (SVR) are widely used as the base-model of Stacking in many studies [8, 9]. Therefore, in this study, the above three models combined with Linear Regression (LR) to construct a Stacking model for the estimation of citrus LNC. The Stacking model architecture is shown in figure 2.
3.3. LNC Estimation Results

All the models in the experiment were given the same data set, and the models were trained by grid search and 5-fold cross validation. The performances of the models are listed in table 2. The scatter plot of the estimated LNC and the measured LNC are shown in figure 3.

| Model   | Training set |          | Testing set |          |
|---------|--------------|----------|-------------|----------|
|         | $R^2$        | MAE      | RMSE        | MAPE     | $R^2$   | MAE      | RMSE        | MAPE     |
| KNN     | 0.802        | 1.011    | 1.276       | 3.21     | 0.711   | 1.125    | 1.502       | 3.761    |
| RF      | 0.958        | 0.453    | 0.586       | 1.43     | 0.722   | 1.178    | 1.474       | 3.919    |
| SVR     | 0.579        | 1.428    | 1.861       | 4.51     | 0.610   | 1.448    | 1.745       | 4.792    |
| Stacking | 0.927       | 0.603    | 0.773       | 1.90     | 0.747   | 1.054    | 1.405       | 3.518    |

The stacking model obtained the best estimation results on the test set, with $R^2$ of 0.747, RMSE of 1.405, MAE of 1.054 and MAPE of 3.518%. Compared with the RF, the best individual models, $R^2$ increased by 0.025, RMSE, MAE and MAPE decreased by 0.069, 0.124 and 0.401%, respectively. It indicated that the estimation accuracy of Stacking model for citrus LNC was higher and the estimation error was smaller. The results also revealed that the base-model extracted the spectral features, and the meta-model integrates the key features, which could effectively avoid the inaccurate estimation results of the individual model. The spatial distribution of nitrogen content generated by stacking model is shown in figure 4.

![Figure 3](image3.png)  
**Figure 3.** The scatter plot of estimated LNC and measured LNC of stacking model.

![Figure 4](image4.png)  
**Figure 4.** The spatial distribution of LNC generated by Stacking model. The LNC mostly distributes between 29~31 g/kg, indicating that the most area of nitrogen content level is in an appropriate state.

4. Discussion

From the actual situation of the study area, the planting mode leads to the soil background being mixed in pixels, and this will affect the spectral indices. Based on the concept of soil line [10], the experiment adopted EVI, OSAVI, SAVI to suppress the influence of soil background as much as
possible. In addition, near infrared band was used as a characteristic band to distinguish different nitrogen content levels of plants [4], it participated most calculation of VIs. Meanwhile, a lot of studies have shown that the estimation of crop biochemical components by combining spectral features and texture features can achieve higher accuracy [11, 12]. This study only used spectral features to construct the model for estimating citrus LNC, and the extraction of features may not be sufficient. Therefore, future work will focus on constructing the estimation model by integrating various features.

5. Conclusion
In this study, a Stacking model was built to achieve accurate estimation of citrus LNC and obtain nitrogen content distribution map. The results showed as follows: (1) Because of the influence of nitrogen content changes, the spectral responses of citrus canopy leaves were significantly different, and the adopted VIs were well correlated with LNC. (2) The Stacking model integrated the feature extraction ability of KNN, RF and SVR, and achieved the best performance, which provided a potential way for the use of satellite remote sensing to achieve the accurate estimation of crop parameter.

Acknowledgments
This work has been supported by the National Key Research and Development Program (Grant No. 2016YFA0601504) approved by Ministry of Science and Technology of China, the National Natural Science Foundation of China (Grant No. 41977394), and the Major Project of Science and Technology of Yunnan Province (Grant No. 202002AE090010).

References
[1] Huang C, Lu X, Li J, Xiao Y, Sun M and Xie S 2013 Research progress in nitrogen nutrition physiology of citrus HN Agr Sci (15) 76-9.
[2] Chemura A, Mutanga O, Odindi J, Kutywayo D 2018 Mapping spatial variability of foliar nitrogen in coffee (Coffea arabica L.) plantations with multispectral Sentinel-2 MSI data ISPRS J Photogramm 138 1-11.
[3] Jia Y, Li B, Cheng Y, Liu T, Guo Y, Wu X and Wang L 2015 Comparison between GF-1 images and Landsat-8 images in monitoring maize LAI Transactions of the CSAE 31 (9) 173-9.
[4] Liu X, Lv Q, He S, Yi S, Xie R, Zheng Y, Hu D, Wang Z and Deng L 2015 Estimation of canopy nitrogen and photosynthetic pigment content of citrus by near-ground remote sensing J Remote Sens 19 (6) 1007-18.
[5] Jiang H, Yang H, Chen X, Wang S, Li X, Liu K and Cen Y 2015 Research on accuracy and stability of inversing vegetation chlorophyll content by spectral index method Spectrosc. Spect Anal 35 (4) 975-81.
[6] Min M and Lee W 2005 Determination of significant wavelengths and prediction of nitrogen content for citrus Transactions of the ASAE 48 (2) 455-61.
[7] Feng L, Zhang Z, Ma Y, Du Q, Williams P, Drewry J and Luck B 2020 Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning Remote Sens 12 (12) 2028.
[8] Ding W, Li X, Yang H, An J and Zhang Z 2020 A novel method for damage prediction of stuffed protective structure under hypervelocity impact by stacking multiple models IEEE Access 8 130136-58.
[9] Wu T, Zhang W, Jiao X, Guo W and Alhaj Hamoud Y 2021 Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration Comput. Electron. Agr. 184 106039.
[10] Prudnikova E, Savin I, Vindeker G, Grubina P, Shishkonakova E and Sharychev D 2019 Influence of soil background on spectral reflectance of winter wheat crop canopy Remote Sens 11 (16) 1932.
[11] Meng D, Zhao J, Lan Y, Yan C, Yang D and Wen Y 2020 SPAD inversion model of corn canopy based on UAV visible light image Transactions of the CSAM 51 (S2) 366-74.

[12] Chen P, Feng H, Li C, Yang G, Yang J, Yang W and Liu S 2019 Estimation of chlorophyll content in potato using fusion of texture and spectral features derived from UAV multispectral image Transactions of the CSAE 35 (11) 63-74.