Integration of Crop Growth Model and Random Forest for Winter Wheat Yield Estimation From UAV Hyperspectral Imagery

Siqi Yang, Ling Hu, Haobo Wu, Huazhong Ren, Hongbo Qiao, Peijun Li, and Wenjie Fan

***Abstract*—Accurate and timely crop yield estimation is critical for food security and sustainable development. The rapid development of unmanned aerial vehicles (UAVs) offers a new approach to acquire high spatio-temporal resolution imagery of farmland at a low cost. In order to realize the full potential of UAV platform and sensor, machine learning has been introduced to estimate crop yield, but the shortages of field measurements have troubled researchers. In this article, the CW-RF model, a new wheat yield estimation model suitable for the North China plain, was established using random forest, and the crop growth model (the CERES-wheat model) was chosen to simulate abundant training samples for random forest at field plot scale. According to CERES-wheat model simulation, the leaf area index (LAI) and leaf nitrogen content (LNC) at the wheat jointing and heading stages were selected as the most sensitive parameters, and were retrieved from UAV hyperspectral imagery using the directional second derivative and angular insensitivity vegetation index methods, respectively. Then the retrieved LAI and LNC results were input into the CW-RF model to estimate winter wheat yield. The field validation in Luohe, Henan showed that the root-mean-squared error of the retrieved LAI and LNC were 6.27% and 12.17% at jointing stages, 9.21% and 13.64% at heading stages, respectively. The RMSE of estimated yield was 1,008.08 kg/ha, and the mean absolute percent error of estimated yield was 9.36%, demonstrating the available of the CW-RF model in wheat yield estimation at field plot scale. Apart from Luohe, validations in some other fields (e.g., Xiaotangshan, Beijing), prove the wide applicability of the CW-RF model. In addition, the UAV hyperspectral data were found to significantly improve the retrieval accuracy, and further improve CW-RF model estimation accuracy. In conclusion, this article showed that the CERES-Wheat model simulation can be important data source for machine learning-based wheat yield estimation model at field plot scale, and the hyperspectral sensor mounted on a UAV is a feasible remote sensing data acquisition mode for winter wheat growth monitoring and yield estimation.**

Manuscript received January 20, 2021; revised April 19, 2021 and June 4, 2021; accepted June 4, 2021. Date of publication June 15, 2021; date of current version June 30, 2021. This work was supported by the National Key Research and Development Program of China under Grants 2016YFD0300601 and 2017YFE0122400. *(Corresponding author: Wenjie Fan.)*

Siqi Yang, Ling Hu, Haobo Wu, Peijun Li, and Wenjie Fan are with the School of Earth and Space Sciences, Institute of Remote Sensing and Geographic Information System,Peking University, Beijing 100871, China, and also with the Beijing Key Laboratory of Spatial Information Integration and Its Application, Peking University, Beijing 100871, China (e-mail: sqyang@pku.edu.cn; hul@pku.edu.cn; wuhb@pku.edu.cn; pjli@pku.edu.cn; fanwj@pku.edu.cn).

Huazhong Ren is with the Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871, China (e-mail: renhuazhong@pku.edu.cn).

Hongbo Qiao is with the College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China (e-mail: qiaohb@126.com).

Digital Object Identifier 10.1109/JSTARS.2021.3089203

***Index Terms*—Hyperspectral remote sensing, random forest, the CERES-wheat model, unmanned aerial vehicle (UAV), wheat yield estimation.**

I. Introduction

CROP yield is one of the most critical issues affecting national economic development and food security [1], [2]. Accurate and timely crop yield estimation is essential for precision agriculture and sustainable development, and can provide strong support for agricultural decision-making and management [3]. Traditional crop yield estimation heavily depends on ground field surveys, which are costly, time consuming and prone to large errors [4].

Since the 1970s, satellite remote sensing data have been broadly used for non-destructive crop yield estimation in large region scale [5], [6]. Numerous studies have taken an empirical approach based on vegetation indices [7]–[9], and showed that there is a linear relationship between crop yield and vegetation indices such as the normalized difference vegetation index (NDVI), a soil-adjusted vegetation index (SAVI), and green vegetation index. Previous studies [10]–[13] have also reported that there are relationships between crop yield and crop biophysical and biochemical parameters, such as the leaf area index (LAI), the fraction of absorbed photosynthetic active radiation (FPAR), and the leaf nitrogen content (LNC). Some studies concentrated on the empirical models which depended on the relationships between these parameters retrieved from remote sensing and crop yield to estimate the final crop yield [14], [15]. These models have successfully estimated the crop yield in the region scale from satellite imagery, and have been widely used due to their simplicity, calculation convenience, and acceptable accuracy. However, relationships established in this way are only applicable for local regions and specified time, and seldom involves the mechanism of crop growth.

The crop growth models are process-oriented and dynamic simulation models, providing a useful approach to simulate crop growth processes and obtain agricultural data at field plot scale, such as LAI, biomass, and yield [16], [17]. The widely used crop growth models include CERES [18] and GOSSYM [19] series models from the United States, SUCROS [20] series models from the Netherlands, and CCSODS [21] series models from the China. The accuracy of crop growth models is higher in simulating crop growth parameters, but the structure and process

are more complicated compared with the empirical models [22]. A large number of input parameters are needed to drive the crop growth models. Some input parameters are difficult to collect, and some parameters become unavailable because of the variability on region scale, so that these models are limited in its practical application [23]. Therefore, integration of remote sensing and crop growth models has become the highlights in the frontier of crop growth monitoring and yield estimation [24]. Some studies focused on assimilating remote sensed crop biophysical parameters into the crop growth model to obtain the estimated yield [25]–[27], so that the problem of spatial scale incompatibility can be solved. But this algorithm is complex, and depended on the time-effectiveness of satellite remote sensing.

In recent years, the development of sensor technology and technological advancements in unmanned aerial vehicles (UAVs) provide an advanced platform for data acquisition [28]. Compared with satellite remote sensing, UAV remote sensing has the advantage of high spatial-temporal resolution, low-cost, flexibility and versatility [29]. Some researchers have conducted research on crop yield estimation by UAV remote sensing based on vegetation indices methods. Geipel *et al*. [30] demonstrated that three vegetation indices calculated from UAV RGB images were highly correlated with corn grain yield. Zhou *et al.* [8] also analyzed the relationships between rice grain yield and several vegetation indices at the booting stage and multiple growth periods based on UAV imagery, and showed that there was a high correlation between grain yield and vegetation indices. But, the relationships are still locally applicable for only specified regions and time.

With the rapid development of artificial intelligence, the application of machine learning algorithms in crop yield estimation has been gradually increasing [6], [31], [32]. The machine learning including support vector machine, artificial neural networks, deep learning, and random forest, is an efficient empirical method for classification and prediction [33]. However, the amount of field measured yield data cannot meet the number of samples required for machine learning framework, and direct training with measured data may result in model overfitting. Among the current machine learning algorithms, the random forest algorithm was widely used in crop yield estimation at the large regional scale, and has achieved accurate predictions [34], [35], due to its unbiased estimation and suitability for small sample data [36]. Therefore, in this article, we introduced crop growth model simulation to increase the number of samples, then using the random forest algorithm to build a crop yield estimation model suitable for the UAV imagery.

The primary objectives of this article include: first, developing a new winter wheat yield estimation model (the CW-RF model) using the random forest regression algorithm in combination with the CERES-Wheat model suitable for UAV hyperspectral imagery; and second, testing the performance of the LAI and LNC retrieval methods for the hyperspectral sensor mounted on the UAV; third, assessing the potential for UAV remote sensing in yield estimation and analyze the possible error sources.

The remaining parts of this article are organized as: Section II introduces the study area, experimental design, image acquisition and processing, and field data collection; Section III presents the methods used in this article, including the CERES-wheat model, sensitive analysis, random forest algorithm, the development of CW-RF model, the LAI and LNC retrieval method and model validation; Section IV shows the retrieved and estimated results, and Sections V and VI provide discussion and conclusions, respectively.

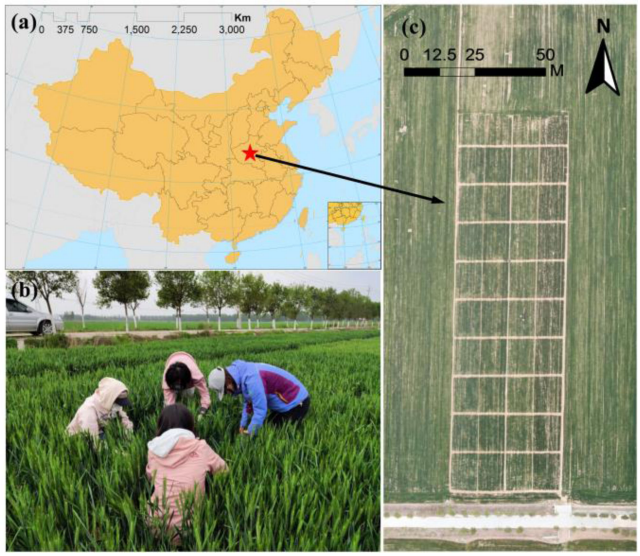


Fig. 1. Overview of the study area. (a) Geographical location of the experimental site. (b) Illustration of field data collection. (c) RGB orthomosaic imagery collected by a UAV on March 22, 2019 showing the spatial location of the 40 plots.

II. Materials

A. Study Area

The study was conducted at the experimental station of National Agriculture Production base for high quality wheat, which is located in Luohe, Henan Province, China, (113°52′54″E, 33°41′59″N) at an altitude of 63 m (see Fig. 1). The study area has a warm temperate humid continental monsoon climate, with mild cold winters, hot rainy summers, and short spring and autumn seasons. The annual average temperature, number of frost-free days, and precipitation are 14.6°C, 220, and 797.2 mm, respectively. The temporal distribution of rainfall is extremely uneven and is concentrated in the summer months, with summer precipitation accounting for approximately 70% of the annual precipitation. The annual sunshine duration is between 2187 and 2359 h. The predominant soil texture is moist and the organic matter content of the soil is 10–20 g/kg in the study area. The main crops planted in the study area are summer corn and winter wheat. The growth period of winter wheat is mainly from October to June of the next year.

B. Experimental Design

An experiment involving four different winter wheat cultivars and four different nitrogen application levels was designed for this article. Aikang 58, Zhoumai 27, Xinong 509, and Yumai 49-198 were selected as the test cultivars (see Fig. 2). The nitrogen

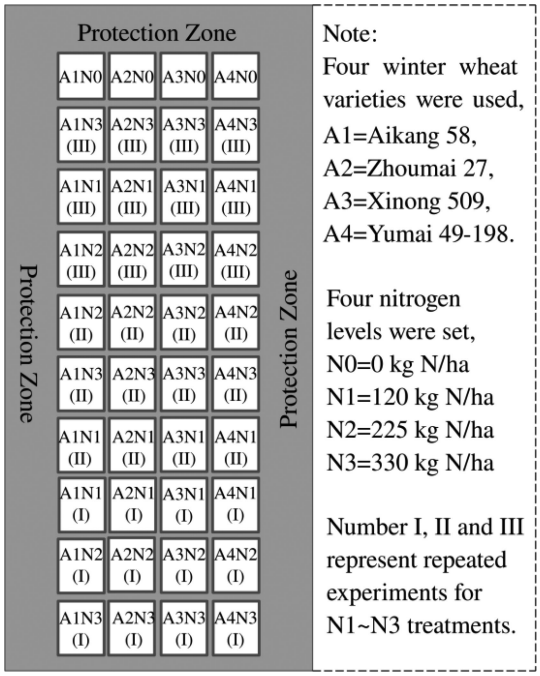


Fig. 2. Spatial distribution of whole plots in experimental area.

treatments were 0, 120, 225, and 330 kg/ha. Half of the nitrogen fertilizer was used as base fertilizer, with the other half applied at the jointing stage of the winter wheat. In addition, 135 kg/ha P_2O_5 and 105 kg/ha K_2O fertilizer supplements were applied to all field plots. The experiment had a randomized complete block design. As the control group, the 0 kg N per hectare treatment had only one block with an area of 100 m^2 (10 $\times$ 10 m), while the other treatments had three replications and each plot had an area of 130 m^2 (10 $\times$ 13 m). A total of 40 plots were used in the article. The spatial distribution of the plots is shown in Fig. 2. The wheat sowing date was October 17, 2018 and the wheat harvesting date was June 2, 2019. The field management practices were the same as those generally used in high yield wheat fields.

C. Image Acquisition and Processing

The UAV used in this article was the DJI Matrice 600 Pro (SZ DJI Technology Co., Ltd., Sham Chun, China) with six rotors, which performs in a very stable manner at low altitudes and low wind speeds [see Fig. 3(a) and (b)]. The maximum payload capacity of the UAV is 6 kg, and the maximum hover time without any load is 38 min. Its working environment temperature is between -10 and 40 °C. A hyperspectral imaging instrument, Pika L (Resonon Inc., Bozeman, MT, USA) was mounted on the UAV to acquire hyperspectral images [see Fig. 3(c)]. The Pika L has a short exposure time for push-broom imaging, weighs 0.6 kg, and measures 10.0 $\times$ 12.5 $\times$ 5.3 cm^3. It's operating range spans from the visible to the near-infrared, with a spectral range of 400$\sim$1000 nm, and its spectral resolution was 2.1 nm. Collected radiation data was recorded as a hyperspectral cube, with 900 samples and up to 2000 lines. Synchronously with the UAV-based hyperspectral data acquisition, an L1D-20C digital camera (Hasselblad Inc., Gothenburg, Sweden) mounted on another UAV (DJI Mavic 2 Pro; SZ DJI Technology Co., Ltd., Sham Chun, China), with four rotors, also acquired digital images. These digital images were used for the orthorectification of hyperspectral images.

The UAV flight was conducted under a clear sky, with cloudless and windless environmental conditions between 10:00 and 14:00 local time during the wheat tilling, jointing and heading stages (December 17, 2018, March 23, 2019 and April 19, 2019). An automatic control system was used to plan flight path. The flight altitude was 100 m and the speed was 3 m/s for the acquisition of hyperspectral and digital images. A total of 150 bands were selected for the hyperspectral imaging instrument and the exposure parameters of the sensor were set manually according to sunlight conditions. The hyperspectral images were continuously acquired during the flight at 90 fp/s and saved to a mobile hard drive. Before each flight, a calibration white plate (100% reflectance) and four 1.2 $\times$ 1.2 m standard targets with a fixed reflectance of 5%, 20%, 40%, and 60%, respectively, were placed on the ground within the UAV flight path and captured in the hyperspectral image. This was used for the radiation correction of hyperspectral images.

The process workflow for the hyperspectral images included radiometric calibration, radiometric correction, atmospheric correction, geometric correction and image mosaicking [37], [38]. Radiometric calibration was performed by a hyperspectral sensor calibration file using Spectronon software (Resonon Inc., Bozeman, MT, USA,), and the original digital number values of the images were converted to an apparent radiance. Radiometric correction was based on four standard targets. According to the relationship between the radiance of targets in airborne images and fixed reflectance, the apparent radiance data was transformed into surface reflectance data using a least square linear method. Second Simulation of a Satellite Signal in the solar spectrum (6S) model was adopted to carry out atmospheric correction. Geometric correction included both lens distortion correction and orthorectification. The georectify tool in Spectronon software was used for lens distortion correction. In addition, a digital orthophoto map (DOM) was generated from UAV-based digital images using Photoscan Professional software (Agisoft LLC., ST. Petersburg, Russia). Then, an orthographic correction was applied based on ground control points that were extracted manually between DOM and hyperspectral images. Finally, Spectronon software was used to mosaic images. The ortho hyperspectral images obtained previously were resampled to 0.1 m and projected into the WGS84 coordination system. Color composite images of hyperspectral data corresponding to the jointing and heading stages are shown in Fig. 4.

D. Field Data Collection

Field measurements and UAV flights were conducted simultaneously. The collected data mainly included the LAI, LNC, and the spectral reflectance of the canopy during the growth stage of winter wheat. Repeated nondestructive sampling was carried out in each plot for the determination of the LAI and spectra, while

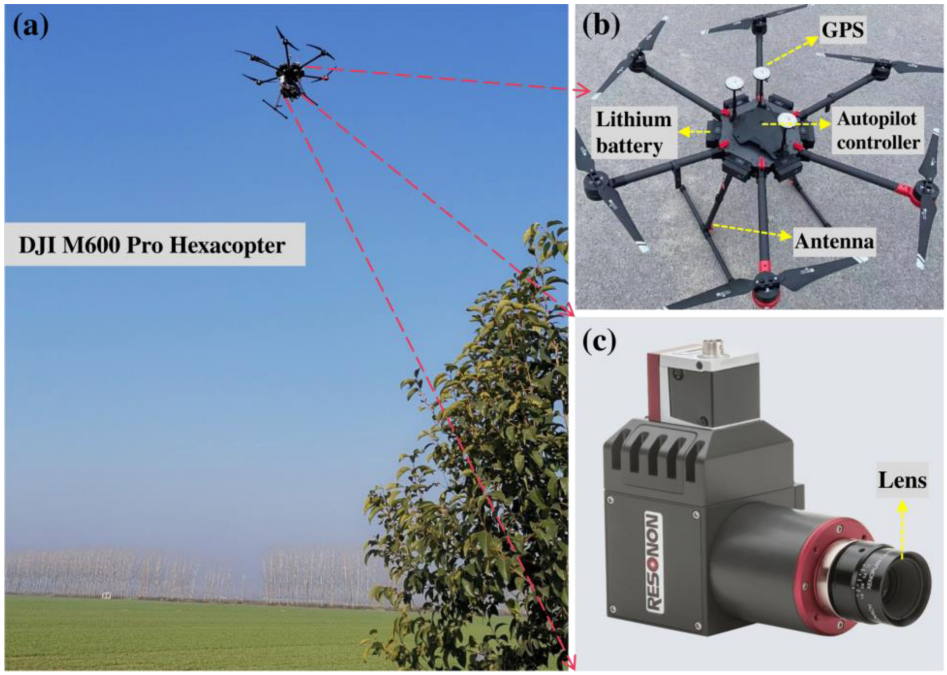


Fig. 3. UAV systems and hyperspectral sensor. (a) DJI M600 Pro Hexacopter platform. (b) Autonomous flight control system and telemetry and telecontrol system. (c) Pika L hyperspectral sensor.

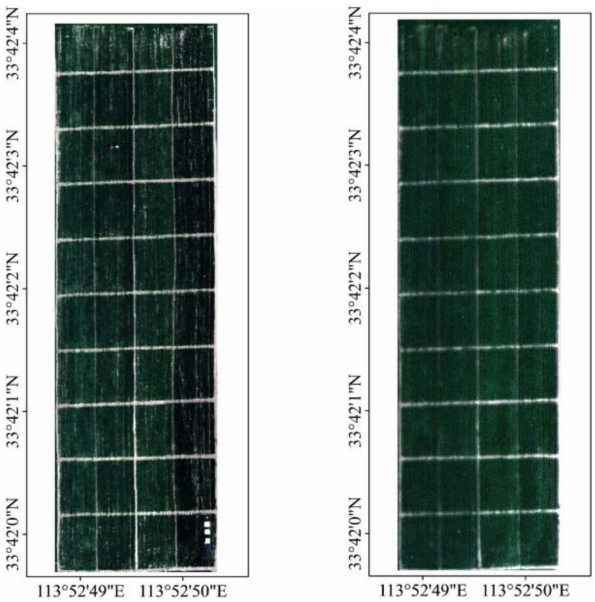


Fig. 4. True color composite images based on UAV hyperspectral data. (a) Jointing stage. (b) Heading stage. The wavelength of the red band is 659 nm, the wavelength of the green band is 549 nm, and the wavelength of the blue band is 479 nm.

destructive sampling was carried out in each plot for the determination of LNC. Three clusters from each plot were randomly selected to determine the LAI using a plant canopy analyzer (LAI-2000; LI-COR Inc., Lincoln, NE, USA). The LAI for each plot was calculated as an average. The ground hyperspectral data was measured using a field spectrometer (FieldSpec 3; Analytica Spectra Devices Inc., Longmont, CO, USA), with a spectral range of 350–2500 nm and spectral resampling interval of 1 nm. Calibration was conducted before and after the measurement using a calibration plate and was repeated ten times to obtain an average. Ten wheat plants were randomly cut with scissors in each plot, and their green leaves were killed at 105 °C for 30 min in the laboratory, then kept at 70–80 °C, dried, and crushed, before the total nitrogen content of leaves was determined. At winter wheat maturity, a 1 m^2 sampling area from each plot was harvested manually for the calculation of yield. The number of winter wheat spikes and grains in the sampling area was counted. The thousand grain weight was calculated after drying under sunlight, and the yield of winter wheat was determined.

III. Methodology

Fig. 5 shows a concept map for winter wheat yield estimation based on the integration of the CERES-Wheat model and random forest algorithm.

First, the parameters selected on the basis of typical conditions for the North China Plain, including local weather and soil conditions, plant characteristics and the general management information were input into the CERES-Wheat model, and abundant simulated data was produced. Second, based on the simulation results, the optimal growth parameters and periods for the wheat yield were selected using sensitive analysis. Then, the relationship between the optimal growth parameters at best stages and wheat yield was developed with the use of the random forest regression algorithm, which was called the CW-RF model. Besides, the parameters were retrieved from UAV hyperspectral

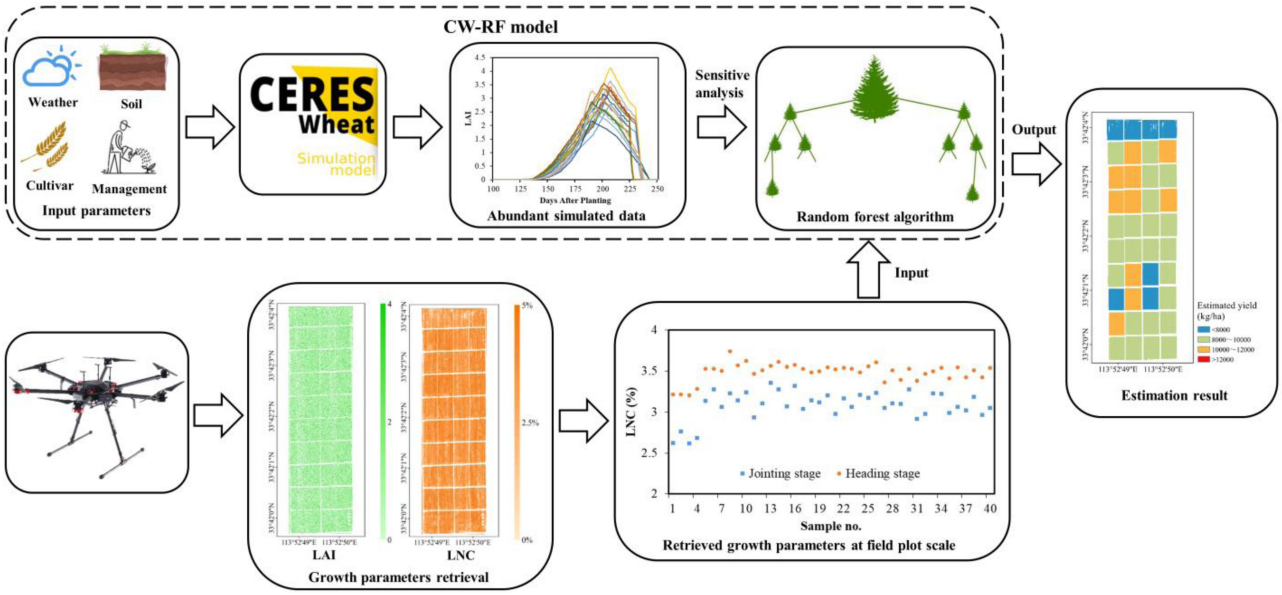


Fig. 5. Concept map for winter wheat yield estimation based on the integration of the CERES-Wheat model and random forest algorithm.

imagery using the proper method. Finally, the retrieved results were input into the CW-RF model, and the estimated winter wheat yield was obtained. Compared with the measured yield, the accuracy of the model was verified.

A. CERES-Wheat Model

The CERES-Wheat model, employed in this article, is a deterministic model to simulate the growth of wheat. It was originally developed under the auspices of the USDA-ARS Wheat Yield Project and the U.S. government multiagency AGRI-STARS program in the late 1970s [39]. The CERES-Wheat model is one of the main models incorporated in the decision support system for agrotechnology transfer (DSSAT), and simulates wheat growth, development, and yield data by considering photosynthesis, nutrient absorption, water absorption and transpiration, organ formation and senescence. DSSAT v.4.5 model was used for the study.

The main inputs of CERES-Wheat include weather and soil conditions, plant characteristics, and field management. In order to make the simulation data more suitable for the situation of the North China Plain, we considered inputting several sets of weather, soil, plant characteristics, and general management parameters adopted in the North China Plain into the CERES-Wheat model.

The minimum weather inputs of the model are daily solar radiation (SRAD, MJ m^{-2} day^{-1}), minimum and maximum air temperature (TMIN and TMAX,°C), and precipitation (RAIN, mm). These temperature and precipitation values were obtained from the China meteorological data sharing service system website, with the exception of solar radiation. The approximate daily solar radiation was calculated from the daily sunshine hours recorded on the same website. According to the Angstrom empirical formula [40], the expression of daily solar radiation is as follows:

$$R_s = R_{\max} \left(a_s + b_s \frac{n}{N}\right) \tag{1}$$

where R_s is daily solar radiation; $R_{\max}$ is daily astronomical radiation; n is daily sunshine hours; N is the theoretical daily maximum sunshine hours; and a_s and b_s are empirical coefficients values of $a_s = 0.18$ and $b_s = 0.55$ for the North China Plain, as recommended by the Food and Agriculture Organization of the United Nations (FAO).

The input soil dataset included general information, such as soil type, color, albedo, drainage, and runoff potential, and for each soil layer it included information such as the proportion of clay, silt, stones, the content of organic carbon, cation exchange capacity, total nitrogen, pH, bulk density, saturated soil water, and other factors.

The input field management information was obtained from the exact practices adopted in the experiment. In addition to basic planting information, such as the date of planting and harvesting, planting population, and planting depth, irrigation and fertilizer information were recorded and used in the model. The irrigation method was sprinkler irrigation and the main nitrogen fertilizer was urea. The irrigation and fertilizer amounts were set as variables.

Plant characteristics are generally determined by debugging the cultivar parameters file in the CERES-Wheat model. The definition, unit, and range of these parameters are given in Table I. The program is usually debugged by the generalized likelihood uncertainty estimation method, but in this article, different sets of calibrated wheat cultivar parameters were derived from several literature sources [41]–[43] and empirical values to simulate more universal results.

The output data of the CERES-Wheat model included the result variables, such as wheat yield and grain protein content,

and process variables such as LAI, FPAR, LNC, above-ground biomass, stem nitrogen content, root nitrogen content and other factors. The process data was recorded on a daily timescale.

B. Sensitive Analysis

Establishing the relationship between simulated multi-period crop growth parameters and simulated yield may lead to model redundancy. Therefore, we considered using a correlation analysis to select the optimal parameters of the best growth period.

The Sensitive analysis was utilized for select optimal growth parameters and periods. Sensitivity analysis is realized by calculating Pearson correlation coefficient. The basic function is represented as

$$\mathrm{r} = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{2}$$

where the r represents the Pearson correlation coefficient, X_i represents simulated wheat growth parameter in this article, Y_i represents simulated wheat yield, $\bar{X}$ and $\bar{Y}$ denotes the mean value of X_i and Y_i respectively.

First, we divided the whole growth process of wheat from sowing to maturity into eight stages: tilling stage, jointing stage, flagging stage, heading stage, flowering stage, filling stage, and maturing stage. Second, considering the growth parameters that can be effectively retrieved by remote sensing, the LAI, LNC, and FPAR were selected among the numerous simulated parameters in this article. Then, the data of each growth stage were averaged. Finally, the Pearson correlation coefficients of these parameters during different growth stages were calculated.

C. Random Forest Algorithm

The random forest is a popular machine learning algorithm based on a classification and regression tree (CART). The CART represents a decision tree structure in the form of binary tree formed by hierarchical organization of training datasets under a series of conditions or restrictions. The random forest can be used for estimating a categorical variable (classification), and also can be used for estimating a continuous variable (regression). The random forest regression algorithm has good performance in predicting high dimensional datasets and is extremely insensitive to noisy datasets [44].

The random forest begins with many bootstrap samples that are extracted randomly from the original training dataset. The regression trees are created by extracting a part of training samples through replacement, and some samples can be selected repeatedly while some samples are out of bag. A regression tree is fitted by a set of the bootstrap samples and each node per tree is chosen from a small set of input variables, which are selected randomly from the total dataset. The estimated value of an observation is calculated through averaging all of the trees.

The random forest regression algorithm has achieved effective results in various remote sensing studies, including agricultural monitoring and management applications. In this article, to model the relationship between wheat growth parameters and wheat yield, random forest algorithm was run using the random forest package (v.4.6-14) in R software (v.3.6.1). Two main parameters were defined and optimized in the random forest algorithm: the number of trees that was created ("ntree" parameter; default value is 500 trees), and the number of different variables for tree node splitting ("mtry" parameter; default value is one-third of the total number of variables). The random forest algorithm was conducted as follows.

1) *Setting proper ntree*: For each regression tree, ntree bootstrap samples are randomly extracted from the original data as the training dataset;
2) *Setting proper mtry*: If the feature dimension of samples is M, a constant mtry $<$ M needs to be specified, and mtry variables are randomly selected from M variables. For each node per tree, the best split is chosen among mtry variables.
3) Each tree is developed to its maximum expansion, and new data is estimated by averaging the estimation results of all trees.

D. Development of CW-RF Model

In this article, we designed four sets of different weather parameters, which were obtained from the annual mean weather data of four provincial meteorological stations on the North China Plain. These weather stations are located in Beijing, Luohe (Henan Province), Weifang (Shandong Province) and Xuzhou (Jiangsu Province). These four sites are all located in the North China Plain with certain representativeness.

In addition, we designed five sets of different soil parameters, which were selected from typical soil types of five provinces in the North China Plain. The five provinces are Beijing, Henan, Shandong, Jiangsu and Anhui Province. The soil data in Beijing came from field measurements, and Henan Province came from local soil records, and Shandong, Anhui and Jiangsu Province were from the China Soil Science Database website (http://vdb3.soil.csdb.cn/).

Ten sets of different wheat cultivar parameters, which came from literature and prior knowledge, seven sets of different nitrogen fertilizer parameters, which were 0, 60, 120, 180, 225, 270, 330 kg/ha, respectively, and three set of different irrigation parameters, which were 10, 30, and 50 mm, were set in this article. These four different types of data were combined and input into the DSSAT software (with the CERES-Wheat model encapsulated), and a total of 4200 groups of output data are produced.

The Pearson correlation coefficient between these growth parameters and simulated yield were analyzed (see Table II). The correlation between both the LAI and LNC and yield were significantly higher than the correlation between FPAR and yield, and therefore the LAI and LNC were selected as the best growth parameters. The strongest correlation between the LAI and yield was found in the heading period of wheat, while the strongest correlation between the LNC and yield was found in the jointing period of wheat. Therefore, the jointing and heading period were selected as sensitive growth periods in this article. The data distribution of the simulated LAI, LNC, and yield is shown in Fig. 6. It shows the simulated data was suitable for training the random forest algorithm.

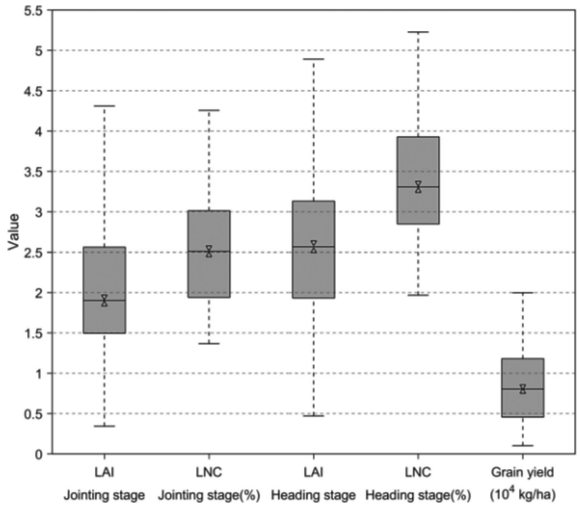


Fig. 6. Box-plot of simulated LAI, LNC, and wheat yield.

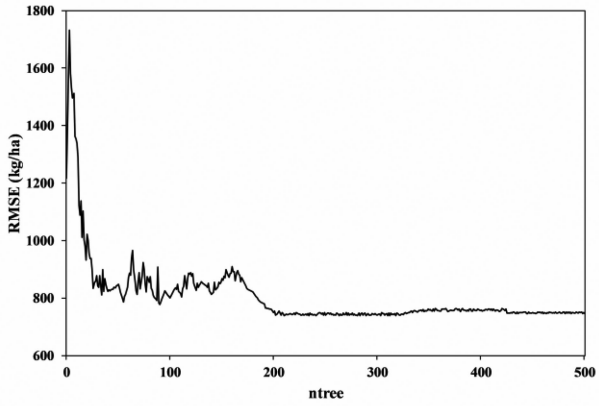


Fig. 7. Association between RMSE and ntree used in random forest.

The simulated LAI and LNC at jointing and heading stages and wheat yield data were input into the random forest regression algorithm, the CW-RF model was developed. For the three stages of random forest algorithm, the parameter values (ntree and mtry) were optimized using the simulated training dataset. The RMSE of the model and out-of-bag error were used to find the values that could best estimate the winter wheat yield. The out-of-bag error was lowest when mtry was set to half of the total variables. When ntree $>$ 200, RMSE roughly reached the minimum and changed smoothly (see Fig. 7). The ntree was set to 200 to improve the calculation efficiency. According to the results, ntree was set to 200, and mtry was set to 2.

Optimal parameters were used to generate variable importance. The importance of variables was evaluated by the influence of variables on estimation accuracy, which allowed for a quick assessment of the relevance between predictors and wheat yield. In this article, percentage increase in mean square error (%IncMSE) was used to measure the importance of variables (see Fig. 8). The results show that LNC at heading stage is the most important explanatory variable in the CW-RF model,

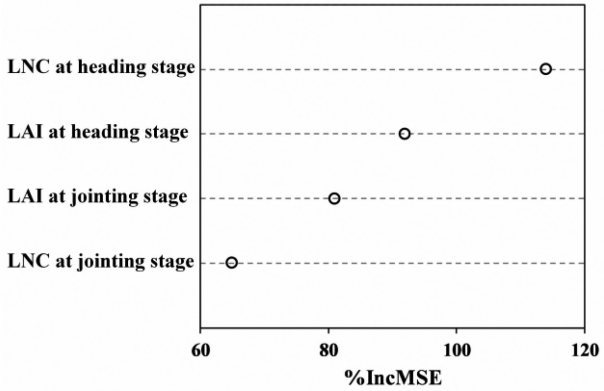


Fig. 8. Variable importance plot generated by the random forest algorithm.

followed by LAI at heading stage and LAI at jointing stage, and finally LNC at jointing stage.

E. Remote Sensing Retrieval Method

The LAI of winter wheat in UAV hyperspectral remote sensing images was retrieved by the directional second derivative (DSD) method [45], while the LNC of winter wheat was retrieved based on the novel angular insensitivity vegetation index (AIVI) [46].

The DSD is an algorithm that can effectively eliminate the soil background effect and bidirection effect (view direction and solar direction), and then retrieve the LAI. The spectrum analysis of leaf and soil indicated that the DSD value of leaf $\rho_v^{''}$ was much larger than that of the soil spectrum $\rho_g^{''}$ within the 0.68–0.71 and 0.73–0.75 μm bands. When $\rho_v^{''} \gg \rho_g^{''}$, the second derivative of reflectance for the object ρ'' can be expressed as

$$\frac{\rho''}{\rho_v^{''}} = 1 - \left[1 - \frac{E_d}{\mu_0 F_0 + E_d}\left(1 - e^{-\lambda_0 \frac{G_v}{\mu_v}(1-\Gamma(\Phi))LAI}\right)\right] e^{-\lambda_0 \frac{G_v}{\mu_v} LAI \Gamma(\Phi)} \tag{3}$$

where G_v is the G function of the view direction; $\mu_v = \cos\theta_v$, θ_v is the viewing zenith angle; λ_0 is the Nilson parameter, which is used to describe the clumping effect of foliage; $\Gamma(\Phi)$ is an empirical function used to describe the hot-spot effect; $\mu_0 F_0$ is the direct irradiance of the sun, and E_d is the diffuse irradiance of the atmosphere. When $\Phi = 0$ and $\Gamma(\Phi) = 1$, if $\mathrm{b} = \lambda_0 G_v/\mu_v$, the expression can be simplified as follows:

$$\frac{\rho''}{\rho_v^{''}} = 1 - e^{-b\mathrm{LAI}}. \tag{4}$$

Equation (4) shows the approximate correlation between the second derivative and the LAI. Based on this equation, the winter wheat LAI was retrieved rapidly and accurately, removing the angle and soil background effects.

In this article, the AIVI was chosen to retrieve LNC [46]. The AIVI is not sensitive to angle and has a strong correlation with LNC of winter wheat. Based on red-edge, blue, and green bands, AIVI overcomes the influence of different experimental conditions and view zenith angles. The results showed that the

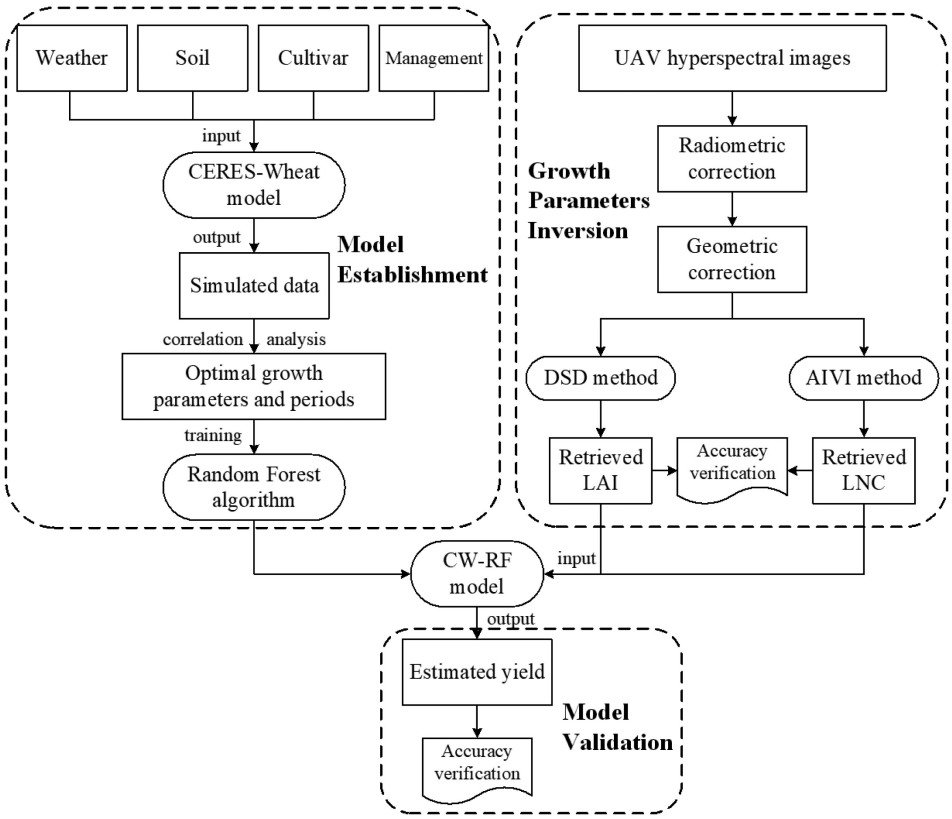


Fig. 9. Framework for the development of CW-RF model, crop growth parameters retrieval and model validation.

AIVI had a stronger association with LNC compared to traditional vegetation indices, suggesting that the linear relationship between AIVI and LNC would be more stable and accurate. The AIVI is expressed as follows:

$$\text{AIVI} = \frac{R_{445} \times (R_{720} + R_{735}) - R_{573} \times (R_{720} - R_{735})}{R_{720} \times (R_{573} + R_{445})} \tag{5}$$

where R_i is reflectance at i wavelength (nm).

F. Model Validation

The accuracy verification mainly consisted of two indexes, the mean-absolute-percent error (MAPE, %) and the root-mean-squared error (RMSE). The MAPE and RMSE were calculated using the following formula

$$\text{MAPE} = \sum_{i=1}^{m} \left| \frac{\hat{y} - y}{y} \right| \times \frac{100}{m} \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)^2} \tag{7}$$

where $\hat{y}$ represents estimated value, y represents measured value, and m is the number of samples.

Fig. 9 shows the framework for winter wheat yield estimation in this article. The algorithm framework mainly includes the development of CW-RF model, crop growth parameters retrieval and model validation.

IV. Results

A. Retrieval Results

1) Wheat LAI Retrieved From UAV Hyperspectral Data: The pixels of the UAV-based hyperspectral images were resampled to 1 m. The DSD method was applied to the UAV hyperspectral images when parameter *b* in simplified formula approximately was 0.5 [45], inversion maps of wheat LAI during the jointing and heading stage were obtained (see Fig. 10).

It can be seen from the ridges between the plots that the soil information was mostly removed and the vegetation information was retained. Each plot was basically covered by winter wheat and the wheat LAI varied from 0 to 4. In addition, the wheat LAI of each plot was obtained by averaging the LAI of each pixel in the plot.

The measured LAI of 40 field plots were all used to validate the accuracy of wheat LAI retrieval. A comparison of the LAI retrieval and measured values are shown in Fig. 11. In both the jointing stage and heading stage of the wheat, the scattered points were evenly distributed around the 1:1 line, the RMSE were 6.27% and 9.21% respectively, and the MAPE were 2.62% and 3.85% respectively. This showed that the retrieved LAI values were close to the measured values, which indicated the reliability of the DSD algorithm.

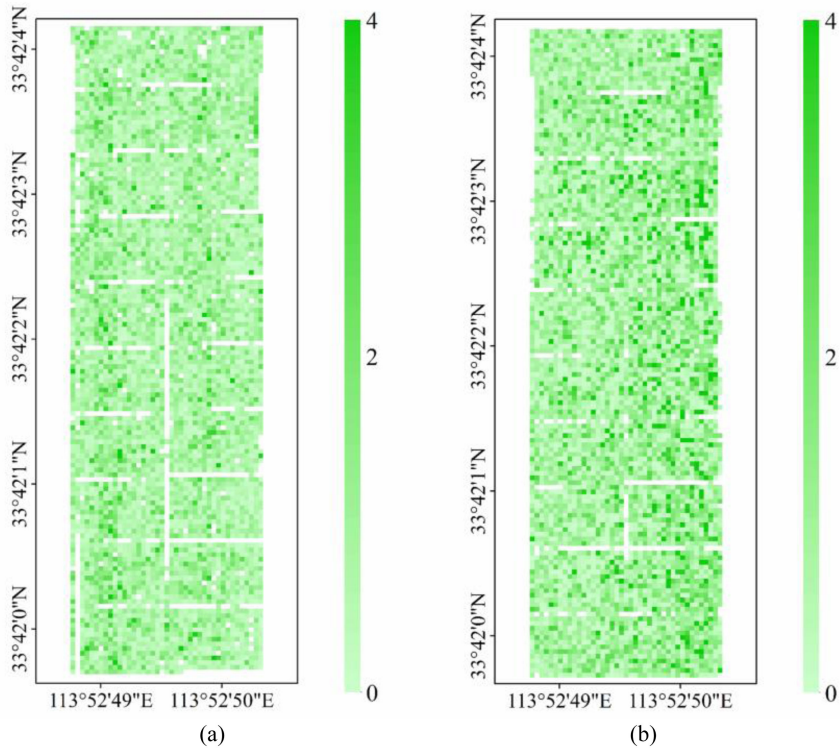


Fig. 10. LAI spatial distribution map retrieved from UAV-based hyperspectral data. (a) Jointing stage. (b) Heading stage.

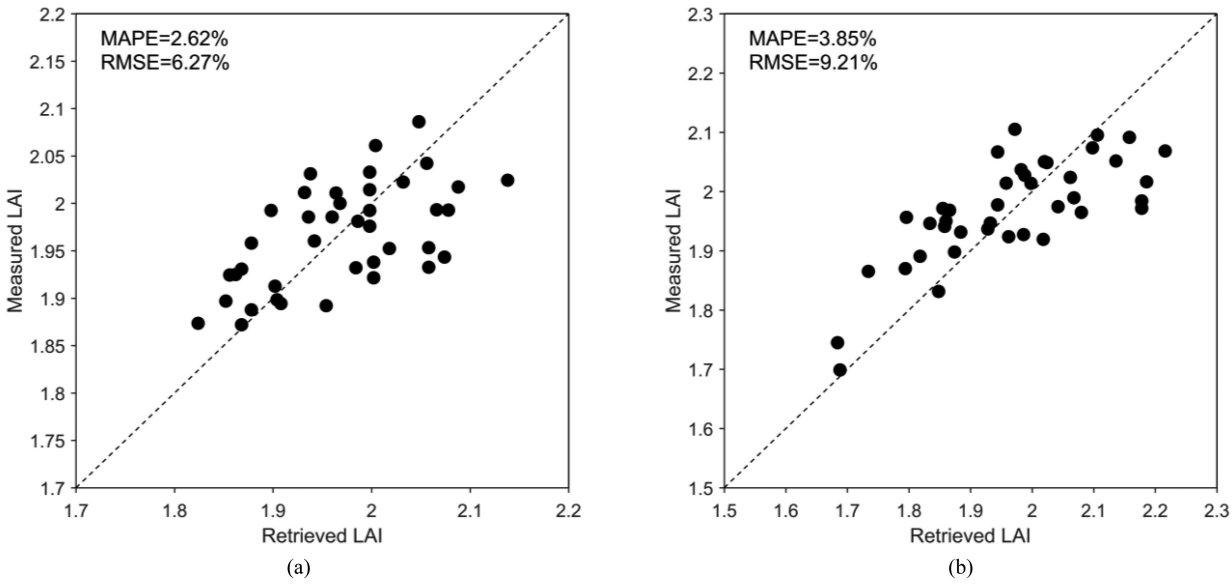


Fig. 11. Comparison between the retrieved and measured LAI at the jointing and heading stages. (a) Jointing stage. (b) Heading stage.

2) Wheat LNC Retrieved from UAV Hyperspectral Data: The pixels of UAV-based hyperspectral images were resampled to 1 m, and the AIVI was calculated from the UAV hyperspectral images at the jointing and heading stages of wheat.

Due to the small number of measured samples, the leave-one-out cross validation was applied in LNC retrieval. One measured sample was left for validation at a time, and the remaining measured samples were used to calibrate the LNC model in one training. According to the measured LNC and AIVI of the corresponding pixels, the relationship between the wheat LNC and AIVI was obtained by linear regression. The validation results were combined into a complete inversion map, and Fig. 12 showed the retrieved wheat LNC during the jointing and heading stages.

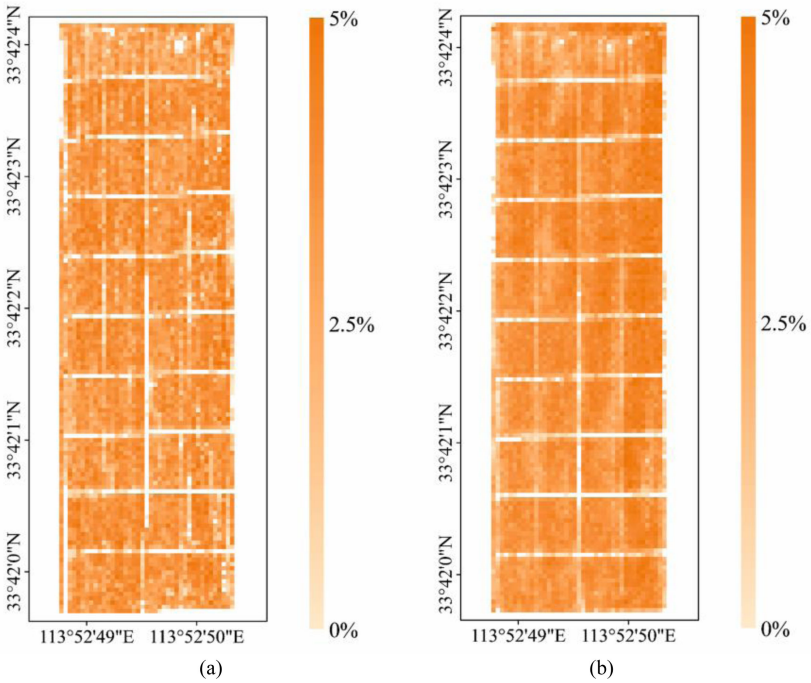


Fig. 12. LNC spatial distribution map retrieved from UAV-based hyperspectral data. (a) Jointing stage. (b) Heading stage.

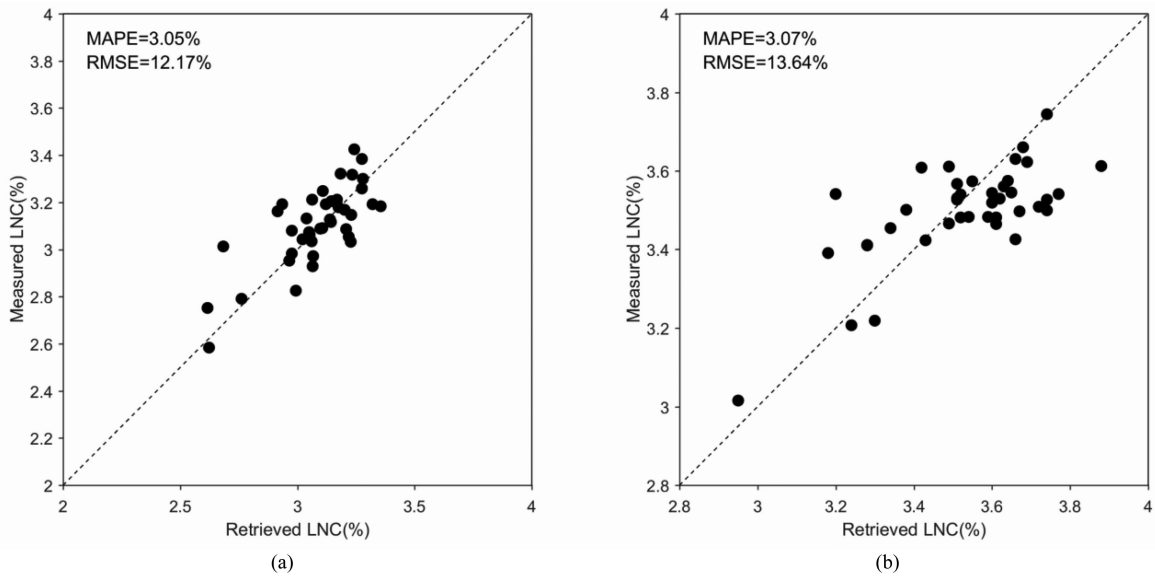


Fig. 13. Comparison between the retrieved and measured LNC at the jointing and heading stages. (a) Jointing stage. (b) Heading stage.

The LNC retrieval results also retained vegetation information and eliminated soil information by setting thresholds for the AIVI. Fig. 12 shows that the wheat LNC varied from 0 to 5%. The wheat LNC of each plot was obtained by averaging the LNC of each pixel in the plot. A comparison of the results of the LNC retrieval and measured values are shown in Fig. 13. The RMSE at the jointing and heading stages of wheat were 12.17% and 13.64%, respectively, and MAPE were 3.05% and 3.07%, respectively, which confirmed the reliability of winter wheat LNC retrieval using AIVI method.

B. *Estimated Yield Results*

Based on the retrieval values of winter wheat LAI and LNC in each plot, the estimated yield in each plot was acquired using the CW-RF model. Fig. 14 shows a spatial distribution map of the measured yield, estimated yield, and error percentage,

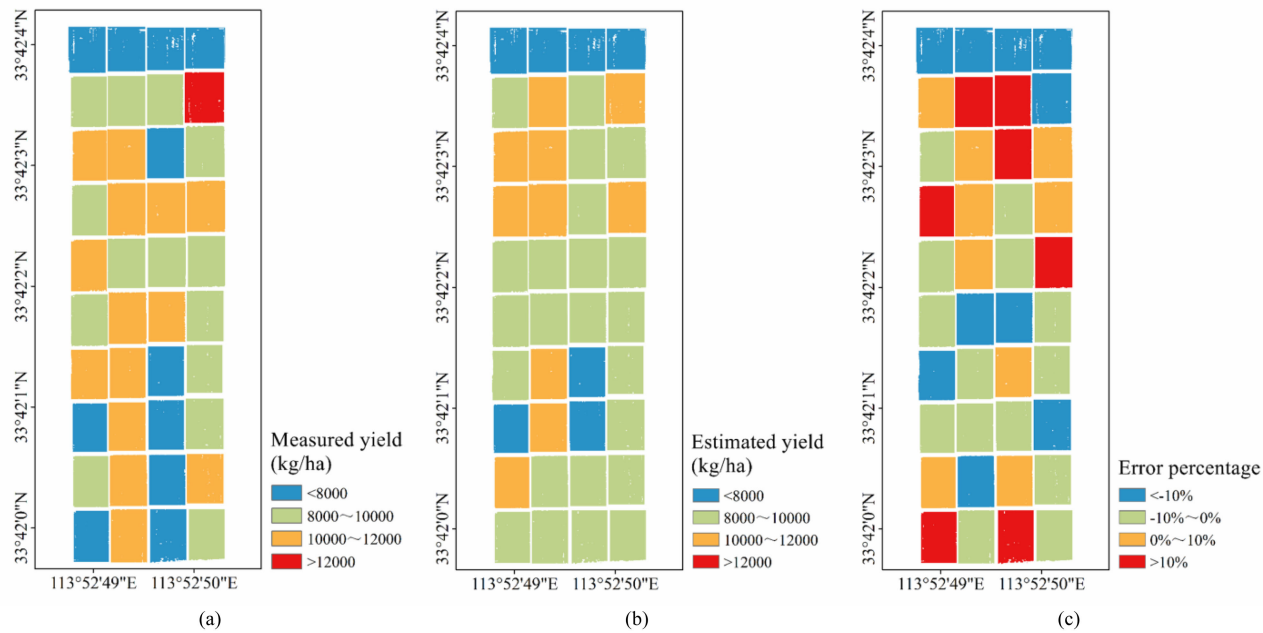


Fig. 14. Yield spatial distribution map. (a) Measured yield. (b) Estimated yield. (c) Error percentage.

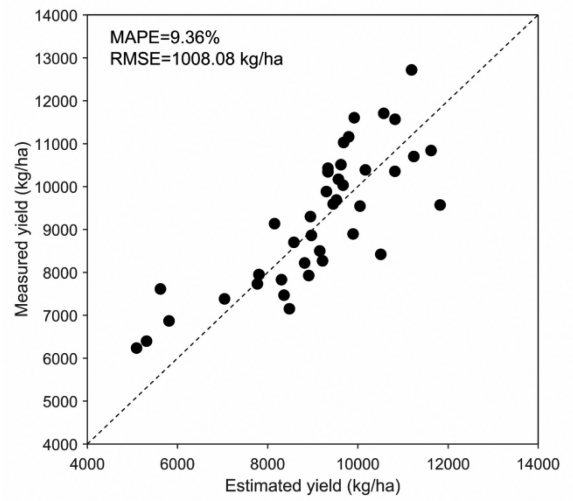


Fig. 15. Relationship between measured and estimated winter wheat yield.

respectively. It can be seen that the absolute error of most plots was less than 10%. Fig. 15 shows the comparison of estimated and measured yield. According to the statistical analysis, the MAPE was 9.36% and RMSE was 1008.08 kg/ha, which showed that the model established in this article performed well and was capable of plot scale yield estimation.

V. Discussion

A. Error Analysis

According to the concept of CW-RF model, crop growth model simulation, sample representation and observation are three main aspects that affect the uncertainties of crop yield estimation [47]. Although the crop growth models (e.g., the CERES-Wheat model) have robust physical and biological explanations for specific crop yield, the bias of these models may be increased due to the large uncertainties in the spatial distribution of weather and soil conditions, plant characteristics, and field management [48]. The uncertainties of input data may be transferred to the model simulation results [49]. Therefore, it is necessary to analyze and summarize the uncertainties of the CERES-Wheat model.

The simplification of the actual growth process of winter wheat, the driving parameters given by experience and prior knowledge, and the regional heterogeneity of some initial conditions, lead to large uncertainties in simulation results of the CERES-Wheat model. Especially in some extreme weather and crop growth conditions, the simulation results deviate greatly from the reality of crop growth. Xiong [50] pointed out that the CERES-Wheat model could not reflect the effects of diseases, pests and weeds, and could not simulate satisfactory results under meteorological disaster and extreme environment stress situation. In addition, the uncertainties caused by spatial heterogeneity of driving parameters are aggravated when crop growth models are applied to yield estimation in regional scale. Observations can provide relatively real growth conditions of crops, and integrating the advantages of crop growth models and observations reduce the uncertainties of yield estimation.

Remote sensing provides larger range of observations compared to the site scale, which can reduce the uncertainties from spatial heterogeneity. Therefore, observation data from remote sensing is currently widely used in crop growth models for yield estimation, and most of studies on the uncertainty of the crop growth model were focus on the model input variables which were closely related to remote sensing [51]. Li *et al.* [52] found that the higher crop parameters retrieval accuracy, the smaller

uncertainties of crop growth models and yield estimation error. Therefore, accurate retrieval of crop parameters is a necessary guarantee for crop growth models.

Since the data source of the CW-RF model was the simulation results of the CERES-Wheat model, the simulation errors of the CERES-Wheat model could be transferred to the CW-RF model. In order to deal with these problems, we hope to introduce more field measurements in the future research, which has two benefits: (the field measurements could be used as the training data of the CW-RF model to reduce the simulation errors and uncertainties brought by the CERES-wheat model; and the field measurements could be used to calibrate the CERES-wheat model to reduce the simulation error and uncertainties.

UAV-based retrieval LAI and LNC were used as the input of the CW-RF model, the retrieval accuracy directly affected the accuracy of CW-RF model for yield estimation. Although UAV operation was strictly in accordance with the standards during the stage of hyperspectral image acquisition, retrieval errors were still unavoidable due to uncertain factors, such as atmospheric conditions and mixed pixels. The influence of parameters retrieval error on the CW-RF model was further quantitatively analyzed. Additional 5% and 10% random MAPE were added to the remote sensed LAI and LNC of winter wheat at the jointing and heading stages, respectively. The RMSE of the LAI and LNC were 8.81% and 15.37% under 5% random MAPE, and 13.72% and 22.83% under 10% random MAPE, respectively. A total of eight groups of data with random errors were generated and input into the CW-RF model. A comparison of the results was given in Table III. The results showed that with an increase in the LAI and LNC retrieval error, the MAPE and RMSE of the CW-RF model also increased, and the accuracy of CW-RF model decreased, demonstrating that the accuracy of the remote sensed LAI and LNC was essential for ensuring the estimation accuracy of the CW-RF model. The importance of the UAV hyperspectral remote sensing retrieval accuracy to the CW-RF model was also indicated. Based on a standard MAPE of less than 20% and RMSE of less than 1500 kg/ha, we believed that the accuracy of the CW-RF model could be guaranteed when the MAPE and RMSE of the remote sensed LAI were less than 10% and 13.72%, respectively, and the MAPE and RMSE of the LNC was less than 5% and 15.37%, respectively.

In addition, the empirical methods for the LAI and LNC retrieval may dilute transferability of the CW-RF model and also affect the accuracy. The concise and physical model to estimate LAI and LNC from UAV-based hyperspectral data should be recommended to avoid the effect of solar-sensor bidirectional reflectance distribution function (BRDF) and background reflectance variation.

The selection of the growth period has a very important influence on the accuracy of winter wheat yield estimation models [53]. The jointing and flagging stages are the key periods for determining the number of ears and grains of winter wheat, while the heading and filling stages are the key periods determining the final grain weight. The CERES-Wheat model simulation results showed that the LAI and LNC at the jointing and heading stages had the highest correlation with yield, followed by the flagging and filling stages, but the difference was not significant. The results coincided with the previous study by Yue *et al.* [54]. The study reported that during the later stages of crop growth, the chlorophyll content of leaves decreases, and the correlation between crop growth parameters and dry matter accumulation decreases. The selection of growth periods may have an impact on the retrieval results, therefore we added more growth periods into the CW-RF model. The MAPE and RMSE were calculated after adding the flagging and flowering stages based on the jointing and heading stages in the CW-RF model (see Table IV).

The results showed that compared with the jointing and heading stages in the CW-RF model (MAPE = 9.36%, RMSE = 1008.08 kg/ha), adding one or more of the flagging and flowering stages would slightly improve the estimation accuracy, which was consistent with the conclusion that the accuracy of crop yield estimation can be improved by using multitemporal data in some studies [2], [8], [55]. Considering the cost of UAV flights and the time requirements for earlier yield estimation for agricultural decision-making, the use of the jointing and heading stages were identified as a tradeoff for yield estimation in the CW-RF model.

B. Applicability of the CW-RF Model

In this article, the CW-RF model was used to estimate the wheat yield. Hyperspectral remote sensing data collected by UAV and available from the National Precision Agriculture Research and Demonstration Base in Xiaotangshan, Beijing (116°23′50″E, 39°54′59″N, located in the North China Plain) was used to verify the applicability of the model. The data was acquired in 2018, and the UAV type, hyperspectral sensor type, data type, and processing workflow were same as those used in this article in Luohe.

Several prevailing winter wheat yield estimation models based on a statistical correlation analysis [2], [56], [57] were selected for comparison with the CW-RF model. These empirical models were based on the relationship between wheat yield and remote sensed LAI, single-growth period vegetation index, and multi-growth periods vegetation index, respectively. LAI was obtained based on the UAV hyperspectral images during the wheat heading period according to Ren *et al*. [56], and NDVI was selected as the vegetation index for modeling. The wavelength of red band used to calculate NDVI was 650 nm, and the wavelength of near-infrared band was 857 nm. The calculation of single-growth period NDVI was based on the images of wheat heading stage according to Reyniers *et al*. [57], while the multigrowth period NDVI was based on the images of wheat jointing and heading stages according to Wang *et al*. [2]. The linear relationships among these models and their RMSE and MAPE values are given in Table V.

These empirical statistical models are easy to implement and have a high calculation efficiency, but they lack the ability to describe physical mechanisms. Taking the empirical model 1 in Table V as an example. The RMSE of the model built in Luohe was 1025.42 kg/ha, and the MAPE is 9.74%. The RMSE of the model directly applied to Xiaotangshan is 2112.99 kg/ha, the MAPE is 34.03%, the RMSE has increased by 106.06%, and the MAPE has increased by 249.38%. Similar situations also

TABLE I
CULTIVAR PARAMETERS OF WHEAT IN THE CERES-WHEAT MODEL

Parameters	Definition	Unit	range	Initial value
P1V	Days required for vernalization at the optimum temperature	days	5~65	35
P1D	Photoperiodic response parameter	%	0~95	50
P5	Accumulated temperature at filling stage	°C· day	300~800	600
G1	Kernel number per unit canopy weight at anthesis	kernel·g^{-1}	15~30	26
G2	Standard kernel size under optimum conditions	mg	20~65	26
G3	Standard, non-stressed mature tiller weight	g	1~2	1.5
PHINT	Accumulated temperature required for the growth of leaves	°C· day	60~100	100

TABLE II
CORRELATIONS BETWEEN WHEAT YIELD AND GROWTH PARAMETERS IN THE DIFFERENT GROWTH STAGES

Parameter	Tilling	Jointing	Flagging	Heading	Flowering	Filling	Maturing
LAI	0.1375	0.7446	0.7263	0.7353	0.6693	0.7332	0.7249
LNC	0.1699	0.6979	0.7586	0.8185	0.7768	0.8052	0.8067
FPAR	0.1557	0.2130	0.3162	0.3885	0.3621	0.3938	0.3813

TABLE III
MAPE AND RMSE OF THE CW-RF MODEL UNDER THE DIFFERENT MAPE ERROR AND RMSE ERROR OF PARAMETERS

LAI		LNC		CW-RF model	
MAPE (%)	RMSE (%)	MAPE (%)	RMSE (%)	MAPE (%)	RMSE (kg/ha)
0	0	5	15.37	12.20	985.05
0	0	10	22.83	18.76	1297.56
5	8.81	0	0	8.51	626.30
5	8.81	5	15.37	15.86	1076.80
5	8.81	10	22.83	22.67	1707.94
10	13.72	0	0	11.87	1081.35
10	13.72	5	15.37	18.17	1270.81
10	13.72	10	22.83	29.06	2102.11

TABLE IV
MAPE AND RMSE VALUES OF THE CW-RF MODEL WHEN MORE GROWTH STAGES WERE ADDED

Growth stages considered in CW-RF model	MAPE (%)	RMSE (kg/ha)
Jointing and heading stages	9.36	1008.08
Jointing, flagging and heading stages	8.83	972.14
Jointing, heading and flowering stages	9.16	996.46
Jointing, flagging, heading and flowering stages	8.98	985.83

TABLE V
COMPARISON BETWEEN PREVIOUS STATISTICAL MODELS AND THE MODEL USED IN THIS ARTICLE

Model	Data source	Linear relation	Luohe		XiaoTangshan	
			RMSE (kg/ha)	MAPE (%)	RMSE (kg/ha)	MAPE (%)
CW-RF	CERES-Wheat	Non-linear	1008.08	9.36	841.36	13.64
Model1	Luohe	Yield=11256×LAI-12971	1025.42	9.74	2112.99	34.03
	XiaoTangshan	Yield=12770×LAI-17775	2205.08	20.11	1428.08	18.65
Model2	Luohe	Yield=33759×NDVI-19108	1283.52	10.62	1869.77	22.47
	XiaoTangshan	Yield=22933×NDVI-11971	2397.25	20.14	1689.49	19.34
Model3	Luohe	Yield=$NDVI_{joint}$×37822.74-$NDVI_{heading}$×6564.65-12029.9	1020.24	9.68	1617.60	20.60
	XiaoTangshan	Yield=$NDVI_{joint}$×30254.15+$NDVI_{heading}$×10659.88-21197.3	1071.39	10.16	1129.97	16.36

Model 1: Used for establishing the relationship between wheat yield and remote sensed LAI.
Model 2: Used for establishing the relationship between wheat yield and the single-growth period vegetation index.
Model 3: Used for establishing the relationship between wheat yield and the multi-growth periods vegetation index.

TABLE VI
COMPARISON BETWEEN UAV-BASED HYPERSPECTRAL AND MULTISPECTRAL DATA

Crop growth parameter	Spectral types	Jointing stage		Heading stage	
		RMSE (%)	MAPE (%)	RMSE (%)	MAPE (%)
LAI	multispectral	6.81	2.81	9.86	4.06
	hyperspectral	6.27	2.62	9.21	3.85
LNC	multispectral	15.38	4.11	15.82	3.66
	hyperspectral	12.17	3.05	13.64	3.07

appeared in several other empirical models. It can be deduced that the previous statistical models tend to be applicable locally. In this article, the yield estimation model had a high accuracy in both Luohe and Xiaotangshan, and performed well in these two areas. The input data of the CW-RF model originated from the CERES-Wheat model, which can reflect a series of physiological and biochemical processes of crops, and it was not limited by empirical statistics.

The spatial scale of the CERES-Wheat model was consistent with the field plot scale used in this article. This matching of scale ensured that the yield estimation model trained by simulated data was suitable for UAV remote sensing without assimilation.

In theory, the idea of UAV-based CW-RF model in this article could be applied to satellite data, but scale suitability problem still existed when the CW-RF model was directly transferred to coarse resolution satellite data. It should be further explored how the CW-RF models and remote sensing retrieval methods in the study could adapt to high-resolution satellite data in the future.

In addition, as a representative integrated learning technique, the random forest algorithm used for modeling in this article was inclusive of all the samples. The few outliers and missing values did not affect the results, which also ensured the stability of the model. Thus, the CW-RF model has a wide applicability and is considered to perform well for winter wheat yield estimation in the North China Plain.

C. Advantages of Hyperspectral Data

With the continuous development of hyperspectral technology, the application of hyperspectral remote sensing in precision agriculture has become an area of active research. Hyperspectral data can accurately describe the spectral details and provide abundant spectral information. Based on sensitive band analysis, vegetation index construction, and red-edge parameter analysis methods, some progress has been made in the retrieval of crop growth parameters, such as the LAI [58], nitrogen content [59] and yield [60]. Some previously published results [61], [62] indicated that hyperspectral data can improve the inversion accuracy of crop growth parameters compared with multispectral data, but a few studies like [63] have come to the opposite conclusion. Liu *et al*. [58] believed that this was due to the low signal-to-noise ratio of sensors, the low spatial resolution of the remote sensing, and the adoption of methods such as vegetation indexes that failed to reflect the advantages of hyperspectral data.

For the UAV platform, whether the hyperspectral data can improve the retrieval accuracy is worth exploring. The RMSE and MAPE of the retrieved LAI and LNC with hyperspectral and multispectral data is given in Table VI. The multispectral data was acquired by convoluting hyperspectral data with the spectral response function of the Pika L sensor, and vegetation indices, such as the NDVI, medium resolution imaging spectrometer, medium terrestrial chlorophyll index , and RED-EDGE NDVI [64] were calculated by multispectral data. To obtain the LAI and LNC, empirical linear relationships between the LAI and LNC of winter wheat and these vegetation indices were established, which were based on local training, without considering the BRDF characteristics of the canopy and background.

The results showed that the precision of the LAI and LNC retrieved from hyperspectral data was higher than that retrieved from multispectral data (see Table VI). The DSD method for LAI retrieval made full use of the advantages of the red-edge band in hyperspectral data, and effectively restrained the influence of soil background noise. The selection and combination of sensitive bands of AIVI for LNC retrieval was reliant on the hyperspectral data. In addition, the high spatial resolution and stability of the UAV platform also improved the stability of hyperspectral data. According to previous discussion, the accuracy of remote sensing data directly affected the accuracy of the CW-RF model, and this effect was positively correlated. The spectral resolution of the hyperspectral sensor on the UAV had the potential to improve the wheat LAI and LNC estimation accuracy, and then the accuracy of CW-RF model may be improved, which required further study.

VI. CONCLUSION

Estimating winter wheat yield early and accurately at field plot scale is of great significance for field management and agricultural operation. In this article, a winter wheat yield estimation model, the CW-RF model, was established based on the CERES-Wheat model simulation data using the random forest regression algorithm. A total of 4200 groups of samples were simulated using the CERES-Wheat model, involving four different weather parameters, five different soil parameters, ten different wheat cultivars, seven different nitrogen fertilizer parameters and three irrigation parameters that were based on the current situation in the North China Plain. The jointing and heading stages were identified as the two key growth periods, and the LAI and LNC were chosen as the main growth parameters for winter wheat yield estimation. Therefore, the LAI and LNC of winter wheat during the jointing and heading stages retrieved

from UAV hyperspectral images were input into the CW-RF model to estimate winter wheat yield.

Field validation shows that the CW-RF model has a high accuracy and could provide an accurate yield estimation at the field plot scale. The CERES-Wheat model simulation could solve the problem of few samples in the application of random forest algorithm for crop yield estimation, and could also ensure estimation accuracy. The model performed well in two typical areas of the North China Plain, Luohe (Henan) and Xiaotangshan (Beijing). Compared with the traditional winter wheat yield estimation model, the CW-RF model owns a more general applicability. More UAV flights and ground measurement experiments will be conducted in other locations to confirm the applicability of the model to the North China Plain.

The prior knowledge and empirical methods for the LAI and LNC retrieval may dilute the transferability of the model and reduce the physical interpretability of the CW-RF model. The quantitatively remote sensed models for LAI and LNC deserve future study.

As the simulation errors and uncertainty of the CERES-Wheat model could be transferred to the CW-RF model, and this effect was positively correlated, more field measurements should be introduced into the CW-RF model to reduce the errors and uncertainties. In addition, simulation and analysis results showed that the UAV hyperspectral data could significantly improve the winter wheat LAI and LNC inversion accuracy, and further improve the accuracy of winter wheat yield estimation.

ACKNOWLEDGMENT

The authors would like to thank the editor who handled our paper and the anonymous reviewers for providing truly outstanding comments and suggestions which are crucial for improving the quality and presentation of this article.

REFERENCES

[1] A. K. Prasad, L. Chai, R. P. Singh, and M. Kafatos, "Crop yield estimation model for iowa using remote sensing and surface parameters," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 8, no. 1, pp. 26–33, Jan. 2006.

[2] L. Wang, Y. Tian, X. Yao, Y. Zhu, and W. Cao, "Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images," *Field Crops Res.*, vol. 164, pp. 178–188, Aug. 2014.

[3] N. D. Mueller, J. S. Gerber, M. Johnston, D. K. Ray, N. Ramankutty, and J. A. Foley, "Closing yield gaps through nutrient and water management," *Nature*, vol. 490, pp. 254–257, Aug. 2012.

[4] D. T. Meshesha and M. Abeje, "Developing crop yield forecasting models for four major Ethiopian agricultural commodities," *Remote Sens. Appl. Soc. Environ.*, vol. 11, pp. 83–93, Aug. 2018.

[5] M. Battude *et al.*, "Estimating maize biomass and yield over large areas using high spatial and temporal resolution sentinel-2 like remote sensing data," *Remote Sens. Environ.*, vol. 184, pp. 668–681, Oct. 2016.

[6] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111599.

[7] J. Huang, H. Wang, Q. Dai, and D. Han, "Analysis of NDVI data for crop identification and yield estimation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 11, pp. 4374–4384, Nov. 2014.

[8] X. Zhou *et al.*, "Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 246–255, Aug. 2017.

[9] S. S. Panda, D. P. Ames, and S. Panigrahi, "Application of vegetation indices for agricultural crop yield prediction using neural network techniques," *Remote Sens.*, vol. 2, no. 3, pp. 673–696, Mar. 2010.

[10] A. D. Baez-Gonzalez *et al.*, "Large-area maize yield forecasting using leaf area index based yield model," *Agronomy J.,* vol. 97, no. 2, pp. 418–425, Mar. 2005.

[11] P. C. Doraiswamy, T. R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, and J. Prueger, "Application of MODIS derived parameters for regional crop yield assessment," *Remote Sens. Environ.*, vol. 97, no. 2, pp. 192–202, Jul. 2005.

[12] J. G. P. W. Clevers, "A simplified approach for yield prediction of sugar beet based on optical remote sensing data," *Remote Sens. Environ.*, vol. 61, no. 2, pp. 221–228, Aug. 1997.

[13] R. Shibles and D. N. Sundberg, "Relation of leaf nitrogen content and other traits with seed yield of soybean," *Plant Prod. Sci.*, vol. 1, no. 1, pp. 3–7, Jan. 1998.

[14] J. Ren, Z. Chen, H. Tang, Q. Zhou, and J. Qin, "Regional crop yield simulation based on crop growth model and remote sensing data," *Trans. Chin. Soc. Agricultural Eng.*, vol. 27, no. 8, pp. 257–264, Aug. 2011.

[15] W. Feng, Y. Zhu, Y. Tian, X. Yao, T. Guo, and W. Cao, "Model for predicting grain yield with canopy hyper-spectral remote sensing in wheat," *J. Triticeae Crops*, vol. 27, no. 6, pp. 1076–1084, 2007.

[16] R. Confalonieri, M. Acutis, G. Bellocchi, and M. Donatelli, "Multi-metric evaluation of the models WARM, cropsyst, and WOFOST for rice," *Ecol. Model.*, vol. 220, no. 11, pp. 1395–1410, Jun. 2009.

[17] Y. Zhu *et al.*, "Assimilation technique of remote sensing information and rice growth model based on particle swarm optimization," *J. Remote Sens.*, vol. 14, no. 6, pp. 1226–1240, Nov. 2010.

[18] J. T. Ritchie, "Model for predicting evaporation from a row crop with incomplete cover," *Water Resour. Res.*, vol. 8, no. 5, pp. 1204–1213, 1972.

[19] J. M. McKinion, D. N. Baker, F. D. Whisler, and J. R. Lambert, "Application of the GOSSYM/COMAX system to cotton crop management," *Agrcultutal Syst.*, vol. 31, no. 1, pp. 55–65, 1989.

[20] H. V. Keulen, F. W. T. Penning De Vries, and E. M. Drees, "A summary model for crop growth," in *Simulation of Plant Growth and Crop Production,* Wageningen, The Netherlands*:* Pudoc, 1982, pp. 87–97.

[21] L. Z. Gao *et al.*, "Wheat cultivational simulation-optimization-decision making system (WCSODS)," *Jiangsu J. Agricultural Sci.*, vol. 16, no. 2, pp. 65–72, 000.

[22] Y. Yan, Q. Liu, Q. Liu, J. Li, and L. Chen, "Methodolagy of winter wheat yield prediction based on assimilation of remote sensing data with crop growth model," *J. Remote Sens.*, vol. 10, no. 5, pp. 804–811, Sep. 2006.

[23] T. B. Pathak, J. W. Jones, C. W. Fraisse, D. Wright, and G. Hoogenboom, "Uncertainty analysis and parameter estimation for the CSM-CROPGRO-Cotton model," *Agronomy J.*, vol. 104, no. 5, pp. 1363–1373, Sep. 2012.

[24] J. Huang *et al.*, "Assimilation of remote sensing into crop growth models: Current status and perspectives," *Agricultural Forest Meteorol.*, vol. 276, Oct. 2019, Art. no. 107609.

[25] J. Huang *et al.*, "Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 4060–4071, Aug. 2015.

[26] Y. Xie, P. Wang, H. Sun, S. Zhang, and L. Li, "Assimilation of leaf area index and surface soil moisture with the CERES-wheat model for winter wheat yield estimation using a particle filter algorithm," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1303–1316, Apr. 2017.

[27] H. Li, Z. Chen, G. Liu, Z. Jiang, and C. Huang, "Improving winter wheat yield estimation from the CERES-wheat model to assimilate leaf area index with different assimilation methods and spatio-temporal scales," *Remote Sens,* vol. 9, no. 3, Mar. 2017.

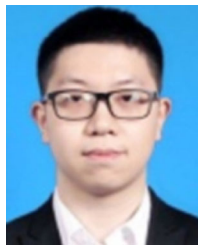
[28] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 92, pp. 79–97, Jun. 2014.

[29] C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: A review," *Precis. Agricultral*, vol. 13, no. 6, pp. 693–712, Jul. 2012.

[30] J. Geipel, J. Link, and W. Claupein, "Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system," *Remote Sens.*, vol. 6, no. 11, pp. 10335–10355, Nov. 2014.

[31] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4559–4565.

[32] H. Jiang *et al.*, "A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US corn belt at the county level," *Glob. Change Biol.*, vol. 26, no. 3, pp. 1754–1766, Mar. 2020.
[33] I. H. Witten and E. Frank, "Data mining: Practical machine learning tools and techniques with java implementations," *ACM Sigmod Rec.*, vol. 31, no. 1, pp. 76–77, Mar. 2002.
[34] Y. Everingham, J. Sexton, D. Skocaj, and G. Inman-Bamber, "Accurate prediction of sugarcane yield using a random forest algorithm," *Agronomy Sustain. Develop.*, vol. 36, no. 2, Jun. 2016.
[35] P. X. Wang, X. Qi, and L. Li, "Estimation of corn yield based on random forest regression," *Trans. Chin. Soc. Agricultural Mach.*, vol. 50, no. 7, pp. 237–245, 2019.
[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[37] S. D. Pozo, P. Rodríguez-Gonzálvez, D. Hernández-López, and B. Felipe-García, "Vicarious radiometric calibration of a multispectral camera on board an unmanned aerial system," *Remote Sens.*, vol. 6, no. 3, pp. 1918–1937, Mar. 2014.
[38] F. Iqbal, A. Lucieer, and K. Barry, "Simplified radiometric calibration for UAS-mounted multispectral sensor," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 301–313, Jan. 2018.
[39] J. T. Ritchie, "Description and performance of CERES wheat: A user-oriented wheat yield model," *ARS Wheat Yield Project*, vol. 38, pp. 159–175, 1985.
[40] A. Angstrom, "Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation," *Quart. J. Roy. Meteorolog. Soc.*, vol. 50, no. 210, pp. 121–126, Apr. 1924.
[41] W. Xiong, D. Conway, I. Holman, and E. Lin, "Evaluation of CERES-wheat simulation of wheat production in china," *Agronomy J.*, vol. 100, no. 6, pp. 1720–1728, Nov. 2008.
[42] H. Zhao *et al.*, "Risk assessment of agricultural drought using the CERES-wheat model: A case study of Henan plain, China," *Clim. Res.*, vol. 50, no. 2/3, pp. 247–256, Dec. 2011.
[43] N. Yao *et al.*, "Parameter estimation and verification of DSSAT-CERES-Wheat model for simulation of growth and development of winter wheat under water stresses at different growth stages," *Trans. Chin. Soc. Agricultural Eng.*, vol. 31, no. 12, pp. 138–150, Jun. 2015.
[44] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: A classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comp. Sci.*, vol. 43, no. 6, pp. 1947–1958, Nov. 2003.
[45] W. Fan, X. Xu, X. Liu, B. Yan, and Y. Cui, "Accurate LAI retrieval method based on PROBA/CHRIS data," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 8, pp. 1499–1507, Aug. 2010.
[46] L. He *et al.*, "Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data," *Remote Sens. Environ.*, vol. 174, pp. 122–133, Mar. 2016.
[47] Y. Curnel, A. J. de Wit, G. Duveiller, and P. Defourny, "Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS experiment," *Agricultural Forest Meteorol.*, vol. 151, no. 12, pp. 1843–1855, Dec. 2011.
[48] H. Li *et al.*, "Assimilation of temporal-spatial leaf area index into the CERES-Wheat model with ensemble Kalman filter and uncertainty assessment for improving winter wheat yield estimation," *J. Integr. Agriculture*, vol. 16, no. 10, pp. 2283–2299, Oct. 2017.
[49] L. Chen, J. Huang, and J. Xu, "Uncertainty analysis of rice yield estimation by remote sensing," in *China Assoc. Remote Sens. Appl.*, Basel, Switzerland: Multidisciplinary Digit. Publ. Inst., 2007, pp. 93–101.
[50] W. Xiong, "The performance of CERES-wheat model in wheat planting areas and its uncertainties," *J. Appl. Meteorol. Sci.*, vol. 20, no. 1, pp. 88–94, 2009.
[51] J. Huang, L. Chen, and X. Wang, "Sensitivity of rice growth model parameters and their uncertainties in yield estimation using remote sensing date," *Trans. Chin. Soc. Agricultural*, vol. 28, no. 19, pp. 119–129, 2012.
[52] H. Li, Z. Jiang, Z. Chen, J. Ren, B. Liu, and Hasituya, "Assimilation of temporal-spatial leaf area index into the CERES-Wheat model with ensemble Kalman filter and uncertainty assessment for improving winter wheat yield estimation," *J. Integr. Agriculture*, vol. 16, no. 10, pp. 2283–2299, 2017.
[53] S. H. Qader, J. Dash, and P. M. Atkinson, "Forecasting wheat and barley crop production in arid and semi-arid regions using remotely sensed primary productivity and crop phenology: A case study in Iraq," *Sci. Total Environ.*, vol. 613, pp. 250–262, Feb. 2018.
[54] J. Yue *et al.*, "Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models," *Remote Sens.*, vol. 9, no. 7, Jul. 2017.
[55] L. Serrano, I. Filella, and J. Penuelas, "Remote sensing of biomass and yield of winter wheat under different nitrogen supplies," *Crop Sci.*, vol. 40, no. 3, pp. 723–731, May. 2000.
[56] J. Ren, Z. Chen, Q. Zhou, and H. Tang, "LAI-based regional winter wheat yield estimation by remote sensing," *Chin. J. Appl. Ecol.*, vol. 21, no. 11, pp. 2883–2888, Nov. 2010.
[57] M. Reyniers, E. Vrindts, and J. D. Baerdemaeker, "Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat," *Eur. J. Agronomy*, vol. 24, no. 2, pp. 87–94, Feb. 2006.
[58] K. Liu, Q. Zhou, W. Wu, Z. Chen, and H. Tang, "Comparison between multispectral and hyperspectral remote sensing for LAI estimation," *Trans. Chin. Soc. Agricultural Eng.*, vol. 32, no. 3, pp. 155–162, Feb. 2016.
[59] W. Feng, X. Yao, Y. Zhu, Y. Tian, and W. Cao, "Monitoring leaf nitrogen status with hyperspectral reflectance in wheat," *Eur. J. Agronomy*, vol. 28, no. 3, pp. 394–404, Apr. 2008.
[60] B. Luo, C. Yang, J. Chanussot, and L. Zhang, "Crop yield estimation based on unsupervised linear unmixing of multidate hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 162–173, Jun. 2013.
[61] W. Fan, B. Yan, and X. Xu, "Crop area and leaf area index simultaneous retrieval based on spatial scaling transformation," *Sci. China-Earth Sci.*, vol. 53, no. 11, pp. 1709–1716, Nov. 2010.
[62] R. Pu, P. Gong, and Q. Yu, "Comparative analysis of EO-1 ALI and hyperion, and landsat ETM+ data for mapping forest crown closure and leaf area index," *Sensors*, vol. 8, no. 6, pp. 3744–3766, Jun. 2008.
[63] N. H. Broge and 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.*, vol. 76, no. 2, pp. 156–172, May. 2001.
[64] H. Zhao, X. Song, G. Yang, Z. Li, D. Zhang, and H. Feng, "Monitoring of nitrogen and grain protein content in winter wheat based on sentinel-2A data," *Remote Sens.*, vol. 11, no. 14, Jan. 2019.



Siqi Yang received the B.S. degree in geographical information science from Beijing Normal University, Beijing, China, in 2018. He is currently working toward the M.S. degree in photogrammetry and remote sensing at the Institute of remote sensing and GIS, Peking University, Beijing.

His research interests include quantitative remote sensing of vegetation, hyperspectral remote sensing, and machine learning.

Ling Hu received the B.S. degree in geographical information science from Beijing Normal University, Beijing, China, in 2017. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing at the Institute of remote sensing and GIS, Peking University, Beijing.

Her research interests include quantitative remote sensing of vegetation, and application of remote sensing in ecology and environment management.

Haobo Wu received the B.S. degree in geographical information science in 2018 from Peking University, Beijing, China, where he is currently working toward the M.S. degree in photogrammetry and remote sensing at the Institute of remote sensing and GIS.

His research interests include quantitative remote sensing, hyperspectral remote sensing, and high-resolution remote sensing.

Huazhong Ren received the Ph.D. degree from Beijing Normal University, China and from Université de Strasbourg, Strasbourg, France in 2013.

Since 2016, he has been an Assistant Professor with the Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing, China. His research interests focus on land surface temperature/emissivity estimate from thermal infrared remote sensing data.

Hongbo Qiao received the B.S. degrees in plant protection from the Henan Agricultural University, Zhengzhou, China, in 2001 and the M.S. and Ph.D. degrees from the China Academy of Agricultural Science, Beijing, China, in 2004 and 2007, respectively.

From 2007 to 2011, he was a Research Assistant with Henan Agricultural University. From 2012 to 2018, he was an Assistant Professor with Henan Agricultural University. Since 2019, he has been a Professor with the Department of Computer Science and Technology, Henan Agricultural University. He is the author of three books, more than 50 articles. His research interests include hyperspectral remote sensing in agriculture.

Peijun Li received the Ph.D. degree in geology from the Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, in 1995.

He is currently a Professor with Peking University, Beijing. He was a Visiting Researcher/Professor with several institutions, including Seoul National University, Seoul, South Korea; University of Bonn, Bonn, Germany; and Columbia University, New York, NY, USA. His research interests include land cover classification and change detection, very high-resolution image analysis, urban remote sensing analysis, and geologic remote sensing.

Dr. Li was a Co-Chair of ISPRS Working Group VII/4 "Methods for landcover classification" from 2008 to 2016. He is currently a Member of Editorial Advisory Board, ISPRS Journal of Photogrammetry and Remote Sensing.

Wenjie Fan received the Ph.D. degree in physical geography from the College of Urban and Environmental Sciences, Peking University, Beijing, China, in 2000.

He was a Postdoctorate Research with the Institute of Remote Sensing and GIS, Peking University, in 2002. She is currently a Professor with the IRSGIS, PKU. She has continued to work on and manage research projects on vegetation remote sensing, scale effects of remote sensing, hyperspectral remote sensing, and application of remote sensing in ecology and environment management for 20 years. She has authored or coauthored more than 90 papers (more than 40 papers indexed by SCI). She has also been the PI of more than seven National Natural funds and other research projects.