Remote Prediction of Oilseed Rape Yield via Gaofen-1 Images and a Crop Model

Wenchao Tang 1, Rongxin Tang 1,2,3*, Tao Guo 2 and Jingbo Wei 1,*

1 Institute of Space Science and Technology, Nanchang University, Nanchang 330031, China; tangwc@ncu.edu.cn (W.T.); rongxint@ncu.edu.cn (R.T.)
2 School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China; 401030919007@email.ncu.edu.cn
3 Jiangxi Provincial Key Laboratory of Interdisciplinary Science, Nanchang University, Nanchang 330031, China
* Correspondence: weijingbo@ncu.edu.cn

Abstract: The fast and accurate prediction of crop yield at the regional scale is of great significance to food policies or trade. In this study, a new model is developed to predict the yield of oilseed rape from high-resolution remote sensing images. In order to derive this model, the ground experiment and remote sensing data analysis are carried out successively. In the ground experiment, the leaf area index (LAI) of four growing stages are measured, and a regression model is established to predict yield from ground LAI. In the remote sensing analysis, a new model is built to predict ground LAI from Gaofen-1 images where the simple ratio vegetation index at the bolting stage and the VARIgreen vegetation index at the flowering stage are used. The WOFOST World Food Study (WOFOST) crop model is used to generate time-series ground LAI from discontinuous ground LAI, which is calibrated coarsely with the MODerate resolution imaging spectroradiometer LAI product and finely with the ground-measured data. By combining the two conclusive formulas, an estimation model is built from Gaofen-1 images to the yield of oilseed rape. The validation shows that the prediction error of the proposed algorithm is less than 5.5%, which highlights the feasibility of our method for accurate prediction of the oilseed rape yield in a large area.

Keywords: oilseed rape; yield; WOFOST; LAI; Gaofen-1

1. Introduction

Crop yield is one of the most important pieces of economic information for a country or a region; it is related to food security and is of great significance for agricultural planning and layout, policy adjustment, and food trade [1]. At present, the accurate prediction of crop yield is one of the hot topics in agricultural research and applications. Traditional methods of crop yield estimation mainly include agronomic estimation, statistical estimation, and meteorological statistical estimation, which are applicable to small areas. The development of remote sensing technology provides a fast and accurate method for achieving large scale crop yield estimation, which can obtain the non-destructive measurements of cropping systems at various scales [2,3]. Remote estimation of crop yield has been widely used in agricultural production because of its low cost and high efficiency compared with traditional field surveys.

Generally, two types of approaches are used for remote estimation of crop yield. One type is to establish empirical relationships between crop yield and remote sensing data, such as the relationship between various vegetation indices and yield [4–6]. Siyal et al. [7] developed a model for rice yield estimation by analyzing the correlation between Landsat-7 enhanced thematic mapper plus (ETM+) bands and crop yield. Sakamoto et al. [8] used
time-series MODerate resolution Imaging Spectroradiometer (MODIS) data to construct the wide dynamic range vegetation index (WDRVI) to carry out near real-time prediction of United States corn yields. Gong et al. found that the normalized difference vegetation index (NDVI) could accurately estimate the yield of oilseed rape with an estimation error of less than 13% through unmanned aerial vehicle (UAV) remote sensing data [9]. This approach mainly relies on remote sensing data, the model construction is simple and easy to be applied, but this approach can only be applied to the modeled area, and it is difficult to be extended to other areas or crops, so there are certain limitations. In addition, the algorithm was constructed by statistical analysis, thus lacking the physiological background [10].

The other type attempts to simulate crop yield through crop growth models using remote sensing data as input parameters [11,12]. Crop growth model is an imperfect representation of the real crop growth process by describing the crop growth and development process through a mathematical modeling approach based on meteorology, soil, crop variety characteristics, and crop management practices [13,14]. There are many types and versions of crop growth models due to different design purposes, regional environments, and crop types. According to the different driving factors of the model, the models can be classified into three types: soil factor driven, atmospheric factor driven, and light energy factor driven. AquaCrop [15], soil–water–atmosphere–plant (SWAP), and Agricultural Production Systems sIMulator (APSIM) [16] are typical soil factor driven models. The Simple and Universal CROp growth Simulator (SUROS) and WOrld FOod STudy model (WOFOST) are driven by atmospheric factors. The Carnegie Ames Stanford Approach (CASA) [17] and the Simple Algorithm For Yield estimate (SAFY) model [18] are driven by sunshine factors. This type of approach is based on agronomy models that give good descriptions of crop development thus providing a substantial theoretical framework for the remote estimation of crop yield [19]. However, it usually requires a large number of input parameters, among which some are difficult to obtain by remote sensing techniques and some are complicated or expensive to measure. This yield estimation method provides a reasonable theoretical framework for remote estimation of crop yield by describing the crop growth and development process through a mathematical modeling approach based on the meteorology, soil, crop variety characteristics, and crop management measures [19]. However, it usually requires a large number of input parameters, some of which are difficult to be obtained with remote sensing techniques and some are time-consuming and laborious to measure [20].

In the above mentioned methods, remote sensing data are usually linked to some greenness-related biophysical parameters in the crop, such as chlorophyll content, leaf area index, crop biomass and net primary productivity (NPP) to directly or indirectly reflect crop growth and photosynthetic capacity and, thus, characterize crop yield [21]. In the growth of most crops, leaves are the major components to absorb light for photosynthesis and assimilation [22], so their condition and vigor can directly determine the potential yield [23]. Leaf area index (LAI) is widely used in crop yield estimation as an indicator of crop leaf’s condition and vigor, and can be calculated accurately and easily from remote sensing data [24,25]. For example, Baez-Gonzalez et al. constructed a model for estimating maize yield by using the leaf area index for the growth period from fall to winter, the results showed that the method could achieve good accuracy in large scale crop yield estimation [26]. Moriondo et al. combined NDVI calculated from satellite data with NDVI simulated by CROPSYST model to predict wheat yield, and the simulated values were significantly correlated with the measured values [27]. Dente et al. assimilated the leaf area index simulated by CERES-wheat model with the leaf area index estimated by remote sensing to improve the prediction of regional wheat yields with good experimental results [28].

Oilseed rape is an important agricultural product in many countries, and the yield is estimated with remote sensing. For example, Sulik and Long [29] used the normalized difference yellowness index (NDYI) to predict the yield of oilseed rape, which is better than NDVI during mid-season development stages, especially peak flowering. Zang et al. [30] proposed the enhanced area yellowness index (EAYI) for flowering stage prediction. How-
ever, in these works, the MODIS images are used, and the accuracy is hardly satisfied due to the 500-m data source.

Affected by long temporal resolution and adverse weather conditions, high-resolution remote sensing data are difficult to be observed continuously throughout the growth period, and therefore do not play a role in predicting oilseed rape yield. Data assimilation integrates remote sensing information into the crop growth model, which can reduce the uncertainty of the regionalization of crop growth model parameters [31], thereby it is of potential to high-precision crop growth monitoring and yield simulation at the regional scale. Therefore, a localized growth model is considered to provide critical growth parameters for estimating oilseed rape yield at a large scale.

In this study, we focused on the research of yield estimation in oilseed rape (Brassica napus). Oilseed rape is one of the major cash crops grown mainly in temperate regions and its byproducts can be used for food, biofuel, and medicine [32]. Oilseed rape is a member of the family Brassicaceae, which has distinct developmental stages (i.e., seedling, bolting, flowering, and pod formation) that have different spectra, especially during the flowering stage when the entire canopy has bright yellow flowers [33]. This growth and developmental characteristic increases uncertainties of yield prediction by spectral indices. Since the seed yield largely depends on the number of flowers that translate into pods, flowering is usually considered as an important factor to predict yield [34,35]. Therefore, the timing of acquiring remote sensing data for yield prediction in oilseed rape is important to achieve high accuracy.

To explore the crop yield estimation on a large scale with limited ground observation data, the following contributions are made in our work.

1. An effective yield estimation method was developed based on remote sensing images and a crop growth model.
2. The WOFOST crop growth model was localized for oilseed rape in which the MODIS data, meteorological data, and measured data are involved.
3. A method is proposed to estimate the yield of oilseed rape using Gaofen-1 images.

The remainder of the work is organized as follows. Section 2 introduces the study area, data, research framework, and critical models for LAI assimilation and plant extraction. Section 3 presents the experimental results of ground observation and continuous ground LAI fusion, and the yield prediction model is regressively built with vegetation indices from Gaofen-1 satellite images. The proposed model is verified in Section 4 where the performance is discussed, too. The conclusions are given in Section 5.

2. Materials and Methods

2.1. Study Area

Oilseed rape is an important economic crop in southern China. In the middle and lower reaches of the Yangtze River in China, winter wheat and rapeseed have the same phenological period. Before maturity, the growth of oilseed rape experiences four stages, namely seedling, bolting, flowering, and bolting stages. The seedling stage is throughout the whole of December and January. The bolting stage is from the beginning to the middle of February. The flowering stage is from the end of February to the end of March. The pod stage is from the beginning to the middle of April. The pods will become mature after about twenty days.

Hubei province (Figure 1) is one of the largest oilseed rape production in China by virtue of its suitable climatic conditions. Hubei province is located in central China and belongs to the Yangtze River basin, with longitudes ranging from 108°21′E to 116°07′E and 29°05′N to 33°20′N in latitude. The climate of the region is subtropical and controlled mainly by the East Asian monsoon. The overwintering of oilseed rape seedlings is unhindered, and the growing stage is characterized by abundant rainfall and sunshine. According to the statistical data from the China statistical yearbooks, Hubei province contributes one-sixth of the oilseed rape planting areas and products of China.
2.2. Data Collections

Two groups of data are used in this research. The first group is the field data collected by the ground experiment. The second group is the remote sensing data from Gaofen-1 satellites.

2.2.1. Field Data

The ground base for this study is located at the crop experiment and research base (30°6′43″N, 115°35′22″E), Central China Agricultural University, Wuxue city. The growing season of oilseed rape in our study was from September 2014 to May 2015. The data from an oilseed rape plot were collected for the full growing stages. This plot was specifically designed for oilseed rape research and divided into 48 subplots with similar sizes (approximately 15 m × 2 m). In 2015, these subplots were planted with hybrid oilseed rape (Huayouza 9 [36]), but with different fertilization programs and planting methods. Eight levels of nitrogen fertilizer (0, 45, 90, 135, 180, 225, 270, and 360 kg/ha) were applied. Additionally, each level was randomly replicated on six subplots (see Figure 1d for details).

The ground LAI data were determined by the LAI-2200C Plant Canopy Analyzer [37], which is widely used for LAI measurements in field. The LAI-2200C is equipped with a fisheye optical sensor consisting of five concentric rings centered at zenith angles of 7°, 22°, 38°, 52°, and 68° to measure the canopy above and below radiation to estimate the canopy light
interception and transmittance at the five angles; ultimately LAI can be calculated from inversion of the Beer–Lambert law [38].

In our experiment, the ground LAI was measured in 48 plots of oilseed rape plots at four key developmental stages, with each plot measured in six aliquots and the average value was used as the LAI of the plot. The harvest date of the experimental field oilseed rape plots was 5 May 2015. In each plot, half of the above-ground plant materials (approximately 15 m²) were cut for measuring the final yield. The harvested plant materials were exposed to sunlight for ten days for seed threshed. The seeds were then cleaned and put into an oven at 60 °C until there was no change in their weight (about 4 days). Finally, all seeds were weighed together and the yield of each plot was the ratio of total weight to ground area (kg/ha). The measured yield will be used to train the yield estimation model.

2.2.2. Satellite and Meteorological Data

Spectral vegetation indices derived from satellite observations are highly correlated to field-measured LAI, so satellite images and data were involved for model building. To calculate the vegetable indices and extract crop regions for large areas, satellite images were used. To estimate the continuous field LAI, remotely sensed LAI were used.

The Gaofen-1 wide-field-view (WFV) satellite images were chosen as the satellite source. Gaofen-1 was launched by China in April 2013 equipped with the multispectral WFV sensor spanning the blue, green, red, and near-infrared (NIR) bands. WFV has a 16-meter (16 m) ground resolution and a 4-day revisiting period suitable for agricultural investigation for large areas. In our study, twenty-two WFV images covering various locations of Hubei province were used to build the algorithm, and ten images for Wuxue City over the years were used for validation. Details of the applications, acquisition dates, coverage areas, and phenology stages for these images are listed in Table 1.

Table 1. The used Gaofen-1 WFV images to build and verify the method.

| Purpose   | Coverage           | Acquisition Date                      | Phenology Stage |
|-----------|--------------------|---------------------------------------|-----------------|
|           | Jingzhou City      | 7 January 2015, 21 Match 2015          | Bolting Stage   |
|           |                    | 21 January 2017, 1 April 2017          | Flowering Stage |
|           | Macheng City       | 21 January 2017, 8 Match 2017          | Bolting Stage   |
|           |                    |                                       | Flowering Stage |
|           | Tianmen City       | 22 January 2017, 8 Match 2017          | Bolting Stage   |
|           |                    | 14 January 2018, 12 Match 2018         | Flowering Stage |
|           | Zhongxiang City    | 18 January 2015, 25 Match 2015         | Bolting Stage   |
|           |                    |                                       | Flowering Stage |
|           | Jiayu County       | 22 January 2015, 25 Match 2015         | Bolting Stage   |
|           |                    | 14 January 2018, 28 Match 2018         | Flowering Stage |
|           | Jianli County      | 22 January 2017, 8 Match 2017          | Bolting Stage   |
|           |                    | 14 January 2018, 12 Match 2018         | Flowering Stage |
|           | Yangxin County     | 3 February 2018, 28 Match 2018         | Bolting Stage   |
|           | Wuxue City         | January of 2014, 2015, 2017, 2018, 2019| Bolting Stage   |
|           |                    | March of 2014, 2015, 2017, 2018, 2019  | Flowering Stage |
The Gaofen-1 WFV images were downloaded from the China Center for Resources Satellite Date and Application (CRESDA). The preprocessing of these images mainly included geometric correction, radiometric calibration, and atmospheric correction. With the assistance of ASTER GDEM v2 data, geometric correction was performed according to the method introduced in [39]. Radiometric calibration converts digital numbers to absolute at-sensor radiance, which is performed using the software Environment for Visualizing Images (ENVI, Harris Geospatial Solutions, Inc., Broomfield, CO, USA), where the calibration parameters were from CRESDA. Atmospheric correction was performed using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) model in ENVI, where the cloud cover on the images were masked by visual interpretation before the correction. The FLAASH parameters used for the correction were obtained based on the image acquisition time and imaging conditions.

The satellite LAI data for the growth model were from the MODIS LAI product MCD15A3H (accessed on 6 June 2020 from https://ladsweb.modaps.eosdis.nasa.gov) for the corresponding years with a temporal resolution of 4 days and a spatial resolution of 500 m. Projection transformation, mosaicking, and cropping were performed on the images. A time series MODIS image dataset covering nine meteorological stations during the oilseed rape growing stages from October to May of the next year was constructed.

Ancillary meteorological data are needed to predict continuous LAI. The phenology of oilseed rape was obtained from the meteorological stations in Hubei province (green dots in Figure 1a). The daily meteorological data were from the China weather data network (accessed on 30 August 2020 from http://data.cma.cn). The China statistical yearbooks on oilseed rape acreage and yield were downloaded from the National Statistical Bureau of China and the Hubei Statistical Bureau.

2.3. Research Framework

The previous work [40] shows that the ground LAI of oilseed rape is highly related to the yield without the help of vegetation indices, which is followed in this work. The regression formula from LAI to yield can be established by ground experiments. To this end, we carried out a continuous observation and data collection for the whole growth cycle of rapeseed, including hyperspectral data, UAV images, leaf area index, leaf chlorophyll, leaf water, nitrogen, phosphorus, and potassium in plants, the final yield of each rape plot, and so on. Among them, the leaf area index and final yield data were used in the ground experiment to build the estimation model.

The strong correlation between vegetable index and LAI can be used to predict ground LAI from remote sensing images, which has been demonstrated in many literature studies. To construct the model, accurate ground LAI is needed over a large area to match the resolution of satellite images. In other words, the ground LAI should reflect the mean value within the region of each satellite pixel, which calls for a large number of measurements within a satellite pixel location. On the other hand, it is costly to measure in-situ LAI in large areas. In our ground experiments, we only obtained 11 groups of point LAI values, which could not be used to build the model from remote sensing images to ground LAI. So we had to turn to other solutions to find the ground LAI.

MODIS provides the LAI product. Unfortunately, previous work [41] shows that MODIS LAI tend to be underestimated than measured during the growing season of crops. This is due to the 500 m ground resolution of the MODIS image being too low to account for the pure pixels of crops. Remotely-sensed LAI must be corrected, approaching to ground-measured LAI.

The MODIS LAI and in-situ measured LAI can be fused to obtain the ground LAI in large regions. The crop model describes the growth and potential regarding nutrient and meteorological conditions. One of the crop models, WOFOST, will be used to integrate the two data sources to output the ground LAI. WOFOST has to be localized to the study area before it is used. Two groups of data, the MODIS LAI product and the ground-measured LAI, are used to tune the parameters to adapt. MODIS LAI is first used to coarsely tune the
WOFOST model, and in-situ measured LAI is used to fine-tune the WOFOST model along with the meteorological information to obtain an accurate model for ground LAI of oilseed rape fields. It can also be understood that WOFOST is used to improve MODIS LAI, which uses measured data to convert satellite LAI to ground LAI.

In order to compensate for insufficient LAI data quantity and MODIS LAI deviation in large areas, we first used MOIDS LAI data to calibrate the WOFOST model in a coarse way, and then used experimental field LAI data to conduct a second calibration for finer model. After the WOFOST model is localized and specialized for oilseed rape, we could obtain a more accurate LAI growth curve of rapeseed near meteorological stations, which could be used to build the model between LAI and vegetable indices.

Following the ideas, the workflow of this study is summarized as eight consequent steps and demonstrated in Figure 2.

**Figure 2.** Workflow for the proposed method.

1. The ground experiment is carried out to monitor the LAI changes at four stages with handheld equipment. The final yield is also measured.
2. The regression model between LAI and crop yield is built with the ground collected data.
3. Data are prepared for LAI prediction. Data from remote sensing satellites are collected and preprocessed with radiometric, geometric, and atmospheric corrections. Meteorological data and statistical yearbooks are downloaded.
4. The WOFOST model is calibrated with MODIS LAI and ground collected data for localization.
5. Time-series LAI are predicted with localized WOFOST, and are linked to vegetable indices from satellite images.
6. Separate models are combined to give the final prediction model from satellite images to the yield of oilseed rape.
7. The planting areas of oilseed rape are extracted using the pyramidal bottleneck residual network.
8. A field investigation is performed to validate the remote sensing estimated yield for oilseed rape.

### 2.4. Leaf Area Index Assimilated with the WOFOST Model

The LAI from ground measurements is collected aperiodically, but dense ground LAI is expected to ensure the availability at the moments of satellites images. To simulate the ground LAI in a continuous way, the WOFOST model [42] is adopted. WOFOST is a dynamic and explanatory growth model developed by the Centre for World Food Studies (CWFS) and Wageningen University, Netherlands, aiming at the description of basic physiological processes, such as photosynthesis, respiration, transpiration, and dry matter partitioning, as well as how these processes are affected by the environment. It has
been widely used in yield forecasting [43], climate change [44], and remote sensing data assimilation [45].

WOFOST can simulate crop LAI directly. The calculation about LAI is based on the contribution of different branches (leaves, stems, and fruits). Then it is consistent to the changing characteristics of the canopy during different growth periods of oilseed rape.

WOFOST has to be localized before it is used. Two groups of data, the MODIS LAI product and the ground-measured LAI, are used to tune the parameters to adapt to the study area. MODIS LAI is used to tune the WOFOST model to coarse localization, and ground LAI is used to fine-tune the WOFOST model to obtain an accurate model. WOFOST is used to improve MODIS LAI, which uses measured data to convert satellite LAI to ground LAI.

Crop models usually have many parameters, some of which are difficult to be obtained or calibrated. On the contrary, WOFOST is a general-purpose model that can simulate different kinds of crops by customizing different parameters. The main parameters of the model have been fully calibrated under different meteorological, soil, and management conditions through ground tests and remote sensing data assimilation [21,46], which can be set to default values or ranges. Since only the core parameters of the model need to be calibrated, WOFOST is easy to use.

The WOFOST model is driven by day-by-day meteorological data to build a crop growth and development module based on the cumulative temperature theory, taking into account the sensitivity of the crop to light length and the growth and senescence processes of the leaves. The accumulation and distribution of dry matter in the model is distributed to various organs of the crop after respiratory consumption. The phenology of WOFOST is described by the dimensionless state variable development stage (DVS). For most annual crops, DVS is set to 0 at seedling emergence, 1 at flowering (for cereals), and 2 at maturity. Due to the large number of adjustable parameters of the WOFOST model, such as meteorology, crop, soil, and field management, it is difficult to calibrate all parameters. Therefore, the default values are set based on the model’s default values or from literature studies for those parameters that are of low sensitivity. For the parameters that are species-related or have high sensitivity and large spatial variability, the range of values is first calculated from the observed data and then determined by optimization algorithms. However, the inactiveness of some parameters is possibly linked to geographic locations and planting methods, so they have to be tuned carefully before used for various locations or crops.

Before WOFOST is calibrated, preparation has to be made on the MODIS LAI sequence and location of oilseed rape fields. The MODIS LAI data of the study area were overlaid in chronological order to generate the corresponding LAI time series curves on selected pixel locations, and the noise from clouds, water vapor, and aerosols was smoothed using the Savitzky–Golay (S–G) filtering method. The selected locations should be filled with oilseed rape fields and close to meteorological stations. To this end, the oilseed rape fields were searched within the range of 750 m near the meteorological stations to obtain valid oilseed rape LAI curves.

In the selection of the optimized parameters, TSUM1 (effective cumulative temperature from seedling to flowering) and TSUM2 (effective cumulative temperature from flowering to maturity) represent two effective cumulative temperatures, respectively, which were obtained based on the dynamic calculation of temperature data due to the difficulty obtaining exact timings of seedling, flowering, and pod stages of oilseed rape near each meteorological station in previous years [21]. In addition, TDWI (initial biomass) and SPAN (leaf senescence coefficient) are two important parameters in the WOFOST model, which are also involved in the optimization process. The ranges of critical parameters are listed in Table 2.
Table 2. Main adjustable parameters of WOFOST in predicting oilseed rape LAI.

| Parameter | Description                                      | Unit     | Minimum | Maximum |
|-----------|--------------------------------------------------|----------|---------|---------|
| SPAN      | Life span of leaves growing at 35 °C             | days     | 17      | 21      |
| TSUM1     | Temperature sum from emergence to anthesis       | °C·days  | 150     | 240     |
| TSUM2     | Temperature sum from anthesis to maturity         | °C·days  | 600     | 900     |
| TDWI      | Initial total crop dry weight                    | kg·hm⁻²  | 40      | 90      |
| CVO       | Efficiency of conversion into storage organ      | kg·kg⁻¹  | 0.60    | 0.85    |
| CVS       | Efficiency of conversion into stems              | kg·kg⁻¹  | 0.66    | 0.70    |

The unknown WOFOST parameters are coarsely calibrated with the meteorological data as input to target MODIS LAI. The global optimization algorithm SCE-UA [47,48] is used for optimization. In the calibration, the cost function is normalized with the remote sensing observed LAI and WOFOST simulated LAI as

\[
\sum_{t=1}^{N} \left( \frac{\text{LAI}_{\text{obs}}^t - \text{LAI}_{\text{sim}}^\text{min}}{\text{LAI}_{\text{obs}}^\text{max} - \text{LAI}_{\text{obs}}^\text{min}} \right)^2
\]

(1)

Here \( \text{LAI}_{\text{obs}}^\text{min} \) and \( \text{LAI}_{\text{sim}}^\text{min} \) denote the minimum LAI values observed by remote sensing or simulated with WOFOST, respectively, \( \text{LAI}_{\text{obs}}^\text{max} \) and \( \text{LAI}_{\text{sim}}^\text{max} \) denote the corresponding maximum values, \( \text{LAI}_{\text{obs}}^t \) and \( \text{LAI}_{\text{sim}}^t \) denote the LAI values at the \( t \)th day either observed or simulated, respectively, and \( N \) is the number of spanning days. The optimal parameters are obtained after the SCE-UA iteration stops when the values of the neighboring five cost functions reach the preset threshold or the number of iterations reaches the allowed times.

The WOFOST parameters obtained by rough calibration of MODIS LAI is recalibrated using the ground-measured LAI. The first four parameters in Table 2 are fine-tuned as other parameters have little effect on the model output. Fine calibration is done manually because of the small number of parameters, where only one parameter is adjusted at a time. The parameter value is set based on the rule that the WOFOST output is consistent with the trend of MODIS LAI, and the peak value approaches the peak value of the ground-measured LAI. After the second calibration, the WOFOST model is able to achieve the prediction of the growth and the development process and maturity of oilseed rape by model simulation.

2.5. Extraction of Planting Areas with Convolutional Neural Networks

To estimate crop yield with remote sensing, it is necessary to extract planting areas and identify crop types. The differences of crop spectral characteristics at different growing stages are observed by using surface information collected from remote sensing images. Among the four steps, crop identification and yield assessment are indispensable for yield estimation at a large scale. The accurate identification of crop types is the premise of crop yield estimation, and the remote sensing identification of crops is mainly based on the spectral differences and phenological differences of crops. In addition, with the development of computer image processing technology, spatial features have also become important auxiliary information for remote sensing identification of crops. According to the features utilized, the present crop identification methods can be categorized into three categories, namely spectral and spatial features-based, phenological features-based, and multi-source data-based, where neural networks [49,50] have been commonly used.

In this study, a pyramid bottleneck residual network [51] was used to identify the planted area of oilseed rape. The algorithm was optimized for multispectral image classification based on the pyramid convolutional neural network with bottleneck residual blocks proposed by Paoletti et al. [52]. In order to make full use of the high spatial resolution and rich structural and texture information of multispectral images, the approach enlarged the input image block size to \( 27 \times 27 \) to extract image features in a larger neighbourhood area. In addition, since the number of bands in multispectral images was much smaller
than that in hyperspectral images, the channel depths of convolutional layers were reduced to have a smaller number of convolutional kernels. The detailed network architecture and parameters of the classifier used in this work are shown in Figure 3. Before the classifier runs, some pixels have to be labeled manually to help train the network. The input image patch has the size of 27 × 27, with the label outputted marking the class to which the central pixel belongs. The parameters are optimized using the Adam optimizer with default parameter values. The training iterates 200 epochs. The learning rate is 0.0003 for the first 100 epochs and 0.00015 for the remaining epochs.

![Figure 3. Different residual cells and deep bottleneck pyramid neural network used for ground classification to exact the oilseed rape cultivation area.](image)

(a) Traditional residual unit. (b) Bottleneck residual unit. (c) Pyramid residual unit. (d) Pyramid residual network architecture. The kernel size, depth, and stride are listed in turn for each convolutional layer. BN denotes the batch normalization. Relu denotes the rectified linear unit for activation. FC denotes the fully connected layer, and “conv” denotes the convolutional layer.

3. Results

3.1. Yield Evaluation by Ground LAI at Different Developmental Stages

The analysis of LAI measurements and the final yield data at four different developmental stages of oilseed rape was carried out, and the results are shown in Table 3. From Table 3, one can see that oilseed rape LAI at the bolting and pod formation stages are closely related to the yield with regard to the determination coefficient ($R^2$) above 0.7, and are of much lower correlations between the yield and LAI at the seedling and flowering stages. The highest correlation between LAI and yield was found at the pod formation stage with $R^2$ reaching around 0.84. At this stage, the green pods contribute a lot to the LAI measurement, and they determine the yield directly. However, the time for modeling using data collected at the pod formation stage for yield estimation is of weak meaningfulness because it is too close to harvest, such that no time is ready for effective yield enhancement measures.
Table 3. Determination coefficients ($R^2$) and standard error(s) (SE) of yield estimation using leaf area index (LAI) measured at four developmental stages in oilseed rape.

| Metric                  | Seedling | Bolting | Flowering | Pod       | Bolting + Flowering |
|-------------------------|----------|---------|-----------|-----------|---------------------|
| Determination coefficients ($R^2$) | 0.21     | 0.70    | 0.59      | 0.84      | 0.72                |
| Standard errors (SE, kg/ha) | 764      | 475     | 561       | 352       | 424                 |

Oilseed rape LAI at the seedling stage is weakly related to yield ($R^2 = 0.21$). At this stage, oilseed rape is in early vegetative period and there is much uncertainty about its future growth due to various external environmental stresses and field management.

A significant relationship can be found between the rape yield and LAI at the bolting and flowering stages. At such stages, rape was near or during its early reproductive period, having almost the maximal plant greenness and photosynthetic capacity, which is indicative of the crop’s potential yield [53]. At the same time, the correlation between LAI and yield is much higher at the bolting stage than that at the flowering stage ($R^2$ is 0.70 and 0.59, respectively). This is mainly because the leaves are fully developed at the bolting stage and the LAI remained fairly stable during this period. Many researcher have found that leaves of oilseed rape greatly contributed to seed development (e.g., pod number and seed weight) during ripening phase [54]. When oilseed rape steps into the flowering stage, conspicuous yellow flowers appear throughout the canopy. During this stage, the LAI measurement is confounded by flowers and leaves, which demonstrates significant optical differences. Luckily, the degree of flower cover at the flowering stage is directly related to the number of pods formed and has potential information on final yield in the floral LAI information. In order to predict the accurate yield in advance, we used a stepwise regression method to construct the LAI model at the bolting and flowering stages:

\[
\text{Yield} = 581.82 \times \text{LAI}_{\text{Bolting Stage}} + 91.47 \times \text{LAI}_{\text{Flowering Stage}} - 63.10
\]  

Since there is a high correlation between ground LAI and yield, the problem now is to accurately obtain the ground LAI at a large regional scale. Because of the advantage for large areas, remote sensing observation is used to establish the estimation model from images to LAI, which will be introduced in the next section.

3.2. Assimilated LAI Sequence

Figure 4 shows the LAI assimilation results of Jingzhou City from 2014 to 2015.

![Figure 4](image-url)
3.3. Ground LAI Estimation from Vegetable Indices

Since Table 3 shows that LAI at the bolting and flowering stages is indicative to the yield in oilseed rape, the next step is to estimate LAI using remotely sensed data for remote prediction of yield. We should note that the truth values of LAI used for modeling were derived from the results of the WOFOST model. Some important vegetable indices, including NDVI, two-band enhanced vegetation index (EVI2), simple ratio index (SR), modified soil-adjusted vegetation index (MSAVI), and visible atmospherically resistant index green (VARIgreen), were calculated and the correlations to LAI are evaluated in Table 4.

The vegetable indices were calculated with the Gaofen-1 images. The analysis of multi-year remote sensing data from oilseed rape planting areas with similar phenology around the city that had ground trial fields shows that the tested vegetable indices, except for VARIgreen, are significantly correlated to LAI at the bolting stage with $R^2$ exceeding 0.72. NDVI, EVI2, and SR have non-linear relationships to LAI (Figure 5a,d,e), while MSAVI is linear to LAI (Figure 5c). SR hits the highest correlation to LAI in the test results, with $R^2$ reaching 0.77, which can be used as an LAI inverse model at the bolting stage.

At the flowering stage, only VARIgreen is significantly correlated to LAI with the $R^2$ of 0.71, while all other indices were weakly correlated to LAI because when rape is in its flowering stage, conspicuous yellow flowers appear and occupy the top canopy. LAI at this time is confounded by flowers and leaves that are visually different. Many studies show that vegetable indices are capable of estimating vegetation LAI because the collected canopy reflectance in visible and NIR bands is a result of leaf absorption and scattering, which directly reflects vegetation greenness-related characteristics, such as LAI and chlorophyll content [55]. However, there is a certain degree of irregular decrease in both the green band and the NIR band during the flowering stage [33]. The good performance of VARIgreen at this time is probably due to the fact that the index highlights the green band information, while the yellow band observed at the flowering stage is closer to the green band. Therefore, we chose VARIgreen for the inversion of LAI during the flowering stage.

Table 4. Vegetation indices tested in this study. All indices were calculated from reflectance collected by Gaofen-1 images.

| VI     | Formula                                                                 | Reference |
|--------|-------------------------------------------------------------------------|-----------|
| NDVI   | $(\rho_{NIR} - \rho_{red}) / (\rho_{NIR} + \rho_{red})$                | [56]      |
| VARIgreen | $(\rho_{green} - \rho_{red}) / (\rho_{green} + \rho_{red})$         | [57]      |
| MSAVI  | $\left[ 2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{red})} \right] / 2$ | [58]      |
| EVI2   | $2.5 \times \left( \rho_{NIR} - \rho_{red} \right) / \left( 1 + \rho_{NIR} + 2.4 \times \rho_{red} \right)$ | [59]      |
| SR     | $\rho_{NIR} / \rho_{red}$                                                | [60]      |

After the regression comparison, the LAI at the bolting and flowering stages were calculated with

\[
\text{LAI}_{\text{Bolting Stage}} = -0.1927 \times \text{SR}^2 + 1.5644 \times \text{SR} + 0.4364 \tag{3}
\]

The LAI at the flowering stage is calculated with

\[
\text{LAI}_{\text{Flowering Stage}} = -65.545 \times \text{VARIgreen}^2 + 19.693 \times \text{VARIgreen} + 2.8536 \tag{4}
\]

Combined with Equation (2), which curves the relationship between LAI and yield, the yield is able to be estimated from remote sensing images.
Figure 5. Relationship between ground LAI and vegetable indices from Gaofen-1 images.
3.4. Integrated New Model and Operational Steps

Based on the analysis in the above two sections, we can obtain the yield estimation model of oilseed rape in the Hubei region with

\[
\text{Yield} = 581.82 \times (-0.1927 \times SR^2 + 1.5644 \times SR + 0.4364)_{\text{BoltingStage}} \\
+ 91.47 \times (-65.545 \times \text{VARIgreen}^2 + 19.693 \times \text{VARIgreen} + 2.8536)_{\text{FloweringStage}} - 63.10
\] (5)

To summarize, the yield estimation for oilseed rape can be achieved through the following steps.

1. At least two available Gaofen-1 WFV images are downloaded and preprocessed. The images are captured at the bolting and flowering stages, respectively.
2. The planting areas of oilseed rapes are extracted from the image captured at the flowering stage with the pyramid bottleneck residual network.
3. The vegetable indices are calculated separately using the equations in Table 4, including SR at the bolting stage and VARIgreen at the flowering stage.
4. The yield in each pixel location is calculated with Equation (5).
5. The total yield of oilseed rape is obtained by summing up the yields whose pixel locations belong to both planting areas and the interest region.

4. Validation and Discussion

To validate the prediction model (Equation (5)), we used the pyramid bottleneck residual convolutional neural network to extract the oilseed rape planting area in the validation area, Wuxiang city, Hubei province, from 2014 to 2019 (of which Gaofen-1 data for 2016 are missing), calculated the planting area of oilseed rape, and compared it with statistical yearbook. The used Gaofen-1 images are listed in Table 1 where the images are captured in January for the bolting stage and March for the flowering stage. The constructed rape yield estimation model was used to calculate the annual oilseed rape yield and the results were compared to the statistical yearbook for validation. The results are shown in Table 5. The overall error rate is less than 5.5% and the estimation accuracy is over 94%. The digital evaluations show that the algorithm achieved good performance in estimating the yield of oilseed rape on a large scale in Hubei province.

Table 5. Comparison of estimated yield with statistical yearbook data in Wuxue City.

| Year | Yearbook Planting Area (km²) | Extracted Cultivated Area (km²) | Yearbook Yield (tons) | Estimated Yield (tons) | Error Rate (%) |
|------|-----------------------------|---------------------------------|-----------------------|-----------------------|----------------|
| 2014 | 30                          | 31.6                            | 77,318                | 75,115                | 2.85           |
| 2015 | 28.96                       | 28.22                           | 73,574                | 74,674                | 1.5            |
| 2017 | 27.13                       | 27.81                           | 67,852                | 71,460                | 5.32           |
| 2018 | 18.74                       | 18.21                           | 45,531                | 44,499                | 2.23           |
| 2019 | 21.23                       | 20.08                           | 51,390                | 49,770                | 3.14           |

The insufficient satellite data are possible adverse conditions. In our method, satellite imageries at the blotting and flowering stages of rape oilseed are used. The flowering stage is around mid-March, and the bolting stage is at the end of January and the beginning of February. The main planting regions in China are central and southern areas, which may suffer from continuous rain in February for more than three weeks. Therefore, satellite data with high temporal resolution is suggested for yield estimation. In contrast, Landsat-8 images are not a good choice because of the 16-day revisit period.

The use of the proposed model is constrained by climatic factors. In this experiment, the rape oilseed yield was estimated for a city. However, the suitable areas for rape oilseed span all of China, together with the inconsistency of phenological differences, which are reflected in differences in sowing time and field management. For example, the flowering
stage in a few areas may be delayed due to the colder weather and late planting, then the planting scope and vegetable indices cannot be accurately extracted from Gaofen-1 images from March, which leads to inaccuracy of the overall yield estimation. Therefore, before reproducing the work, it is necessary to consider whether the local climatic factors are the same as our study area, or adjust the time of data collection according to the local climate.

Some other indices were used to predict the yield of oilseed rape. NDYI and BNDVI at the flowering period were constructed in [29] using MODIS data to fit the oilseed rape yield. However, the model based on MODIS data failed to obtain accurate rapeseed yield, indicated by the low $R^2$ values from the linear regression model. Therefore, the effectiveness of this approach needs to be re-evaluated when applied to high-resolution satellite data. The red edge vegetation index ($CI_{\text{red edge}}$) was used in [40] to estimate the yield of oilseed rape, and the error rates were between 5.7% and 7.1%. In contrast, our validation error was below 5.5%. NDYI and $CI_{\text{red edge}}$ may be useful to construct the LAI at the flowering stage. However, given that the overall performance with our step-by-step modeling was already excellent, the significance of this substitution requires enough field measurement data to be more fully examined.

The core of our model is the LAI at critical stages, as the work reveals that accurate LAI is dominant for rapeseed yield prediction. However, our method is still limited by observations, i.e., it requires a specific phase of satellite data acquisition before yield estimation can be made. Constrained by adverse weather conditions and satellite revisit cycles, these data are often missing. In addition, yield estimates only on observations are still not perfect because they are timed so close to harvest time that earlier predictive estimates of the final yield are lacking. These works remain to be improved in future work.

5. Conclusions

In this study, we developed a new algorithm for estimating rapeseed yield over a large area. Our method uses remote sensing images as the input and describes the regression relationship between LAI and oilseed rape yield. Plenty of ground experiments were performed along with the data analysis. The time-series ground LAI was obtained with the localized WOFOST model to extract the LAI of the critical growth periods of oilseed rape. The Gaofen-1 WFV data covering the bolting stage and flowering stage of rapeseed were used to estimate the oilseed rape yield in a large area.

Some conclusions were drawn through the experiment. (1) The combined LAI of the bolting stage and the flowering stage can achieve accurate yield prediction. (2) The simple ratio index is the most correlated with LAI at the bolting stage. (3) At the flowering period, VARIgreen is highly correlated to LAI. The effectiveness of the yield estimation algorithm is verified in Wuxue City, Hubei Province, which shows that the yield estimation error in 2014–2019 is less than 5.5% when compared with the digits in statistical yearbooks. This highlights the feasibility of our method to accurately predict oilseed rape yields at a large regional scale.

Author Contributions: Methodology, W.T.; software, T.G.; validation, T.G.; formal analysis, W.T. and T.G.; investigation, W.T. and J.W.; resources, R.T.; data curation, T.G.; writing—original draft preparation, W.T. and T.G.; writing—review and editing, J.W.; visualization, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (no. 61860130), the Jiangxi Provincial Department of Science and Technology (no. 20204ABC03A40), and the Jiangxi Provincial Institute of Water Sciences (no. 2021SKTR07 and 202224ZDKT11).

Acknowledgments: We thank the research groups led by Jianwei Lu and Shanqin Wang, from the College of Resources and Environment, Huazhong Agricultural University, China, for their hard work in collecting the yield data and their generosity in sharing the data. We acknowledge the data support from the lab from the School of Remote Sensing and Information Engineering, Wuhan University, China.

Conflicts of Interest: The authors declare no conflict of interest.
55. Kross, A.; McNairn, H.; Lapen, D.; Sunohara, M.; Champagne, C. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int. J. Appl. Earth Obs. Geoinf.* 2015, 34, 235–248. [CrossRef]

56. Rouse, J.W.J.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In NASA Special Publication; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1974; Volume 351, p. 309.

57. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 2002, 80, 76–87. [CrossRef]

58. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 1994, 48, 119–126. [CrossRef]

59. Jiang, Z.Y.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 2008, 112, 3833–3845. [CrossRef]

60. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 1969, 50, 663–666. [CrossRef]