Estimation of Winter Wheat Grain Protein Content Based on Multisource Data Assimilation

Pengfei Chen 1,2
1 State Key Laboratory of Resources and Environment Information System, Institute of Geographic Science and Natural Resources Research of Chinese Academy of Sciences, Beijing 100101, China; pengfeichen@igsnrr.ac.cn
2 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

Received: 17 August 2020; Accepted: 29 September 2020; Published: 30 September 2020

Abstract: Data assimilation is a robust method to predict crop biophysical and biochemical parameters. However, no previous study has attempted to predict grain protein content (GPC) at a regional scale using this method. This study explored the feasibility of designing an assimilation model for wheat GPC estimation using remote sensing, a crop growth model, and a priori knowledge. The data included a field experiment and regional sampling data, and Landsat Operational Land Imager images were employed, with the CERES (Crop Environment RESource Synthesis)-Wheat model used as simulation model. To select an optimal method for data assimilation in GPC prediction, different state variable scenarios and cost function solving algorithm scenarios were compared. Additionally, to determine whether a priori information could improve GPC prediction, the collected leaf area index (LAI) and leaf N content sampling data and the range of GPC in the study region were used to constrain the data assimilation process. Furthermore, the data assimilation method was compared to the use of only the CERES-Wheat model. The results showed that GPC could be predicted by remote sensing observation, a crop growth model, and a priori knowledge at regional scale, where the use of data assimilation improved the GPC prediction compared to using only the CERES-Wheat model.

Keywords: data assimilation; grain protein content; winter wheat; CERES-wheat model; satellite image

1. Introduction

Globally, wheat is a very important grain crop. The grain protein content (GPC, %) is a critical indicator that determines the use and commercial value of the wheat [1]. A combination of factors, such as the plant genetics, field management, and weather conditions determine the GPC [2], which show significant spatial differences [3,4]. Therefore, detecting wheat GPC in an accurate, non-destructive, and timely manner at the regional scale is essential for harvesting wheat according to crop quality, thereby ensuring that high-quality crops are sold at appropriate prices. Further, determining the wheat GPC is crucial to support the timely adoption of optimized management measures to modify the growth conditions (e.g., fertilizer, water) to ensure the desired GPC outcome.

Currently, two main types of non-destructive detection methods are used for wheat GPC prediction, namely the crop growth model simulation method and the remote sensing inversion method. Using weather, soil, cultivar, and field management parameters as inputs, wheat growth models, such as APSIM (Agricultural Production Systems Simulator)-Wheat and CERES-Wheat, can simulate the growth and development of wheat, as well as predict the wheat GPC [5–7]. In contrast, remote sensing technology uses the vegetation canopy spectrum acquired by sensors to predict the biophysical and biochemical parameters of vegetation based on their spectral characteristics of the reflectance curve.
Several studies have been conducted to predict wheat GPC using remote sensing data based on the theory that canopy reflectance can predict parameters related to crop growth vigor during the growth season, which, in turn, can be used to estimate the GPC [8,9]. These two methods have various advantages and disadvantages. The growth model method has a superior mechanistic explanation and can be used to dynamically simulate the accumulation of wheat grain nitrogen (N). However, when it is applied at a regional scale, owing to the large number of input parameters in the model, the non-uniformity of the Earth surface environment makes it difficult to obtain various parameters at a spatial scale [10]. When inaccurate parameters are used to simulate GPC, the results are clearly not accurate. The remote sensing method can well reflect the spatial variability of the predicted parameters; however, the crop canopy reflectance spectrum in the growing season contains no direct information regarding the GPC at the end of the season. The spectral data and GPC that are connected by the intermediate variables in the growing season can be determined by remote sensing technology. However, this empirical relationship may not be robust, implying greater uncertainty in estimating the GPC. The characteristics of these two methods indicate their potential to be used together, where combining remote sensing technology and crop growth models to predict crop quality is considered to have great potential [11,12].

The data assimilation method can be used to combine the growth model and remote sensing. Data assimilation is the process of incorporating observations with simulations to improve prediction accuracy and identify the level of uncertainty [13]. The data assimilation methods have been classified into three categories: namely, the forcing, calibration, and updating methods [14]. When remote sensing observations are available, forcing methods use the estimated variables from remote sensing to replace the crop model simulation variable when running the model to estimate the target variable. The updating method continuously updates crop model simulation variables according to the uncertainty of the remote sensing variables and crop model variables, and drives the model to estimate the target variable. The calibration method adjusts the input parameters of the crop models to ensure optimal consistency between the remote sensing data and the simulated state variables and, subsequently, uses the adjusted input parameters to estimate the target variable [15]. The forcing method is easy to operate; however, it is assumed that the model parameters are calibrated and the accuracy of the observed data is high [16]. The updating method is more flexible than the forcing method, but it is influenced by the measurement uncertainty of the remote sensing and simulated variables, and the choice of observed remote sensing variables [17]; theoretically, the calibration method is superior to the other methods, but it requires numerous optimization iterations, resulting in a long computing time [15]. These methods have been employed in several studies to estimate biomass [18,19], leaf area index (LAI) [20,21], and crop yield [22,23] values by combining remote sensing data and a crop growth model.

However, few studies have estimated crop quality by employing the data assimilation method. Only Li et al. [16,24], coupling ground-measured winter wheat spectral data with the CERES-Wheat model, have used the calibration data assimilation method to estimate GPC. Their results showed that assimilating remote sensing data into the CERES-Wheat model could successfully predict the GPC. However, the assimilation method based on ground-sampled remote sensing data is flawed in that it is difficult to accomplish monitoring of large areas over a short observation period. In contrast, satellite imaging clearly has an advantage for regional monitoring. However, until recently, no study has attempted to estimate wheat GPC by assimilating satellite images and a crop growth model. Further, national, and other experimental stations, such as those at research institutes or universities, have long-term observation data on crop phenology, yield, and the like. These historical data can be regarded as a priori knowledge, and could be helpful in GPC prediction. A priori knowledge is helpful to overcome the ill-posedness of the mechanism model inversion when the number of remote sensing observations is insufficient [25]. As stated by Liang and Qin [26], a priori information can be incorporated into a data assimilation model to help predict the target variable. However, this aspect has not been discussed in existing studies on crop quality estimation.
Therefore, this study was conducted to estimate winter wheat GPC by coupling satellite images, a crop growth model, and a priori information using the data assimilation method. The objectives of this study were: i) to study the possibility of estimating wheat GPC by assimilating satellite images with the crop growth model; ii) to evaluate the effect of adding a priori information into the assimilation process on the accuracy of GPC estimation; and iii) to evaluate if the data assimilation method performed better than only using the crop growth model for GPC prediction. Remote sensing prediction models for the LAI and leaf N content (LNC) were developed, and data assimilation under different designed scenarios was performed. Then, the predicted GPC values for data assimilation under different scenarios were compared with use of only the crop growth model. The selection of optimal state variables and cost function solving algorithms, and the effect of a priori knowledge on GPC prediction using data assimilation are discussed. These findings are expected to promote the use of data assimilation methods with crop growth models and contribute to the improved management of grain crops to optimize their yield and quality.

2. Materials and Methods

2.1. Study Site

The study was conducted in the city of Yucheng (36°33′36″, 116°23′24″), which covers an area of 990 km² in the northwestern part of Shandong Province, China (Figure 1). The site is characterized by a warm, temperate, semi-humid monsoon climate, with 13.3 °C annual averaged temperature, 556 mm of annual averaged precipitation, and 2546.2 h of annual averaged total sunshine [27]. The main farming type in the region is the winter wheat and summer maize double-cropping system, with winter wheat generally planted at the beginning of October and harvested in June of the next year.

2.2. Crop Growth Model

The decision support system for agrotechnology transfer (DSSAT) has been used widely over the last decades by numerous researchers in several different applications. The DSSAT version 4.6 can simulate more than 28 crops, including maize, wheat, and rice, and was used in the current study. The CERES-Wheat model in DSSAT is used especially for wheat phenology, biomass, yield, simulating water and N balances, and using meteorological data, soil information, crop management data, and
crop genotype parameters as input [28]. In the CERES-Wheat model, daily carbon accumulation was simulated as a function of incoming solar radiation, LAI, plant population, canopy extinction coefficient, and radiation-use efficiency and, subsequently, allocated into different plant organs according to potential growth and developmental stage [19]. The N accumulation of the grain was divided into phases [29]. In each phase, the grain N was derived from the direct root uptake of N during the grain filling stage and the retranslocation of N from the pre-stored N in other organs, which are functions of environmental factors, such as rainfall, temperature, and solar radiation [16]. The CERES-Wheat model comprises several cultivar and ecological parameters, which need to be calibrated using local field experimental data before the model can be used to simulate wheat growth. More information regarding the calibration of the cultivar and ecological parameters in this study is given in the following section.

2.3. Data Acquisition

2.3.1. Field Experimental Data

A two-year (2010–2012) winter wheat N field experiment was conducted on a farm in the city of Yucheng, China. The N experiment comprised five N treatments in a randomized complete block design, with four replications. The soil texture (0–20 cm) is loamy, with a pH value of 8.0 (in water). The average soil organic matter, NO3-N, available P, and available K levels were 3.1%, 26.3 mg kg−1, 22.8 mg kg−1, and 126.0 mg kg−1, respectively, in the 0–20 cm soil layer before the N experiment. The five N fertilization treatments comprised 0, 70, 140, 210, and 280 kg N ha−1. The cultivar “JiMai 22” was mainly used in the research area, sown at a density of 4,500,000 plants ha−1 on 17 October 2010 and 13 October 2011. The plot size was 10 × 5 m, and all treated plots received 120 kg P2O5 ha−1 and 80 kg K2O ha−1 before sowing. During the wheat growth stage, detailed field management measures and the duration of each wheat growth stage were recorded. In addition, field campaigns were conducted to measure LAI in the jointing, booting, anthesis, and grain filling stages. During data collection, representative sites were selected in each plot and the LAI was measured using the LAI-2200 plant canopy analyzer (LI-COR, Nebraska, America). When the wheat reached maturity, grain in an area of 1.6 × 3.0 m2, located in a representative area of each plot, was harvested, and subsequently dried at 65 °C until a constant weight was reached. A subsample of grain of approximately 100 g was collected for each plot and milled into powder. The N concentration of the powder was measured using the Dumas combustion method and a Vario MACRO cube analyzer (Elementar, Langenselbold, Germany). The GPC was calculated by multiplying the measured powder N concentration by 6.25.

2.3.2. Regional Sampling Data

Considering the spatial distribution of wheat in the study area, 23 sampling points were pre-selected using a simple random sampling method. In 2014 and 2015, LAI and LNC (g m−2) were measured at the jointing (29 March 2014; 30 March 2015), booting (16 April 2014; 17 April 2015), anthesis (29 April 2014; 4 May 2015), and grain filling (9 May 2014; 21 May 2015) growth stages at each selected location. During the field campaign, a representative area near the center of each pre-selected site was chosen and the latitude and longitude values of the position were obtained using a handheld GPS device (Juno SD Handheld, Trimble, America). For LAI measurements, the LAI-2200 instrument was used. After LAI measurements, two rows of wheat with a length of 1 m were harvested at each sampling point and taken to the laboratory. The leaves were separated from the stems, placed into an oven, dried to a constant weight, and weighed using a balance. Subsequently, the samples were crushed and the leaf N concentration (%) was measured using the Vario MACRO cube analyzer. The LNC was calculated by multiplying the leaf N concentration by the total weight of the sampled leaf, divided by the sampling area. Additionally, in 2014, when the wheat reached maturity, another 13 sampling points were added randomly between the above-mentioned 23 sampling points, i.e., a total of 36 sampling points were used to determine the wheat GPC. For the GPC measurements, four rows of wheat with a length of 1 m were harvested at each sampling point, after which the grain was dried, milled into powder, and
measured by the method described above. As the “JiMai 22” cultivar accounted for more than 85% of the research region, we assumed that most of the collected samples contained this cultivar. The final distribution of the sampling points is shown in Figure 1.

2.3.3. Remote Sensing Data

Optical satellite images from Landsat 8 Operational Land Imager (OLI) were used in this study, which has a spatial resolution of 30 m. In 2014, three scenes of cloudless images were collected on 21 March, 22 April, and 8 May during the wheat growth stage. For each scene, two images (path numbers 122/34 and 122/35) were required to cover the entire research area. The following steps were used for image pre-processing. First, the digital number (DN) values of the two images were converted to top-of-atmosphere radiance values. Second, the FLAASH module of ENVI 5.1 software (EXELIS Company, Mclean, VA, United States) was used for atmospheric correction. Third, geometric correction was done based on the sampled GPS points. Finally, the two images were mosaicked, resized, and masked to retain only Yucheng city. These images were used as remote sensing data during data assimilation.

2.3.4. Meteorological Data

Meteorological data are considered as the critical driving data for the crop growth model. As the experimental farm is located close to the Ecological Comprehensive Experimental Station of Yucheng, meteorological data observed by their automatic weather station over period of 2010–2012 were used in the current study. These data included solar radiation, sunshine hours, maximum and minimum temperatures, and precipitation. During 2013–2014, in addition to the data observed by the automatic weather station in Yucheng, meteorological data from three national weather stations (Nos 54715, 54808, and 54823) in the vicinity of Yucheng were also collected from the daily climate dataset (V3.0) shared by the National Meteorological Information Center (http://data.cma.cn). These data only included the number of sunshine hours, maximum and minimum temperatures, and precipitation. As the CERES-Wheat model requires solar radiation information rather than the number of sunshine hours to perform wheat growth simulation, the sunshine hours were converted into radiation using the method of Teh [30]. After the conversion, to ensure accuracy, a regression model was constructed between the converted radiation and actual radiation at Yucheng experimental station and used for further correction of the converted data. The grid data (30 × 30 m) for radiation, maximum and minimum temperatures, and precipitation in 2013–2014 in Yucheng city were generated using the inverse distance interpolation method to interpolate the meteorological data of the four weather stations to correspond to the spatial resolution of the remote sensing image.

2.4. Data Analysis Method

The calibration data assimilation method was used in this study and the following data analysis processes were performed. First, the two-year field experimental dataset was used to calibrate the cultivar and ecological parameters in the CERES-Wheat model. During these processes, all cultivar parameters were calibrated first, after which seven ecological parameters were calibrated (listed in Table 1). During calibration of the cultivar parameters, the ecological parameters were set at their default values, and the cultivar parameters were changes in given steps within the ranges shown in Table 1. The optimal cultivar parameter values were defined as those with the lowest error for simulating the emergence time of the growth period, LAI, and GPC values. During calibration of the ecological parameters, the cultivar parameters were set at the optimal values and the ecological parameters were changes in given steps within the ranges shown in Table 1. The optimal ecological parameters were defined as those with the lowest error for simulating the same outputs as for previous calibration step. During these processes, the CERES-Wheat model was run for each N treatment with the measured soil parameters and recorded field management parameters. For each N treatment, the actual values of the emergence time of the growth period, LAI, and GPC were the average of four
replications. The determination coefficient ($R^2$) and root mean square error (RMSE) were used to evaluate the model performance.

Table 1. Calibrated cultivar and ecological parameters for CERES (Crop Environment REsource Synthesis)-Wheat model calibration.

| Parameter                                      | Default | Range         |
|------------------------------------------------|---------|---------------|
| Vernalization time parameter (P1V)             | 5.00    | 0.0–60.00     |
| Photoperiod parameter (P1D)                    | 75.00   | 0.0–200.00    |
| Grain filling phase duration parameter (P5)    | 450.00  | 100.00–999.00 |
| Kernel number parameter (G1)                   | 30.00   | 10.00–50.00   |
| Kernel size parameter (G2)                     | 35.00   | 10.00–80.00   |
| Tiller weight parameter (G3)                   | 1.00    | 0.50–8.00     |
| Phyllochron interval parameter (PHINT)         | 60.00   | 30.00–150.00  |
| Photosynthetically active radiation conversion ratio parameter (PARUE) | 2.70    | 1.00–6.00     |
| Root depth growth rate (RDGS)                  | 3.00    | 1.00–5.00     |
| Duration of phase end juvenile to terminal spikelet parameter (P1) | 200.00  | 100.00–500.00 |
| Duration of phase terminal spikelet to end leaf growth parameter (P2) | 200.00  | 100.00–500.00 |
| Vernalization effect parameter (VEFF)          | 0.60    | 0.00–1.00     |
| Vegetative phase parameter (LAFV)              | 0.10    | 0.01–0.80     |
| Specific leaf area parameter (SLAS)            | 400.00  | 100.00–600.00 |

Second, during data assimilation, the CERES-Wheat model was run on a per-grid basis matching the OLI data and a state variable was used to connect the remote sensing data and the crop growth model. Considering the feasibility of the variables being estimated accurately by remote sensing and the connection between these variables and GPC, three scenarios of state variables were designed in this study: namely, using the LAI, LNC, or both the LAI and LNC as state variables. In each scenario, the Extended Fourier Amplitude Sensitivity Test (EFAST) method was used to select the variable among the soil and field management input parameters in the CERES-Wheat model. The assimilation process was used to adjust the selected parameters to improve their accuracy. The criterion for selecting a variable was that it had to be sensitive to the state variables and GPC. EFAST is a variance-based method used to quantify global sensitivity [31,32] that combines the advantage of the FAST method to compute the total sensitivity index and computational efficiency, with the capacity of the Sobol method to compute total effects and second order effects [33]. The soil and field management parameters of the CERES-Wheat model and their values are shown in Table 2. The EFAST analysis was performed by using the Simlab 2.2.1 software package, provided by the EU SCIENCE HUB (https://ec.europa.eu/jrc/en/samo/simlab).

Third, using remote sensing images and regional sampling data, the LAI and LNC prediction model was constructed based on the spectral index method. During this process, commonly used spectral indices were selected to estimate the LAI and LNC (Table 3). Taking the LAI prediction as an example; for each index, linear, logarithmic, exponential, and power models were used to design the LAI prediction model. The optimal model with the highest $R^2$ and lowest RMSE values for prediction of LAI were retained. Then, results from different indices were compared and the best index was selected and subsequently used to estimate LAI. As the acquisition times of field sampling data and remote sensing images were not the same, these data were time matched. For LAI, a simple and accurate LAI simulation model created by Baret was used [34], which describes the relationship between the accumulated temperature and LAI. For each sampling site, LAI values measured by field sampling and accumulated temperature values were used to fit the model to obtain model coefficients. Then, the LAI from remote sensing images on the corresponding acquisition date was estimated by the model. The temperatures of each sampling site were obtained using the GPS data and the grid methodology data described in Section 2.3.4, followed by subsequent calculation of the accumulated temperature. Unlike
Baret’s method, the accumulated temperature was calculated using data above 0 °C instead of above 8 °C, as elaborated by Casa et al. [35]. As there is no simple model to simulate LNC, the CERES-Wheat model was used to estimate the LNC for each sampling site, where the variables selected in the second step were varied over the ranges shown in Table 2. When the difference between the simulated and actual LNCs at the sampling time reached a minimum, these values were selected as the final model input for simulating the LNC at the acquisition time of the remote sensing images. During the process, with the exception of the selected variables, all soil and field management input variables were set to their default values (Table 2). Additionally, the methodology data of each site were extracted from the grid methodology data described in Section 2.3.4 using their GPS information.

Table 2. Soil and field management parameters in the CERES-Wheat model and values.

| Parameter                        | Default            | Ranges                          |
|----------------------------------|--------------------|---------------------------------|
| Planting date                    | 17 October         | 4–28 October #                  |
| Plant population at emergence    | 450.00             | 300.00–450.00 #                 |
| (10,000 ha⁻¹)                    |                    |                                 |
| Row spacing (cm)                 | 20.00              | 15.00–20.00 #                   |
| Plant depth (cm)                 | 3.00               | 3.00–6.00 #                     |
| First irrigation time            | 21 November        | 20 November–10 December #      |
| First irrigation amount (mm)     | 150.00             | 80.00–150.00 #                  |
| Second irrigation time           | 23 March           | 15–31 March #                   |
| Second irrigation amount (mm)    | 100.00             | 80.00–150.00 #                  |
| Third irrigation time            | 19 April           | 17 April–10 May #               |
| Third irrigation amount (mm)     | 100.00             | 80.00–150.00 #                  |
| First fertilization time (same   | 17 October         | 4–28 October #                  |
| as first irrigation time)        |                    |                                 |
| First applied N amount (kg N ha⁻¹)| 105.00            | 140.00–200.00 #                 |
| Fertilizer type for first        | Urea               |                                 |
| fertilization                   |                    |                                 |
| Second fertilization time (same  | 19 April           | 17 April–10 May                 |
| as third irrigation time)        |                    |                                 |
| Second applied N amount (kg N ha⁻¹)| 105.00            | 140.00–200.00 #                 |
| Fertilizer type for second       | Urea               |                                 |
| fertilization                   |                    |                                 |
| Soil nitrate (g N Mg⁻¹)          | 21.60              | 10–30 $                        |
| Root weight (kg ha⁻¹)            | 1,658.00           | 800.00–1,800.00 $               |
| Water table depth (cm)           | 190.00 *           |                                 |
| Albedo                           | 0.11 *             |                                 |
| Evaporation limit (mm)           | 6.00 *             |                                 |
| Drainage rate (% day⁻¹)          | 0.10 *             |                                 |
| Runoff curve number              | 80.00 *            |                                 |
| Mineralization factor            | 1.00 *             |                                 |
| Photosynthesis factor            | 1.00 *             |                                 |
| Soil lower limit (cm⁻³ cm⁻³)     | 0.10 *             |                                 |
| Soil upper limit (cm⁻³ cm⁻³)     | 0.30 *             |                                 |
| Soil upper limit (cm⁻³ cm⁻³)     | 0.45 *             |                                 |
| Bulk density (g cm⁻³)            | 1.54 *             |                                 |
| Soil clay content (%)            | 3.00 *             |                                 |
| Soil silt content (%)            | 90.00 *            |                                 |
| Root growth factor               | 1.00 *             |                                 |
| Soil organic carbon (%)          | 1.00               | 0.30–2.60 $                     |
| Soil total N (%)                 | 0.09               | 0.05–0.15 $                     |
| Soil pH in water                 | 8.20 *             |                                 |
| Cation exchange capacity (mol kg⁻¹)| 10.00 *         |                                 |

# Parameter range defined by the field management strategy in the study area. * Parameter variation was small and the values were calculated from the average of long-term observation data from the Yucheng experimental station. $ Parameter range determined from data from the second national soil census in China.


with their priori knowledge, as several agricultural ecological experimental stations have long-term monitoring value. The terminal condition was reaching the minimal cost function value or the maximum number of iterations, with the simulated GPC within the given range. When the terminal condition was met, the study area was used as a priori information during data assimilation to control the predicted GPC algorithm [46]. Additionally, to prevent ill-posed inversion, the range of GPC (11% ≤ GPC ≤ 16%) was included the simulated annealing (SA) algorithm [44], the shuffled complex evolution algorithm method developed at the University of Arizona (SCE-UA) [45], and the differential evolution (DE) algorithm [46]. Additionally, to prevent ill-posed inversion, the range of GPC (11% ≤ GPC ≤ 16%) in the study area was used as a priori information during data assimilation to control the predicted GPC value. The terminal condition was reaching the minimal cost function value or the maximum number of iterations, with the simulated GPC within the given range. When the terminal condition was met, the values of the adjusted variables were used to drive the model to predict the GPC. To evaluate the model performance, the predicted and actual GPC values at the sampling sites were compared, along with their \( R^2 \) and RMSE values. Then, the results from different state variable scenarios and different cost function solving algorithms were compared to select the optimal method.

Fourth, the calibration data assimilation method was used for coupling the remote sensing images and the CERES-Wheat model to predict GPC on a per-grid basis matching the OLI data, as described earlier. During this process, for each scenario, the CERES-Wheat model was driven with the selected adjusted field management variables varied over the ranges shown in Table 2. The simulated state variables and remote sensing state variables were compared and a cost function was computed. For the different scenarios, the cost functions are expressed in Equation (1).

\[
\sum_{i=1}^{n} \left( \frac{LAI_{\text{sim},i} - LAI_{\text{obs},i}}{n} \right)^2 + \sum_{j=1}^{n1} \left( \frac{LNC_{\text{sim},j} - LNC_{\text{obs},j}}{n1} \right)^2
\]

Here, \( w \) is the cost function when coupling remote sensing images and the crop growth model; \( n \) is the total number of remote sensing LAI values (when only LNC is used as the state variable, \( n = 0 \)); \( LAI_{\text{sim},i} \) is the LAI value at time \( i \) from the CERES-Wheat model; \( LAI_{\text{obs},i} \) is the remote sensing LAI value at time \( i \); \( n1 \) is the total number of remote sensing LNC values (when only LAI is used as a state variable, \( n1 = 0 \)); \( LNC_{\text{sim},j} \) is the LNC at time \( j \) from the CERES-Wheat model; and \( LNC_{\text{obs},j} \) is the remote sensing LNC at time \( j \).

Before calculating the cost function, all parameters were normalized to between 0 and 1 to eliminate errors resulting from differences in the magnitudes of the variables. Three algorithms conventionally used in research were used to update the adjusted parameter to obtain minimum cost function values, including the simulated annealing (SA) algorithm [44], the shuffled complex evolution algorithm method developed at the University of Arizona (SCE-UA) [45], and the differential evolution (DE) algorithm [46]. Additionally, to prevent ill-posed inversion, the range of GPC (11% ≤ GPC ≤ 16%) in the study area was used as a priori information during data assimilation to control the predicted GPC value. The terminal condition was reaching the minimal cost function value or the maximum number of iterations, with the simulated GPC within the given range. When the terminal condition was met, the values of the adjusted variables were used to drive the model to predict the GPC. To evaluate the model performance, the predicted and actual GPC values at the sampling sites were compared, along with their \( R^2 \) and RMSE values. Then, the results from different state variable scenarios and different cost function solving algorithms were compared to select the optimal method.

Finally, the LAI and LNC over two years collected from the regional sampling dataset were averaged for the jointing, booting, and anthesis growth stages. These data were also considered as a priori knowledge, as several agricultural ecological experimental stations have long-term monitoring data on these parameters, especially LAI. The averaged LAI and LNC values were incorporated into the data assimilation process to determine whether this information could improve the accuracy of...
GPC estimation. The selected optimal state variable scenario and cost function solving algorithm in the above step were used and the cost function was changed into Equation (2).

$$w_1 = w + \sqrt{\frac{\sum_{k=1}^{n_2} (LAI_{Sim,k} - LAI_{pri,k})^2}{n_2} + \frac{\sum_{h=1}^{n_3} (LNC_{Sim,h} - LNC_{pri,h})^2}{n_3}}$$

(2)

Here, $w_1$ is the cost function including the long-term average LAI and LNC values as a priori knowledge; $n_2$ is the total number of wheat growth stages for the average LAI (when only LNC is used as a state variable, $n_2 = 0$); $LAI_{Sim,k}$ is the simulated LAI at time $k$ from the CERES-Wheat model; $LAI_{pri,k}$ is the average LAI at time $k$; $n_3$ is the total number of wheat growth stages for the average LNC (when only LAI is used as a state variable, $n_3 = 0$); $LNC_{Sim,h}$ is the simulated LNC value at time $h$; and $LNC_{pri,h}$ is the average LAI at time $h$.

The prediction result was compared with the result in step 4. The flowchart for data assimilation is shown in Figure 2. Additionally, in order to analyse whether the data assimilation method increased the accuracy of GPC estimation compared to using only the crop growth model, the CERES-Wheat model was also run on a per-grid basis to predict GPC and the prediction result was compared with the best result obtained using the data assimilation method.

![Flowchart for data assimilation.](image-url)
3. Results

3.1. CERES-Wheat Model Calibration Results

After the cultivar and ecological parameters were calibrated, the CERES-Wheat model achieved satisfactory prediction results (Figure 3). The $R^2$ value between the predicted and actual appearance time of the wheat growth stage was 0.98, with an RMSE value of 5 d. The $R^2$ value between the predicted and actual LAI was 0.91, with an RMSE value of 0.70. The $R^2$ between the predicted and actual GPC was 0.59, with an RMSE value of 1.32%. These findings, indicating the satisfactory performance of the calibrated CERES-Wheat model, implied that this model could be used with confidence for data assimilation with remote sensing images in the current study.

![Figure 3. Calibration results for the CERES-Wheat model. (a) Appearance time of growth stage; (b) leaf area index (LAI); and (c) grain protein content (GPC).](image)

3.2. Selected Adjusted Soil and Field Management Input Variables for Each State Variable Scenario

When the LAI was used as the state variable, the selected variables must be sensitive to both LAI and GPC. The EFAST sensitivity analysis results shown in Figure 4 indicate that the planting date, amount of water in the third irrigation, soil organic carbon, and soil nitrate content were the soil and field management input parameters sensitive to both the LAI and GPC. Accordingly, these parameters were selected as adjusted variables when LAI was used as a state variable.
Figure 4. Extended fourier amplitude sensitivity test (EFAST) sensitivity analysis results for selecting adjusted variables using LAI as the state variable. Sensitivity of (a) GPC; (b) LAI at the booting growth stage; (c) LAI at anthesis growth stage; (d) LAI at the grain filling growth stage to soil and field management parameters.

When LNC is used as a state variable, the selected adjusted variables must be sensitive to both LNC and GPC. As shown by the EFAST sensitivity analysis results in Figure 5, the planting date, amount of water in the third irrigation, soil organic carbon, and soil nitrate content were the soil and field management input parameters that were sensitive to both LNC and GPC. Hence, these parameters were selected as adjusted variables when the LNC was used as the state variable.

Accordingly, when both the LAI and LNC were used as state variables for data assimilation, the planting date, amount of water in the third irrigation, soil organic carbon, and soil nitrate content were selected as state variables.
Figure 5. EFAST sensitivity analysis results for selecting adjusted variables using leaf nitrogen content (LNC) as the state variable. Sensitivity of (a) GPC; (b) LNC at the booting growth stage; (c) LNC at the anthesis growth stage; (d) LNC at the grain filling growth stage to soil and field management parameters.

3.3. Results for Remote Sensing Observation

The optimal regression results between the spectral indices and LAI and LNC values are shown in Table 4. For LAI prediction, triangular vegetation index (TVI) was the optimal prediction index, with an $R^2$ value of 0.67 and RMSE value of 1.21. For LNC prediction, enhanced vegetation index (EVI) was the optimal indicator, with an $R^2$ value of 0.76 and RMSE value of 26.26 kg ha$^{-1}$. The cross-validation results of these are shown in Figure 6. During cross-validation, the LAI model constructed using the TVI and the LNC model constructed using the EVI performed well, with $R^2$ values of 0.65 and 0.70, and RMSE values of 1.24 and 26.96 kg ha$^{-1}$, respectively. Accordingly, TVI was selected to predict LAI and EVI was selected to predict leaf N content using Landsat OLI data.
Table 4. Results of regression analysis between spectral indices and LAI and LNC values. Linear, logarithm, exponential, and powder models were used to conduct regression. The results of the optimal calibrated model are shown in the table. Regional sampling data for 2014 were used.

| Spectral Index | LAI $R^2$ | LAI RMSE | LNC $R^2$ | LNC RMSE (kg ha$^{-1}$) |
|---------------|-----------|----------|-----------|------------------------|
| EVI           | 0.63      | 1.44     | 0.76      | 26.26                  |
| NDVI          | 0.40      | 1.73     | 0.44      | 39.98                  |
| RVI           | 0.37      | 1.67     | 0.41      | 42.54                  |
| TVI           | 0.67      | 1.21     | 0.77      | 28.06                  |
| MSAVI         | 0.61      | 1.32     | 0.68      | 32.67                  |
| OSAVI         | 0.64      | 1.22     | 0.73      | 25.68                  |
| GNDVI         | 0.29      | 1.86     | 0.29      | 44.35                  |
| MTVI2         | 0.60      | 1.33     | 0.69      | 32.42                  |

Figure 6. Cross-validation results for the (a) LAI model designed by TVI; and (b) LNC model designed by EVI.

3.4. Data Assimilation Results

The data assimilation results for GPC estimation under different state variable and cost function solving algorithm scenarios are shown in Table 5. Optimal results were achieved using both LAI and LNC as the state variables, for all cost function solving algorithms. In this scenario, both the relationship between the actual GPC and that predicted by the DE algorithm or by the SCE-UA algorithm were significant at the 0.01 level, with $R^2$ values of 0.26 and 0.39 and RMSE values of 1.25% and 0.91%, respectively. The relationship between the actual GPC and that predicted by the SA algorithm was not significant at the 0.01 level, with an $R^2$ value of 0.09 and a RMSE value of 1.11%. For each cost function solving algorithm, the results were close when using only LAI or LNC as the state variable. For the DE algorithm, the relationships between the actual GPC and that predicted using LAI (or LNC) as the state variable were not significant at the 0.01 level, with $R^2$ values of 0.14 and 0.11 and RMSE values of 1.20% and 1.44%, respectively. For the SA algorithm, the relationships between the actual GPC and that predicted using LAI (or LNC) as the state variable were not significant at the 0.01 level, with $R^2$ values of 0.07 and 0.05 and RMSE values of 1.27% and 1.29%, respectively. For the SCE-UA algorithm, the relationships between the actual GPC and that predicted using LAI (or LNC) as a state variable were significant at the 0.01 level, with $R^2$ values of 0.18 and 0.21 and RMSE values of 1.08% and 1.29%, respectively. Considering the different cost function solving algorithms, SCE-UA obtained optimal results, followed by DE, with SA obtaining the worst results. Accordingly, the final assimilation model for GPC prediction used both LAI and LNC as state variables, along with the SCE-UA algorithm to solve the cost function, to give the prediction results shown in Figure 7a.
Table 5. Data assimilation results for GPC estimation under different state variable and cost function solving algorithm scenarios.

| State Variable Scenario | Cost Function Solving Algorithm Scenario | $R^2$ | RMSE (%) |
|-------------------------|------------------------------------------|-------|----------|
| LAI                     | DE                                       | 0.14  | 1.20     |
|                         | SCE-UA                                   | 0.18**| 1.08     |
|                         | SA                                       | 0.07  | 1.27     |
| Leaf N content          | DE                                       | 0.11  | 1.44     |
|                         | SCE-UA                                   | 0.21**| 1.29     |
|                         | SA                                       | 0.05  | 1.29     |
| LAI and leaf N content  | DE                                       | 0.26**| 1.25     |
|                         | SCE-UA                                   | 0.39**| 0.91     |
|                         | SA                                       | 0.09  | 1.11     |

** means significant at 0.01 level; SA means simulated annealing algorithm; SCE-UA means the shuffled complex evolution algorithm method developed at the University of Arizona; DE means the differential evolution algorithm.

Figure 7. Predicted vs. actual GPC values for (a) the data assimilation method with both the LAI and LNC as state variables and using the SCE-UA algorithm to solve the cost function; (b) adding historical observed average LAI and LNC data as a priori knowledge; (c) using only the CERES-Wheat model.

Two-year regional sampling LAI and LNC data were averaged for each growth stage and used as a priori knowledge in the data assimilation process. The GPC prediction results after adding this information are shown in Figure 7b. The figure shows that adding this information did not improve the model estimation results. In fact, the performance declined, with an $R^2$ value of 0.31 and a RMSE value of 0.89%.

Additionally, Figure 7c shows the predicted vs. actual GPC values for the CERES-Wheat model without data assimilation. The performance of this model was poor, as the relationship between...
the actual and predicted GPC was not significant at the 0.01 level, and the $R^2$ value was only 0.01. The predicted GPC did not reflect its actual spatial variability in the study region. Hence, the data assimilation method was clearly better than using only the CERES-Wheat model method for GPC prediction.

4. Discussion

4.1. Optimal State Variables and Cost Function Solving Algorithm

As stated before, crop LNC and LAI values can be detected by remote sensing technology with high accuracy [47,48], where these parameters during the wheat growth stage affect the GPC value at the end of the season [49]. Accordingly, these factors were considered as candidates for state variables when assimilating the CERES-Wheat model with remote sensing images. The optimal results were achieved when using both LAI and LNC as state variables, which is consistent with the findings of Li et al., who conducted wheat GPC prediction based on combining the CERES-Wheat model with field sampled hyperspectral reflectance data [24]. Thus, when estimating wheat GPC using the data assimilation method, it is recommended to use both LAI and LNC as state variables.

Different cost function solving algorithms have been compared in previous studies for predicting various crop parameters, but not for GPC prediction. Xing et al. compared the SA, SCE-UA, and particle swarm optimization (PSO) algorithms for assimilating the AquaCrop model with remote sensing data to estimate wheat above ground biomass, canopy cover, and yield [50]. The authors suggested that SCE-UA was a superior method. In the current study, the SA, DE, and SCE-UA algorithms were compared, and the SCE-UA algorithm was again found to be superior. Consequently, SCE-UA is recommended as the optimal algorithm for solving the cost function in GPC prediction using the data assimilation method.

4.2. Effect of a Priori Knowledge on GPC Prediction Results Using Data Assimilation

The application of a priori knowledge in data assimilation can improve forecasting [51]. Considering a priori LAI values at different crop growth stages as constraint information in the cost function, Wang et al. obtained improved LAI predictions compared with excluding this information during the assimilation of remote sensing data with the CERES-Wheat model [52]. Until recently, no study has considered the effect of adding a priori information to the process of the GPC prediction model using data assimilation. In this study, we included two items of a priori information: the range of GPC in the study area, and the average LAI and LNC values at different wheat growth stages using sampled data obtained during long-term observation. The range of GPC represents crucial a priori information to prevent ill-posed inversion. We attempted data analysis without this information in each state variable scenario and cost function solving algorithm scenario. Without such information, illogical grain protein inversion results sometimes occurred (for clarity, these data are not shown). This is consistent with the findings of Cui et al., who stated that a priori knowledge helped overcome the ill-posedness of the model inversion when the number of observations is insufficient [25]. Thus, it is recommended to use the GPC range as a priori information when estimating GPC using the data assimilation method. In this study, the addition of a priori average LAI and LNC data did not improve the GPC prediction results during the data assimilation process. This conclusion differed from that of the above mentioned previous study estimating LAI values [52]. This was attributed to the estimation accuracy being affected by the remote sensing observation error when using data assimilation. When the error of state variables estimated by remote sensing is larger (smaller) than the difference between the average state variable in the historical observation and the actual value of the state variable, adding a priori data increased (decreased) the estimation accuracy of GPC. Accordingly, the a priori average LAI and LNC data should be used according to the specific situation when estimating GPC using the data assimilation method.
4.3. Data Assimilation vs. Only Crop Growth Model

GPC estimation using the crop growth model has a good mechanistic explanation. However, it contains many soil and field management input parameters. On a regional scale, it is impossible to obtain values for these parameters for each specific piece of land, so average values for the study area are commonly set as default values. In this case, the simulated GPC has the spatial resolution of the meteorological data used in the model. When the study area is small, there is little variation in the meteorological data over the region, and the predicted GPC does not reflect the actual spatial variation. The results of the sensitive analysis performed in this study showed that some input parameters of the CERES-Wheat model (i.e., the planting date, soil organic carbon, soil nitrate content, and amount of water used in the third irrigation) have a significant impact on the final GPC value. These four parameters have a high spatial variation, especially the planting date. Thus, the predictions in this study using only the crop growth model were poor. In contrast, the data assimilation method combined the crop growth model with remote sensing, which can provide the actual biophysical and biochemical parameters of the crops. Based on this information, we were able to correct the values of four critical input parameters to achieve GPC predictions better than those using only the crop growth model, where the predicted GPC reflected the actual spatial variation over the study area. Thus, we conclude that the combination of data assimilation and the crop growth model provides better GPC predictions on a regional scale than use of only the crop growth model.

4.4. Comparison with Existing Studies, and Future Development

Previous research regarding GPC prediction using the data assimilation method was conducted mainly by Li et al. [16,24]. Based on field sampled spectral data and the CERES-Wheat model, they obtained an $R^2$ value of 0.519 and a RMSE value of 1.53%, with LNC selected as the state variable [16]. Further, they obtained an $R^2$ value of 0.758 and a RMSE value of 1.16% when using both LAI and LNC as state variables [24]. However, data assimilation of GPC using ground remote sensing data is limited in observation range. In the current study, Landsat OLI images and the CERES-Wheat model were coupled to predict GPC. As stated by Li et al., coupling satellite data with a growth model is important for GPC prediction in precision agriculture [16]. The GPC results obtained in this study were inferior to those published by Li et al., which can be ascribed to two main factors: i) compared with ground sampled spectra, the problem with satellite data is mixed pixels. The scale mismatch between field measured LAI and LNC data and image resolution resulted in errors in the remote sensing estimation model for these parameters. ii) In the current study, the field measured LAI and LNC was interpolated for the data to correspond with the time the satellite images were collected when constructing the LAI and LNC prediction models employing remote sensing information. An error in the data interpolation process would be transferred to the LAI and LNC prediction models. However, we determined that the wheat GPC could be estimated by the CERES-Wheat model, OLI images, and a priori information with moderate accuracy.

Further studies are required to improve the GPC prediction accuracy at a regional scale, where the following two aspects should be considered. First, the accuracy of remote sensing predictions of state variables should be increased. With the development of earth observation technology, satellite imagery with high spatial and temporal resolution has become available, such as data from European Sentinel-2 data and Chinese Gaofen data. These data could be used instead to provide more accurate estimations of LAI and LNC, resulting in improved estimation of GPC during data assimilation. Further, unmanned aerial vehicles (UAVs) have attracted considerable scientific and public attention in recent years, as they can take off at any time and are manipulated easily. With sensors attached to UAVs, high-resolution spatial and temporal images can be acquired at low cost. The data from UAVs can also be used to improve GPC prediction in future studies. Second, efforts are needed to improve the performance of the crop growth model. The existing models are unable to flawlessly simulate the growth and development of crops under the influence of genetics and external factors [4,15,53].
5. Conclusions

Studies have shown that the wheat GPC can be predicted by the data assimilation method using field sampled spectra and the CERES-Wheat model. However, field sampled remote sensing data have a limited observation area over a short period of time. Long-term observed crop biochemical and biophysical data can be used as a priori information during data assimilation, which can improve the prediction results. These scientific questions have not been discussed in previous studies. In the current study, we showed that GPC could be predicted with moderate accuracy by combing Landsat OLI data with the CERES-Wheat model. In addition, we discussed the effect of the choice of assimilation of state variables and algorithms for solving the cost function, as well as the effect of a priori knowledge on the accuracy of GPC prediction during data assimilation. Additionally, we compared GPC prediction results of data assimilation with values predicted using only the CERES-Wheat model. Using both LAI and LNC as state variables, with the SCE-UA algorithm to solve the cost function, produced optimal results. Considering the a priori knowledge, using the range of GPC values observed in the study area to constrain the output during data assimilation was an effective way to prevent ill-posed inversion. The effect of adding a priori information of average LAI and LNC during data assimilation on the accuracy of GPC prediction depended on the remote sensing estimation error of these parameters. The data assimilation method performed better than using only the CERES-Wheat model for GPC prediction, and is recommended for GPC prediction at the regional scale. In further studies, regional GPC predictions could be improved by increasing the accuracy of remote sensing prediction of state variables and improving the performance of the crop growth model.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41871344, 41301378; the National Research and Development Plan of China, grant number 2017YFD0201501; and the Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDA23100101.

Acknowledgments: The author wants to thank the Yucheng Agricultural Ecological Experimental Station for providing relevant soil and meteorological data for this study. The author also thanks Yingying Sui for her valuable assistance during the field campaign.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Xu, X.; Teng, C.; Zhao, Y.; Du, Y.; Zhao, C.; Yang, G.; Jin, X.; Song, X.; Gu, X.; Casa, R.; et al. Prediction of wheat grain protein by coupling multisource remote sensing imagery and ECMWF data. Remote Sens. 2020, 12, 1349. [CrossRef]
2. Song, X.; Wang, J.; Yang, G.; Feng, H. Winter wheat cropland GPC evaluation through remote sensing. Intell. Autom. Soft Comput. 2014, 20, 599–609. [CrossRef]
3. Asseng, S.; Milroy, S.P. Simulation of environmental and genetic effects on grain protein concentration in wheat. Eur. J. Agron. 2006, 25, 119–128. [CrossRef]
4. Orlando, F.; Marta, A.D.; Mancini, M.; Motha, R.; Qu, J.J.; Orlandini, S. Integration of remote sensing and crop modeling for the early assessment of durum wheat harvest at the field scale. Crop Sci. 2015, 55, 1280–1289. [CrossRef]
5. Asseng, S.; Bar-Tal, A.; Bowden, J.W.; Keating, B.A.; Van Herwaarden, A.; Palta, J.A.; Huth, N.I.; Probert, M.E. Simulation of grain protein content with APSIM-Nwheat. Eur. J. Agron. 2002, 16, 25–42. [CrossRef]
6. Anwar, M.R.; Takahashi, S.; Itoh, S.; Nakatsuji, T. Modeling yield and grain protein of Japanese wheat by DSSAT cropping system model. In Proceeding of Plant Growth Modeling and Applications; Hu, B.G., Jaeger, M., Eds.; Liaima, Chinese Agricultural University: Beijing, China, 2003; pp. 312–320.
7. Pan, J.; Zhu, Y.; Jiang, D.; Dai, T.; Li, Y.; Cao, W. Modeling plant nitrogen uptake and grain nitrogen accumulation in wheat. Field Crop. Res. 2006, 97, 322336. [CrossRef]
8. Basnet, B.B.; Apan, A.A.; Kelly, R.M.; Jensen, T.; Strong, W.M.; Butler, D.G. Relating satellite imagery with gain protein content. In Proceedings of the 2003 Spatial Science Institute Biennial Conference: Spatial Knowledge Without Boundaries (SSC2003); Spatial Sciences Institute: Canberra, Australia, 2003; pp. 22–27.
9. Zhao, H.; Song, X.; Yang, G.; Li, Z.; Zhang, D.; Feng, H. Monitoring of nitrogen and GPC in winter wheat based on sentinel-2A data. *Remote Sens.* 2019, 11, 1724. [CrossRef]

10. Kasampalis, D.A.; Alexandridis, T.K.; Deva, C.; Challinor, A.; Moshou, D.; Zalidis, G. Contribution of remote sensing on crop models: A review. *J. Imaging* 2018, 4, 52. [CrossRef]

11. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.; Liang, S.; Chen, Z.; Xue, J.; Wu, Y.; et al. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agric. For. Meteorol.* 2019, 276–277, 107609. [CrossRef]

12. Zhang, Z.; Li, Z.Y.; Chen, Y.; Zhang, L.; Tao, F. Improving regional wheat yield estimations by multi-step assimilating of a crop model with multi-source data. *Agric. For. Meteorol.* 2020, 290, 107993. [CrossRef]

13. Talagrand, O. Assimilation of observations, an introduction. *J. Meteorol. Soc. Jpn.* 1997, 75, 191–209. [CrossRef]

14. Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.W.; Brazile, J.; Singh, R.; Schaeman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs.* 2007, 9, 165–193. [CrossRef]

15. Jin, X.; Kumar, L.; Li, Z.; Feng, H.; Xu, X.; Yang, G.; Wang, J. A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 2018, 92, 141–152. [CrossRef]

16. Li, Z.; Jin, X.; Zhao, C.; Wang, J.; Xu, X.; Yang, G.; Li, C.; Shen, J. Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *Eur. J. Agron.* 2015, 71, 53–62. [CrossRef]

17. Thorp, K.R.; Wang, G.; West, A.; Moran, M.; Bronson, K.; White, J.; Mon, J. Estimating crop biophysical properties from remote sensing data by inverting liked radiative transfer and ecophysiological models. *Remote Sens. Environ.* 2012, 124, 224–233. [CrossRef]

18. Mass, S.J. Use of remotely sensed information in agricultural crop growth models. *Ecol. Model.* 1988, 41, 247–268. [CrossRef]

19. Thorp, K.R.; Hunsaker, D.J.; French, A.N. Assimilating leaf area index estimates from remote sensing into the simulations of a cropping systems model. *Trans. ASABE* 2010, 53, 251–262. [CrossRef]

20. Li, R.; Li, C.J.; Dong, Y.; Feng, L.; Wang, J.; Yang, X.; Pan, Y. Assimilation of remote sensing and crop model for LAI estimation based on ensemble kaiman filter. *Agric. Sci. China* 2011, 10, 1595–1602. [CrossRef]

21. Dong, Y.; Zhao, C.; Yang, G.; Chen, L.; Wang, J.; Feng, H. Integrating a very fast simulated annealing optimization algorithm for crop leaf area index variational assimilation. *Math. Comput. Model.* 2013, 58, 877–885. [CrossRef]

22. Yao, F.; Tang, Y.; Wang, P.; Zhang, J. Estimation of maize yield by using a process-based model and remote sensing data in the northeast China plain. *Phys. Chem. Earth* 2015, 87, 142–152. [CrossRef]

23. Huang, J.; Ma, H.; Su, W.; Zhang, X.; Huang, Y.; Fan, J.; Wu, W. Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, 8, 4060–4071. [CrossRef]

24. Li, Z.; Wang, J.; Xu, X.; Zhao, C.; Jin, X.; Yang, G.; Feng, H. Assimilation of two variables derived from hyperspectral data into the DSSAT-CERES model for grain yield and quality estimation. *Remote Sens.* 2015, 7, 12400–12418. [CrossRef]

25. Cui, S.; Yang, S.; Zhu, C.; Wen, N. Remote sensing of surface reflective properties: Role of regularization and a priori knowledge. *Optik* 2014, 125, 7106–7112. [CrossRef]

26. Liang, S.; Qin, J. Data Assimilation Methods for Land Surface Variable Estimation, in Advances in Land Remote Sensing: System, Modeling, Inversion and Application; Liang, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 319–339.

27. Chen, P.; Jing, Q. A comparison of two adaptive multivariate analysis methods (PLSR and ANN) for winter wheat yield forecasting using Landsat-8 OLI images. *Adv. Space Res.* 2017, 59, 987–995. [CrossRef]

28. Palosuo, T.; Kersebaum, K.C.; Angulo, C.; Hlavinka, P.; Moriondo, M.; Olesen, J.E.; Patil, R.H.; Rugel, F.; Rumbaur, C.; Takač, J.; et al. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *Eur. J. Agron.* 2011, 35, 103–114. [CrossRef]

29. Nuttall, J.G.; O’Leary, G.J.; Panozzo, J.F.; Walker, C.K.; Barlow, K.M.; Fitzgerald, G.J. Models of grain quality in wheat-A review. *Field Crop. Res.* 2017, 202, 136–145. [CrossRef]

30. Teh, C. Introduction to Mathematical Modeling of Crop Growth; Brown Walker Press: Boca Raton, FL, USA, 2006.

31. Saltelli, A.; Tarantola, S.; Chan, K.P.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 1999, 41, 39–56. [CrossRef]
32. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Carboni, J.; Gatelli, D.; Tarantola, S. *Global Sensitivity Analysis: The Primer*; John Wiley & Sons Ltd: West Sussex, UK, 2008.

33. Jin, X.; Li, Z.; Nie, C.; Xu, X.; Feng, H.; Guo, W.; Wang, J. Parameter sensitivity analysis of the AquaCrop model based on extended Fourier amplitude sensitivity under different agro-meteorological conditions and application. *Field Crop. Res.* 2018, 226, 1–15. [CrossRef]

34. Baret, F. Contribution au Suivi Radiométrique de Cultures de Céréales. Ph.D. Thesis, Université de Paris-Sud Orsay, Paris, France, 1986.

35. Casa, R.; Varella, H.; Buis, S.; Guérif, M.; De Solan, B.; Baret, F. Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *Eur. J. Agron.* 2012, 37, 1–10. [CrossRef]

36. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W.; Harlan, J.C. *Monitoring the Vernal Advancement of Retrogradation (Green Wave Effect) of Natural Vegetation*; Type III, Final Report; NASA/GSFC: Greenbelt, MD, USA, 1974; pp. 1–371.

37. Pearson, R.L.; Miller, L.D. *Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Short-Grass Prairie, Pawnee National Grasslands, Colorado*; ERIM: Ann Arbor, MI, USA, 1972.

38. Huete, A.; Justice, C.; Liu, H. Development of vegetation and soil indices for MODIS. *Remote Sens. Environ.* 1994, 49, 224–234. [CrossRef]

39. Broge, N.H.; Leblanc, E. 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.* 2001, 76, 156–172. [CrossRef]

40. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 1996, 55, 95–107. [CrossRef]

41. Gitelson, A.A.; Kaufman, Y.; Merzlyak, M. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 1996, 58, 289–298. [CrossRef]

42. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 2004, 90, 337–352. [CrossRef]

43. Wang, J.; Li, X.; Lu, L.; Fang, F. Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *Eur. J. Agron.* 2013, 49, 126–140. [CrossRef]

44. Hansen, P.M.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 2003, 86, 542–553. [CrossRef]

45. Mercier, A.; Betbeder, J.; Rapinel, S.; Jegou, N.; Baudry, J.; Hubert-Moy, L. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model. *Eur. J. Agron.* 2019, 102, 1–13. [CrossRef]

46. Shi, L.; Lin, L.; Wang, S.; Chen, R. Image fusion based on data assimilation and differential evolution algorithm. *Acta Photonica Sin.* 2010, 39, 1688–1692, (in Chinese with English abstract).

47. Janusauskaite, D.; Feiziene, D.; Feiza, V. Relationship between spring triticale physiological traits and productivity changes as affected by different N rates. *Acta Agric. Scand. Sect. B Soil Plant Sci.* 2017, 67, 534–541. [CrossRef]

48. Xing, H.; Li, Z.; Xu, X.; Feng, H.; Yang, G.; Chen, Z. Multi-assimilation methods based on AquaCrop model and remote sensing data. *Trans. Chin. Soc. Agric. Eng.* Trans. Csaes. 2017; 33, 183–192, (in Chinese with English abstract).

49. Pinnington, E.M.; Casella, E.; Dance, S.L.; Lawless, A.S.; Marison, J.L.; Nichols, N.K.; Wilkinson, M.; Quaife, T.L. Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using four-dimensional variational data assimilation. *Agric. For. Meteorol* 2016, 228–229, 299–314. [CrossRef]
52. Wang, D.W.; Wang, J.D.; Liang, S.L. Retrieving crop leaf area index by assimilation of MODIS data into a crop growth model. *Sci. China Earth Sci.* **2010**, *53*, 721–730. [CrossRef]

53. Marin, F.; Jones, J.W.; Boote, K.J. A stochastic method for crop models: Including uncertainty in a sugarcane model. *Agron. J.* **2017**, *109*, 483–495. [CrossRef]