

Article

# Improving Spring Maize Yield Estimation at Field Scale by Assimilating Time-Series HJ-1 CCD Data into the WOFOST Model Using a New Method with Fast Algorithms

Zhiqiang Cheng, Jihua Meng \* and Yiming Wang

Key Laboratory of Digital Earth, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; chengzq@radi.ac.cn (Z.C.); wangyiming19920404@163.com (Y.W.)

\* Correspondence: mengjh@radi.ac.cn; Tel.: +86-10-6486-9473

Academic Editors: Mutlu Ozdogan, Yoshio Inoue and Prasad S. Thenkabail

Received: 31 December 2015; Accepted: 22 March 2016; Published: 4 April 2016

**Abstract:** Field crop yield prediction is crucial to grain storage, agricultural field management, and national agricultural decision-making. Currently, crop models are widely used for crop yield prediction. However, they are hampered by the uncertainty or similarity of input parameters when extrapolated to field scale. Data assimilation methods that combine crop models and remote sensing are the most effective methods for field yield estimation. In this study, the World Food Studies (WOFOST) model is used to simulate the growing process of spring maize. Common assimilation methods face some difficulties due to the scarce, constant, or similar nature of the input parameters. For example, yield spatial heterogeneity simulation, coexistence of common assimilation methods and the nutrient module, and time cost are relatively important limiting factors. To address the yield simulation problems at field scale, a simple yet effective method with fast algorithms is presented for assimilating the time-series HJ-1 A/B data into the WOFOST model in order to improve the spring maize yield simulation. First, the WOFOST model is calibrated and validated to obtain the precise mean yield. Second, the time-series leaf area index (LAI) is calculated from the HJ data using an empirical regression model. Third, some fast algorithms are developed to complete assimilation. Finally, several experiments are conducted in a large farmland (Hongxing) to evaluate the yield simulation results. In general, the results indicate that the proposed method reliably improves spring maize yield estimation in terms of spatial heterogeneity simulation ability and prediction accuracy without affecting the simulation efficiency.

**Keywords:** yield simulation; prediction accuracy; spatial heterogeneity; WOFOST model; fast assimilation algorithms; remote sensing

---

## 1. Introduction

Because of its vast territory and sparse population, Northeast China is home to many big farms. Owing to the fertile soil and modern agriculture management in these big farms, this region has become the main maize planting area and contributed the maximum growth of maize yield in China, the world's second-largest maize producer (FAO, 2013). Modern agriculture comprises more than 70% of China's total maize production (National Bureau of Statistics of China, 2014). Therefore, timely and precise simulation is important for farms to optimize management and boost yield. However, most yield prediction methods depend on conventional techniques including forecasting based on agro-meteorological models or establishing a relationship between remote sensing (RS) spectral vegetation indexes (VIs) and field-measured yields. One of the main drawbacks of these methods is that they are only applicable to specific crop cultivars, crop growth stages, or geographical regions [1,2].

In contrast, a crop model can effectively address these problems owing to its comprehensive mathematical descriptions of key physical and physiological processes, simulation of soil processes, and ability to overcome issues such as abnormal weather conditions and natural disasters [3–6]. Thus, crop models have emerged as the most potent tools for yield estimation, and they have been shown to be effective for various types of crops in different countries. Several studies have confirmed that such crop models can be successfully applied to crop yield prediction at field scale [7–13]. However, they are also found to be limited under regional-scale extrapolation for estimating crop yield owing to uncertainties in input parameters or initial conditions [1,14]. In particular, these uncertainties include two different cases: parameters that are hard to obtain at pixel or field scale and parameters that change in a single growth season or over years. Obtaining additional unknown or variational input parameters is an effective way to address these problems; however, in general, it is extremely expensive and even infeasible.

Optical remote sensing techniques provide extensive and diverse spatial information on the actual growth status of a crop. For example, HJ-CCD has high revisit frequency, moderate-resolution imaging spectroradiometer (MODIS) has wide coverage, and Quickbird has high spatial resolution. Fully exploiting RS data is useful for addressing problems in many research fields, including field yield simulation [15,16]. Since the 1980s, when Steve Maas first attempted to combine RS data for modeling [17], several methods for assimilating RS data into crop models have been explored. The ensemble Kalman filter (EnKF) is used in the CERES-Wheat model [18], WOFOST model [1,5,19–21] and SWAP (soil–water–atmosphere–plant) model [22] for many different crops. Particle filtering (PF) is used to reduce uncertainty in the STICS (Simulateur multi-disciplinaire pour les Cultures Standard) model prediction [23]. Four-dimensional variational data assimilation (4DVar) is also used in the CERES (crop–environment resource synthesis)-Wheat model [24] to improve prediction accuracy. However, many factors may influence the assimilation performance. In our study, three main factors when we applied common data assimilation method (EnKF) into the model for yield simulation. Firstly, the spatial heterogeneity between or within fields was worse. Secondly, the common data assimilation method encountered problems when the nutrient module, which is used after growth simulation [11], was applied to obtain precise yields. Spatial heterogeneity and the nutrient module are both important in our study when the simulation yield is adopted for further application. At the meantime, the efficiency of simulation is another factor when the model is applied to a larger farm. Thus, such problems have to be addressed in order to obtain useful yield estimates by using crop models and RS data.

This paper presents a new method with fast algorithms to address the problems of yield simulation at field scale when using crop models and RS data. The fast algorithms aim to avoid errors in RS data and assimilate time-series HJ-1 CCD data into the calibrated and validated WOFOST model. Thus, two major objectives are achieved: improving the spatial heterogeneity simulation ability and prediction accuracy of the crop model (nutrient-limited level) for a large- or medium-sized farmland without affecting the simulation efficiency. Several experiments and analyses are carried out in Hongxing farm to evaluate the spatial heterogeneity and prediction accuracy. In addition, the simulation time is recorded. Complete details of the proposed assimilation method and analysis results are presented in the following sections.

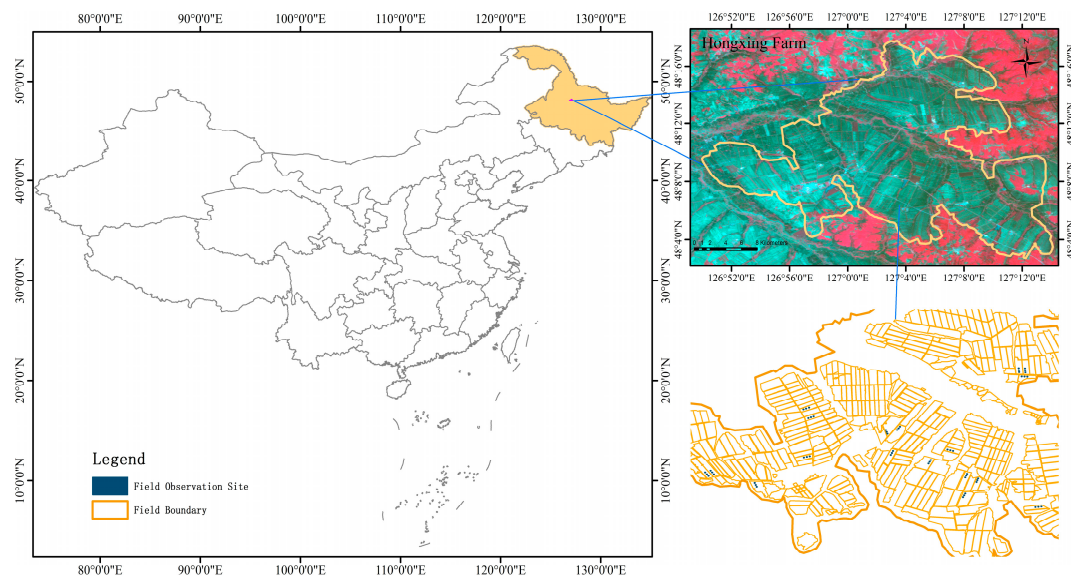
## 2. Materials and Methods

### 2.1. Study Area

The study was conducted in a large farmland, namely, Hongxing Farm (Hongxing), which is located in the central region of Heilongjiang (48°09'N, 127°03'E), Northeast China. The farm is in the mid-temperate zone, characterized by average annual precipitation of 548.8 mm and average annual accumulative temperature (>10 °C) of 2293 °C. The major crops are soybean, spring maize,

and spring wheat. In 2014, the proportion of the spring maize planting area in Hongxing was nearly 50%. The growing season in the study area extends from the beginning of May to mid-October.

The average size of the fields is more than 55 hectares and moderate-resolution remote sensing data such as HJ-CCD data with a spatial resolution of 30 m can be easily obtained. The farm management is scheduled by an independent organization. Hence, each field is usually planted with only one type of crop, and the sowing and harvest times are easily simulated. Figure 1 shows the location and field distribution of the study area.



**Figure 1.** Location of study area and the field observation sites in 2014.

## 2.2. WOFOST Model

In this study, the WOFOST model [11,25] was used to simulate the growing process of spring maize. The model, which originated from the Center of World Food Studies located in the Netherlands, is a member of the family of Wageningen crop models [5,26], and it is the central component of the Crop Growth Monitoring System (CGMS) [9]. WOFOST can simulate daily crop growth (two or more days can also be selected for the simulation step) and compute the yield at the end of the growing season. The types of yield calculated include the potential yield ( $Y_{pt}$ ), water-limited yield ( $Y_{wl}$ ), and nutrient-limited yield ( $Y_{nl}$ ).  $Y_{pt}$  is determined on the basis of meteorological conditions and crop physiological characteristics by assuming the absence of water and nutrient stress factors. In other words,  $Y_{pt}$  is the highest yield level that can be obtained in theory.  $Y_{wl}$  is derived from  $Y_{pt}$  by considering the water supply, which is influenced by rainfall, soil type, and field topography. Thus,  $Y_{wl}$  is usually lower than  $Y_{pt}$ . Finally, the model can also consider soil nutrients to obtain  $Y_{nl}$ .  $Y_{nl}$  is calculated statically on the basis of soil characteristics and  $Y_{wl}$  ( $Y_{pt}$  can also be used).

## 2.3. Field Campaign and Model Calibration

The WOFOST model has been developed mainly for European conditions and it has been used to simulate the production of the chief annual crops in Europe [27,28]. Although it has also been used for regional land evaluation, yield potential simulation, risk analysis, and yield forecasting studies in Europe, Africa, and China [9,29,30], new calibration is necessary for crop growth simulation in the present study. In general, the main input parameters are weather, soil, crop, and management parameters.

These parameters were calibrated in four ways: documentary method, farm data collection, field observation, and RS estimation. Parameters calibrated by the documentary method included

previously researched parameters, WOFOST default parameters, and some soil parameters from the 1:1,000,000 Chinese soil database (<http://www.soil.csdb.cn>).

Data collected from Hongxing farm included historical field measurement data and management data. Because the farm carries out surveys in the fields each year, we could obtain the yield and soil nutrient at field level for more than five years. The farm also requested Heilongjiang Bayi Agricultural University to investigate the soil fertility of the cultivated land, and some soil parameters could be calibrated on the basis of the corresponding results. The farm has been running a weather station for several years; it can provide daily meteorological data including mean, maximum, and minimum temperature, wind speed, vapor pressure, precipitation, humidity, and radiation data. Most meteorological data were collected at the farm level, except temperature, which could be resolved into different temperature zones by landform and physiognomy. In addition, nitrogen fertilizer data could also be obtained in this manner.

The field campaigns were carried out in both 2014 and 2015. During the entire growing season of spring maize from 11 April to 30 October 2014, the main collected data included yield, LAI, biomass, soil nutrients, and crop growth period. The observed yields in 2014 were obtained using two methods. In the first method, 19 fields were selected to be the experimental plots; three locations were selected in each field, and their average yield was considered as the yield of the field. In the second method, a group of technicians were asked to obtain the total weight after the harvester finished harvesting each field. In addition, the percentage impurity and water content were recorded. In all the yield results, the impurities were removed and the water content was converted to 25%. The observation period was 25 September–25 October. Series observations of biomass and LAI were made at fixed sites for at least one month of the entire growing season. From 25 April to 1 May, soil samples were collected for assays in order to obtain the soil parameters and basic soil nutrient data. Furthermore, crop key growth period data such as sow time, anthesis, and harvest time were collected by technicians. The soil and growth period data were obtained from the same sites as the biomass and LAI.

In 2015, some special campaigns were also carried out for the key parameters of WOFOST. Based on the parameters calibrated in 2014, the selected calibration order was: (1) growth stages; (2) biomass; and (3) yield. Four observation areas (shown in Figure 2) were selected to observe the growth stages every two (during growth stages) or five (before growth stages) days. The growth stage observation information (listed in Table 1) can be used to calculate TSUM1, TSUM2, TBASEM, and TEFFMX.

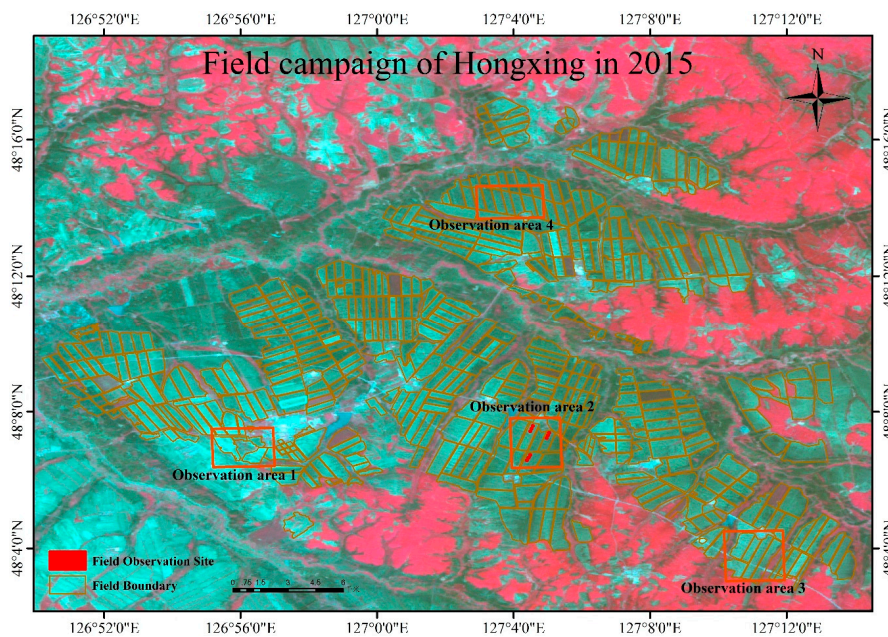


Figure 2. Observation areas of Hongxing in 2015.

**Table 1.** Key growth stages of four observation areas in 2015.

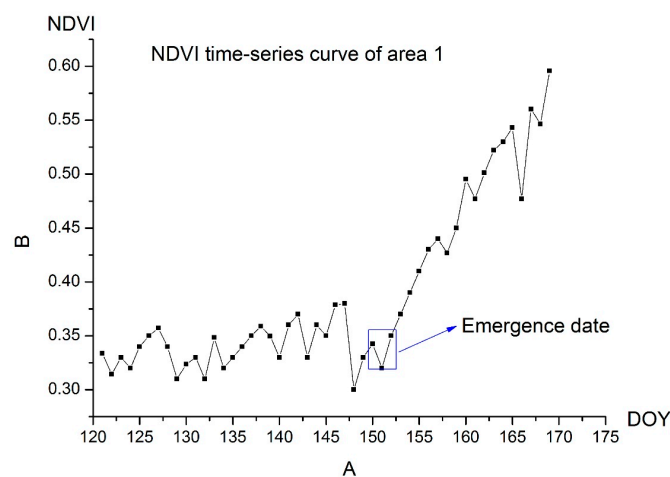
Observation Area	Growth Stages	Date (Month–Day)	DVS	Observation Area	Growth Stages	Date (Month–Day)	DVS
Area 1	First emergence time	22 May	0.04	Area 3	First emergence time	20 May	0.03
	Widespread emergence time	25 May	0.08		Widespread emergence time	26 May	0.07
	Last emergence time	27 May	0.09		Last emergence time	29 May	0.09
	First anthesis time	17 July	0.85		First anthesis time	18 July	0.87
	Widespread anthesis time	22 July	0.94		Widespread anthesis time	22 July	0.95
	Last anthesis time	27 July	1.03		Last anthesis time	31 July	1.12
	First maturity time	20 September	1.92		First maturity time	21 September	1.90
	Widespread maturity time	25 September	1.97		Widespread maturity time	25 September	1.94
	Last maturity time	28 September	1.99		Last maturity time	29 September	1.98
Area 2	First emergence time	27 May	0.09	Area 4	First emergence time	23 May	0.05
	Widespread emergence time	1 June	0.11		Widespread emergence time	28 May	0.09
	Last emergence time	4 June	0.13		Last emergence time	30 May	0.10
	First anthesis time	20 July	0.87		First anthesis time	21 July	0.88
	Widespread anthesis time	25 July	0.96		Widespread anthesis time	25 July	0.99
	Last anthesis time	30 July	1.04		Last anthesis time	1 August	1.15
	First maturity time	23 September	1.90		First maturity time	22 September	1.90
	Widespread maturity time	27 September	1.97		Widespread maturity time	28 September	1.94
	Last maturity time	30 September	1.99		Last maturity time	30 September	1.97

For biomass and yield, continuous campaigns were carried out in the four observation areas (shown in Figure 2) in order to obtain the dry matter weight of different crop portions, which could be used for calibrating the parameters of conversion efficiency of assimilates and fraction of above ground dry matter. These weight data (listed in Table 2) were collected on 24 May, 27 June, 31 July, and 30 August.

**Table 2.** Dry matter weight of observation Area 2.

Date (Month-Day)	DVS	Portion	Mean Dry Matter Weight (g/plant)
24 May	0.1	leaves	3.1
		stem	0.93
		storage organs	0.0
27 June	0.5	leaves	20.6
		stem	8.6
		storage organs	0.0
31 July	1.1	leaves	36.0
		stem	49.1
		storage organs	7.9
30 August	1.7	leaves	78.0
		stem	61.2
		storage organs	106.5

From Table 1, we can see that different areas in the farm have different growth stages. For example, the widespread emergence time in Area 2 is seven days later than that in Area 1. These necessary parameters can hardly be obtained by field campaigns at field scale; however, RS can overcome this problem. Time-series NDVI data for every day calculated from HJ-CCD, MODIS, and GF-1 were used to obtain the emergence time. The inflection point (shown in Figure 3) at which the NDVI time-series curve starts to rise was selected as the emergence time. We defined the inflection point by a simple but fast and effective algorithms. First, Savitzky–Golay filter was used to obtain trend line of NDVI time-series, and the trend line usually have six inflection points: the first one for emergence time, the second one for anthesis time, which is hard to define, the fifth one for mature time, and sixth one for harvest time. Then, a five days filter was used to define the first and fifth inflection point. For example, before the first point, the slope of trend line was closed to zero which had only small fluctuations caused by water changes and noise of RS data while the field was covered with bare soil. When the line reached this point, the crop began to emerge and the slope of trend line become larger. The five days filter used to find the change which slope began to increase. The third day was selected as the inflection point. The simulation results are listed in Table 3. From the table, we can also find that the date from RS minus nine days usually matches that from field observation. Due to the report from Hongxing, the maize cannot be harvested as soon as mature (influenced by the grain procurement); thus, the mature time was used instead of the harvest time. The maturity time can be calculated by RS [31], which are also listed in Table 3. The accuracy of emergence and harvest time both remained at a high level, with errors less than 2 days and the  $R^2$  more than 0.90; thus, we can obtain ideal emergence and harvest times using this method with daily RS NDVI data.

**Figure 3.** NDVI time-series curve of Area 1 in 2015.

**Table 3.** Widespread emergence and harvest time from field observation and remote sensing.

Area Name	Emergence Date from Field Observation (Month-Day)	Emergence Date from RS (Month-Day)	Harvest Date from Field Observation (Month-Day)	Harvest Date from RS (Month-Day)
Area 1	22 May	31 May	25 September	27 September
Area 2	1 June	12 June	27 September	28 September
Area 3	26 May	5 June	25 September	24 September
Area 4	28 May	6 June	28 September	26 September

Using the methods described above, we can calibrate the parameters of WOFOST at three different levels: meteorological (except temperature) and most crop and soil parameters at farm level, some soil parameters at field level, and some other key parameters (growth stages and LAI) at pixel level. The main crop and soil parameters are listed in Tables 4 and 5 respectively.

**Table 4.** Input crop parameters of the WOFOST model.

Parameters	Description	Values	Unit	Source
TDWI	Initial total crop dry weight	120	kg/ha	Farm data
TSUM1	Temperature sum from emergence to anthesis	890	°C.d	RS and M
TSUM2	Temperature sum from anthesis to maturity	710	°C.d	RS and M
TBASEM	Lower threshold temperature for emergence	7	°C	RS and F
TEFFMX	Maximum threshold temperature for emergence	34	°C	RS and F
CVL	Conversion efficiency of assimilates into leaf	0.65	kg/kg	Field observation
CVO	Conversion efficiency of assimilates into storage organ	0.82	kg/kg	Field observation
CVR	Conversion efficiency of assimilates into root	0.72	kg/kg	WOFOST default
CVS	Conversion efficiency of assimilates into stem	0.69	kg/kg	WOFOST default
FRTB	Fraction of total dry matter to root	0–0.40	kg/kg	Field observation
FOTB01	Fraction of above ground dry matter to storage organs (DVS = 0.1)	0	kg/kg	Field observation
FOTB05	Fraction of above ground dry matter to storage organs (DVS = 0.5)	0	kg/kg	Field observation
FOTB11	Fraction of above ground dry matter to storage organs (DVS = 1.1)	0.20	kg/kg	Field observation
FOTB17	Fraction of above ground dry matter to storage organs (DVS = 1.7)	0.73	kg/kg	Field observation
FLTB01	Fraction of above ground dry matter to leaves (DVS = 0.1)	0.77	kg/kg	Field observation
FLTB05	Fraction of above ground dry matter to leaves (DVS = 0.5)	0.69	kg/kg	Field observation
FLTB11	Fraction of above ground dry matter to leaves (DVS = 1.1)	0.25	kg/kg	Field observation
FLTB17	Fraction of above ground dry matter to leaves (DVS = 1.7)	0.19	kg/kg	Field observation
FSTB01	Fraction of above ground dry matter to stem (DVS = 0.1)	0.23	kg/kg	Field observation
FSTB05	Fraction of above ground dry matter to stem (DVS = 0.5)	0.31	kg/kg	Field observation
FSTB11	Fraction of above ground dry matter to stem (DVS = 1.1)	0.55	kg/kg	Field observation
FSTB17	Fraction of above ground dry matter to stem (DVS = 1.7)	0.08	kg/kg	Field observation
Q10	Relative change in respiration rate per 10 °C	2	-	WOFOST default
SPAN	Life span of leaves growing at 35 °C	33	day	WOFOST default
AMAX	Maximum leaf CO <sub>2</sub> assimilation rate	25–75	Kg(CO <sub>2</sub> )/ha(leaf)/h	WOFOST default
TMPFTB	Reduce factor of AMAX by temperature	0–42	—	WOFOST default
EFF	Initial light-use efficiency of single leaves	0.47	kgCO <sub>2</sub> /J/ha/hm <sup>2</sup> s	Farm's database
CFET	Correction factor transpiration rate	1	—	Farm data
RML	Relative maintenance respiration rate leaves	0.025	—	WOFOST default
RMO	Relative maintenance respiration rate storage organs	0.009	—	WOFOST default
RMR	Relative maintenance respiration rate roots	0.01	mm	WOFOST default
RMS	Relative maintenance respiration rate stems	0.015	day	WOFOST default

RS and M denotes remote sensing and meteorological data, RS and F denotes remote sensing and farm data.

**Table 5.** Input soil parameters of the WOFOST model.

Parameters	Description	Values	Unit	Source
NBASE	Mean basic soil nitrogen content	316	mg/kg	Field observation
PBASE	Mean basic phosphorus content	40	mg/kg	Field observation
KBASE	Mean basic potassium content	176	mg/kg	Field observation
NF	Quantity of nitrogen fertilizer	261.5	kg/ha	Farm data
PF	Quantity of phosphorus fertilizer	138	kg/ha	Farm data
KF	Quantity of potassium fertilizer	150.5	kg/ha	Farm data
SMTAB-1	Volumetric moisture content (pF = −1)	0.41	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SMTAB01	Volumetric moisture content (pF = 1)	0.395	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SMTAB02	Volumetric moisture content (pF = 2)	0.265	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SMTAB06	Volumetric moisture content (pF = 6)	0.084	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SMFCF	Soil moisture content at field capacity	0.295	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SMW	Soil moisture content at wilting point	0.084	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
SM0	Soil moisture content of saturated soil	0.41	cm <sup>3</sup> /cm <sup>3</sup>	Field observation
RDMCR	Maximum root depth allow by soil	2.5	Cm	Farm's database
CONTAB	10-log hydraulic conductivity(pF = 0–4.2)	−7.2–2.0	cm <sup>3</sup> /cm <sup>3</sup>	WOFOST default
K0	Hydraulic conductivity of saturated soil	10	—	WOFOST default
SOPE	Maximum percolation rate in the root zone Initially	10	cmd <sup>−1</sup>	WOFOST default
WAV	available water in total root-exploitable soil	20	cm	WOFOST default

#### 2.4. Remote Sensing Data and LAI Calculation Model

The HJ-CCD data were selected as the input RS data. HJ-CCD data were collected from two Chinese environmental remote-sensing satellites, HJ-1A and HJ-1B. The former employs a CCD camera and an infrared multi-spectral camera, whereas the latter employs a CCD camera and a hyperspectral camera. The onboard imaging systems and infrared cameras provide a global scan every two days. The two identical CCD cameras observe a broad coverage of 360 km with a spatial resolution of 30 m. They have four visible and near-infrared bands, which include B1 (430–520 nm), B2 (520–600 nm), B3 (630–690 nm), and B4 (760–900 nm). The useful data in this study were all collected by CCD cameras, and we obtained 15 images of Hongxing in 2014. Precise geometric correction, projection transformation, and atmospheric correction were performed using ERDAS, ENVI, and ArcGIS. For example, the time-series HJ-CCD data of Hongxing are listed in Table 6.

**Table 6.** Time-series HJ-CCD data of Hongxing.

Satellite Sensors	Image Date	Solar Zenith	Solar Azimuth	Sensor Azimuth	Sensor Zenith
HJ-1A CCD1	13 April 2014	45.13	318.79	−76.55	9.34
HJ-1A CCD1	29 April 2014	49.97	315.24	−75.92	13.48
HJ-1B CCD1	24 May 2014	55.22	309.54	−74.45	6.45
HJ-1B CCD2	31 May 2014	53.28	303.31	99.17	22.90
HJ-1B CCD2	12 June 2014	54.31	302.24	99.75	8.16
HJ-1A CCD2	22 June 2014	54.98	302.93	99.97	16.37
HJ-1B CCD1	7 July 2014	54.40	304.15	98.64	9.65
HJ-1B CCD1	14 July 2014	56.20	309.67	−75.08	9.90
HJ-1A CCD1	26 July 2014	54.34	312.03	−74.87	6.82
HJ-1A CCD2	19 August 2014	46.48	313.58	99.12	15.75
HJ-1A CCD2	12 September 2014	38.96	319.35	98.59	17.82
HJ-1B CCD2	18 September 2014	36.57	319.87	98.75	25.43
HJ-1A CCD1	24 September 2014	37.13	326.88	−76.43	9.95
HJ-1B CCD1	30 September 2014	34.77	327.28	−75.24	7.64
HJ-1B CCD1	4 October 2014	33.34	328.01	−71.60	2.58

Before assimilation, time-series LAI data are required. Physical-model-based methods [8,32] and empirical regression methods [33] are widely used in LAI calculation. Radiative transfer models are the main physical models for LAI simulation. Thus, for comparison, the PROSAIL model (physical-model-based method) and a simple empirical regression model (empirical regression method) were used to obtain the LAI time-series data from time-series HJ-CCD data. The LAI data were



observed for spring maize and soybean of the whole farm (shown in Figure 1) and the observed LAI data were collected on 10 June, 8 July, 25 July, 25 August, and 5 September. Accordingly, the corresponding LAI were calculated using RS images captured on 12 June, 7 July, 26 July, 19 August, and 12 September. The accuracy was analyzed on the basis of the  $R^2$ , RMSE, and F values; the results, which are listed in Table 7, showed that the accuracy of the empirical regression method is much higher than that of the PROSAIL model. Considering that the study conditions are homogeneous and that the crop type remains stable for years, the disadvantage of the empirical regression method such as poor performance due to change in time, crop type, and geographic position can be disregarded. Based on the analyses described above, the empirical regression method was selected to build a model for LAI calculation. However, the physical model is recommended when the study area is large or the crop type changes several times annually.

**Table 7.** Analysis results of PROSAIL and regression methods for LAI calculation

Time	Model	$R^2$	F	RMSE
10 June	$Y = 2.437X + 0.409$	0.33	10.80	0.118
	PROSAIL	0.13	4.13	0.216
8 July	$Y = 4.305X - 0.7002$	0.21	9.34	0.121
	PROSAIL	0.11	3.19	0.154
25 July	$Y = 5.711X - 0.247$	0.40	19.32	0.123
	PROSAIL	0.26	13.87	0.124
25 August	$Y = 21.845X - 14.212$	0.59	41.15	0.178
	PROSAIL	0.03	0.73	0.174
5 September	$Y = 1.677X + 1.701$	0.67	74.34	0.115
	PROSAIL	0.01	0.20	0.212
DVS = 0–1	$Y = 5.828X - 0.784$	0.96	980.02	0.218
DVS = 1–2	$Y = 4.564X + 0.026$	0.80	233.64	0.193

The normalized differential vegetation index (NDVI) was selected as the independent variable and the calculated LAI was selected as the dependent variable to construct an equation of linear regression. LAI values from all key growth periods of the growing season were included in the empirical regression model in order to obtain better  $R^2$ , F, and RMSE values. However, when we tried to build the model, we faced another problem: for the entire growing season, NDVI and LAI of one crop showed a similar trend of first rising then falling, whereas the rate did not remain the same. Thus, one NDVI value might match two different LAI values, one from an earlier stage and the other from a later stage of the growing season. Therefore, two models were selected to simulate LAI for earlier and later stages. Further, DVS = 1 (for anthesis) was considered as the limit of both stages because the LAI reached its peak value at that time. We also analyzed the accuracy by calculating the  $R^2$ , F, and RMSE values. The results, which are listed in Table 7, showed that the empirical regression model could calculate the LAI with higher accuracy than other methods.

### 3. Proposed Assimilation Method

Assimilation is a useful method for combining RS data and crop models in order to avoid errors when the crop model lacks the necessary parameters. LAI usually serves as the bridge for assimilation, as it can be calculated using both RS data and the crop model. In the WOFOST model, leaf simulation is an important process. The model calculates the dry matter increase of the leaves by considering  $CO_2$  assimilation, conversion, and distribution, as well as the dry matter decrease by considering maintenance, respiration, and death. In particular, death simulation is an independent process that calculates the lifespan of new and old leaves every day. The maximum lifespan of leaves is an input parameter (SPAN) of the model, and all the leaves are assumed to die when they reach the SPAN value.

Based on the biomass of the leaves, stems, and storage organs as well as their conversion coefficients (SLATB, SPA, and SSATB), the model can precisely obtain the LAI at the potential or water-limited yield for each simulation step [11]. Thus, assimilation methods, including EnKF, can be easily applied to the WOFOST model [5,18].

Furthermore, certain obstacles exist, and three of them are obvious. In the WOFOST model, the nutrient-limited yield is calculated only when the growing period is over; the nutrient-limited LAI of each step cannot be obtained. Therefore, because nutrient data are a necessary factor in the study, common assimilation methods encounter a problem that the  $Y_{nl}$  is lower than the measured results. One main reason is that the LAI obtained from RS data is usually lower than that obtained from the model without considering the effect of soil nutrients. The second problem is the spatial heterogeneity. In the study area, many input parameters such as most weather data and crop parameters are similar or constant for the entire farm, and it is difficult to obtain these parameters at field or pixel scale. The soil nutrients and fertilization are the only soil parameters that can be obtained at field scale, but common assimilation methods may face other problems when these data are used. Thus, the spatial heterogeneity at both field level and within the fields cannot meet the practical requirements. Lastly, the computational efficiency must be considered when simulation has to be performed at pixel scale. The complex algorithms of common assimilation methods entail a long simulation time. In summary, common assimilation methods face three problems: poor spatial heterogeneity simulation ability, conflict with nutrient modules, and relatively low computational efficiency. Therefore, a new assimilation method that has good spatial heterogeneity simulation ability and that can coexist well with the nutrient module while providing high computational efficiency is necessary for obtaining useful yield estimates to guide the field action.

For this purpose, a new method is presented to assimilate the RS data into the crop model. The method is simple yet effective in improving the spatial heterogeneity and prediction accuracy. The flow of the proposed method and the reasons for selecting it are shown in Figures 4 and 5 respectively.

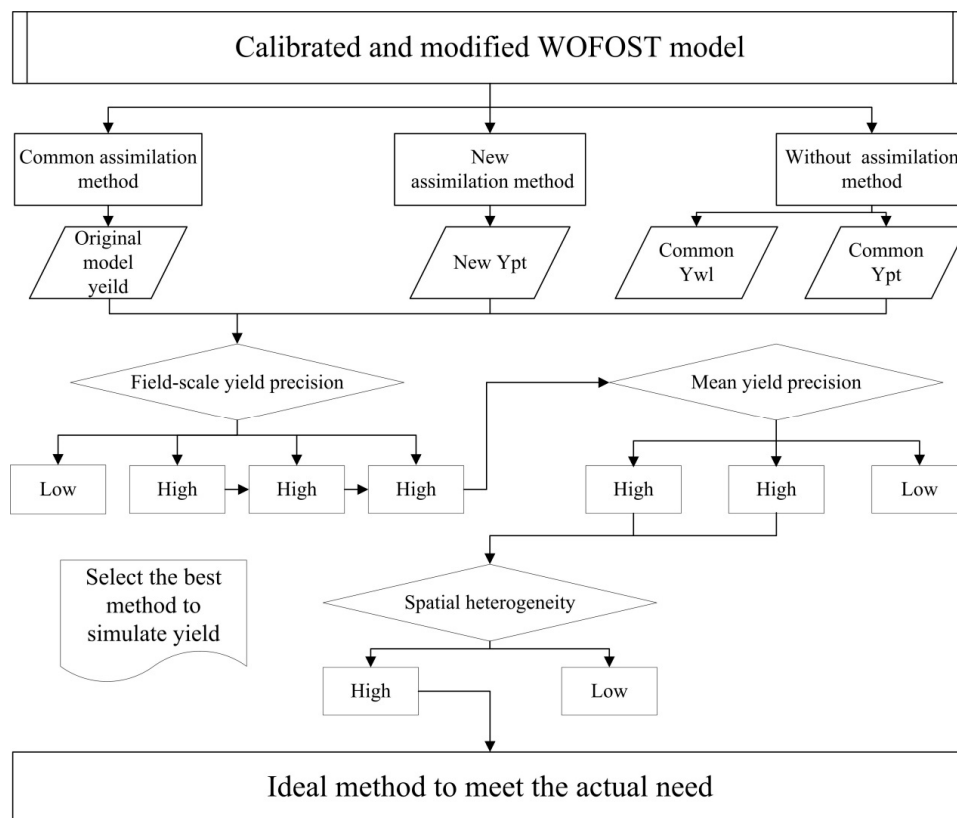


Figure 4. Process of selecting the new method.

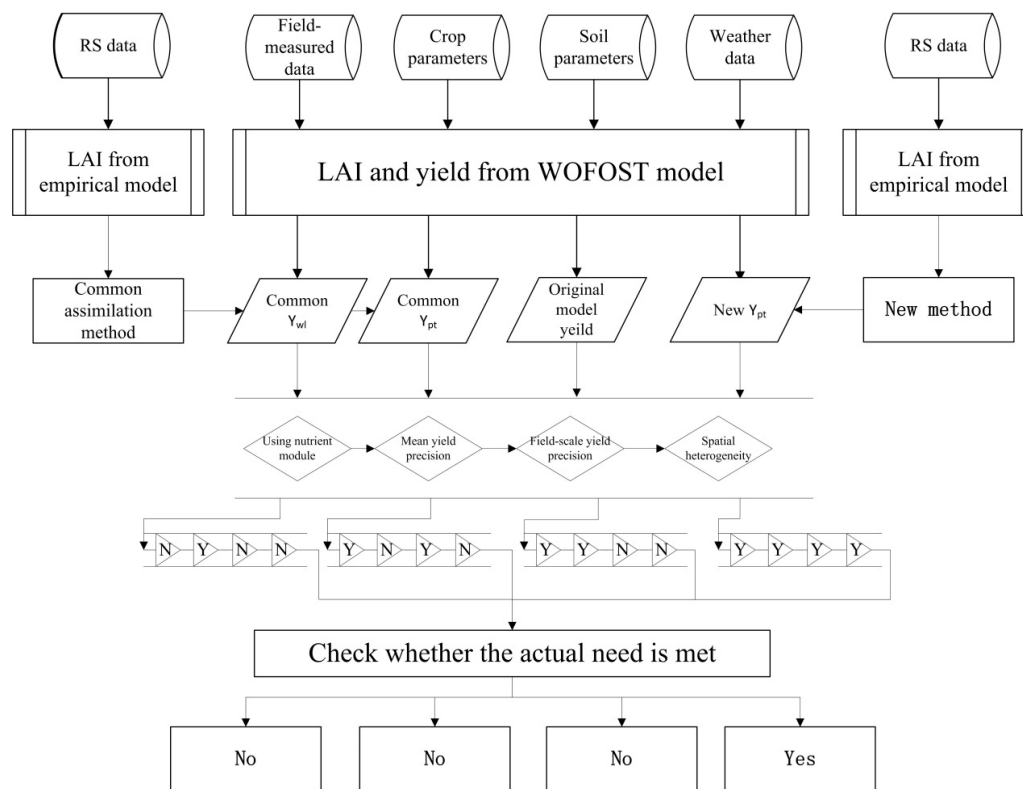


Figure 5. Reasons for selecting the new assimilation method.

The theoretical bases and primary concepts underlying the construction of the proposed method are listed below:

- (a) It is necessary to use the nutrient parameters in the WOFOST model to obtain the nutrient-limited yield in the study area; therefore, assimilation must be applied at the water-limited level to ensure accuracy of the yield result.
- (b) The spatial heterogeneity of the original model with common assimilation methods is not ideal. Ideal spatial heterogeneity, which is important for managing the field action, should be provided by the new method.
- (c) One important advantage of RS is its ability to distinguish surface features in a single image, and we should exploit it fully.
- (d) The new method should have high computational efficiency because the crop model itself is a complex system.

Based on these theoretical bases, the core algorithm of the new method is expressed by the following equation:

$$LAI = \frac{LAI_{WF} + \left(1 + \frac{LAI_{RS} - LAI_{RSM}}{LAI_{RSMX} - LAI_{RSMN}} \times a\right) \times LAI_{WFM}}{2} \tag{1}$$

where  $LAI$  is the assimilation result that will be used in the simulation of the next step,  $LAI_{WF}$  is the LAI simulation result of the previous step in WOFOST,  $LAI_{RS}$  is obtained from the empirical regression model and the corresponding time of  $LAI_{WF}$ ,  $LAI_{WFM}$  is the mean value of  $LAI_{WF}$  in the study area,  $LAI_{RSM}$  is the mean value of  $LAI_{RS}$ ,  $LAI_{RSMX}$  and  $LAI_{RSMN}$  are the predefined maximum and minimum  $LAI_{RS}$  values, respectively, and  $a$  is a coefficient that can magnify the effect of the RS data to improve the spatial heterogeneity.

The core equation includes two important parts, namely,  $WF_{info}$  and  $RS_{info}$ .  $WF_{info}$  includes the LAI information of WOFOST, obtained from each step of the model simulation. The crop model executes each simulation step at the water-limited level. Therefore,  $WF_{info}$  is equal to  $LAI_{WF}$ .  $RS_{info}$  includes the LAI information of RS, obtained from the simulation result of the empirical regression model. The value of  $RS_{info}$  is given by

$$RS_{info} = 1 + \frac{LAI_{RS} - LAI_{RSM}}{LAI_{RSMX} - LAI_{RSMN}} \times a \quad (2)$$

$RS_{info}$  is the LAI information normalized by  $LAI_{RSMX}$  and  $LAI_{RSMN}$ , which are computed as

$$LAI_{RSMX} = LAI_{RSM} + 2 \times \delta \quad (3)$$

$$LAI_{RSMN} = LAI_{RSM} - 2 \times \delta \quad (4)$$

where  $\delta$  is the standard deviation of LAI from RS.

The usage rate of the RS data is controlled by the coefficient  $a$ . The value of coefficient  $a$  was suggested to be determined by the method's performance of spatial heterogeneity. In this study, when  $a$  was set 11, we can obtain the closest value of CV compared with the observed yield (shown in the section of results). Further,  $RS_{info}$  is applied to the core equation through  $LAI_{WFM}$ , and we set two coefficients ( $b$ ,  $c$ ) to change the proportion of  $WF_{info}$  and  $RS_{info}$ . The final core equation is given by

$$LAI = b \times WF_{info} + c \times RS_{info} \times LAI_{WFM} \quad (5)$$

The values of the two coefficients depend on coefficient  $a$ , the accuracy of the WOFOST model, and the quality of RS data. Usually,  $c$  can be given a greater value if we have higher quality of RS and a lower value may be considered if we can calibrate more parameters of WOFOST. For simplicity, we can consider only  $a \times c$  to meet the spatial heterogeneity requirement. The sum of coefficients  $b$  and  $c$  is always equal to 1. When  $b$  and  $c$  are both 0.5, we can obtain the same equation as Equation (1).

Before the method can be applied to the calibrated WOFOST, two problems remain: the scale, field, distribution, and crop type of the study area and the quality of RS images. The condition of the study area is the premise for applying the new assimilation method. The study area should not be too large and the fields in this area should have a centralized distribution and uniform sowing plan; Hongxing meets these requirements. In short, the crop should have similar DVS in WOFOST. The main reason is that the spatial difference of the RS images is the most important information for the new method. Therefore, it is necessary to ensure that the spatial difference is caused only by the limiting factors and not by different growth seasons. The maize variety can influence crop phenology and growing condition and then influence the spatial difference. The crop parameters need to be recalibrated when the variety has changed. The crop growth simulation should be done respectively if there is more than one varieties of maize planting in study area. Base on the farming report of Hongxing, they have never planted two or more varieties maize for one year since the year 2008. We hence ignored the influence of maize variety to phenology and growing condition. The same climatic conditions and management measures are also required in the study area to have to ensure that the crop has similar growth seasons, especially in terms of sow time, emergence time, anthesis, and maturity time. The sub-area method is useful when the scale of the study area is too large or the field distribution is not centralized. The sub-area method can be developed on the basis of the field distribution, DEM, environment, and field management measures for different farms. Accordingly, Hongxing was divided into two different districts.

In terms of the RS data quality, the time-series HJ data makes it impossible to ensure that there are no cloudy or bad pixels. After precise geometric correction and atmospheric correction, a few bad pixels remain, especially in July and August, *i.e.*, during the rainy season in Northeast China. These pixels may significantly influence the assimilation LAI result as the coefficient  $a$  increases.

Two methods are applied before LAI is assimilated into the crop model. First, Savitzky–Golay (S-G) filtering is applied to the LAI results of RS to obtain the filtered LAI. This method can eliminate cloudy pixels using time-series data when a cloudy pixel does not occur continuously. Some bad pixels remain in the filtered LAI; they are checked and corrected using a method that is expressed by the following equation:

$$LAI_{RS} = \begin{cases} LAI_{WF}, LAI_{RS} \leq a_1 LAI_{WF} \text{ or } LAI_{RS} \leq b_1 LAI_{RSM} \\ \frac{LAI_{RS} + LAI_{WF}}{2}, LAI_{RS} > a_2 LAI_{WF} \text{ or } LAI_{RS} > b_2 LAI_{RSM} \end{cases} \quad (6)$$

where we set  $a_1 = 0.65$  and  $b_1 = 0.45$  to account for the cloud cover and its shadow, which can cause the NDVI to be much lower than usual, and  $a_2 = 1.4$  and  $b_2 = 1.8$  to account for the bad pixels, which can cause NDVI to be much higher than usual. The lower pixels do not include useful information; therefore, the LAI simulation result of WOFOST ( $LAI_{WF}$ ) is selected instead of  $LAI_{RS}$ . Although the bad pixels that cause NDVI to be higher than usual cannot be used to obtain LAI, they may include useful information. Therefore, the mean of  $LAI_{RS}$  and  $LAI_{WF}$  is selected instead of  $LAI_{RS}$ . The values of  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  were obtained through intensive analyses and tests using HJ-CCD data and LAI measured data.

#### 4. Results

The accuracy of the calibrated WOFOST model was analyzed firstly before we applied the new assimilation method to Hongxing in Northeast China. The analysis results (listed in Table 8) showed that the calibrated model can simulated phenology, LAI variation and mean yield value with high prediction accuracy. Meanwhile, the  $R^2$  and spatial heterogeneity needs to be improved and assimilation can resolve this problem.

**Table 8.** Analysis results of the calibrated WOFOST model in 2014 (kg/ha).

Name	Observed Result	Simulaton Result	$R^2$	CV
Yield	9808.20	9767.696	0.134	0.039%
LAI (DVS = 0.5)	1.78	1.84	0.783	3.01%
LAI (DVS = 1.0)	3.56	4.34	0.528	1.19%
LAI (DVS = 1.5)	3.07	4.18	0.638	1.56%
Emergence time	25 May	27 May	-	-
Anthesis time	22 July	23 July	-	-
Maturity time	20 September	24 September	-	-

Some experiments about water stress have also been done in the year 2013 and 2014. From the Hongxing production report, the year 2014 is a bumper year while the year 2013 is an off one under the influence of abnormal meteorology. Hence, the number of stress days, including wet days and dry days, were counted. The results, which are listed in Table 9, verified the production report and showed that the calibrated model can exclude the water stress influence to yield to some extent. The EnKF and new method can obtain similar  $R^2$  and mean yield both at water-limited level and nutrient limited level when the water stress was serious such as the year 2013. However, when the water stress was not serious such as 2014, the EnKF performed obviously worse than the new method for mean yield at nutrient-limited level,  $R^2$  and CV. In other words, the new method is better than EnKF especially when the water stress is not serious.

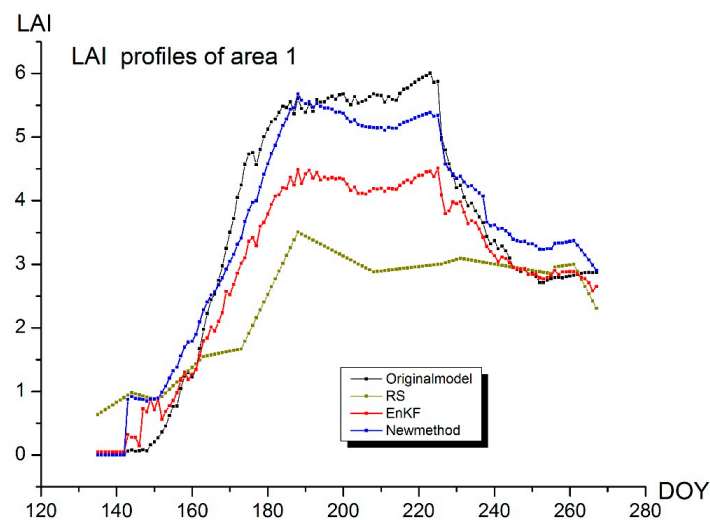
Based on the experiments and analysis results, we can obtain that the new method is useful to obtain more accurate yield simulation results, especially when water stress was not serious. Further experiments were conducted to check the yield simulation accuracy and the ability to improve spatial heterogeneity. Then, the simulation times of different methods were evaluated.

**Table 9.** Analysis results of the yields of Hongxing in 2013 and 2014 (kg/ha).

Method	Year	Stress Days	Mean	R <sup>2</sup>	CV
New method	2014	6	9710.73	0.584	6.55%
EnKF (nutrient-limited)	2014	6	9942.92	0.582	1.81%
EnKF (water-limited)	2014	6	9985.12	0.373	1.41%
New method	2013	34	7819.22	0.498	11.32%
EnKF (nutrient-limited)	2013	34	7791.39	0.456	8.01%
EnKF (water-limited)	2013	34	7821.87	0.447	7.56%

#### 4.1. Simulation Accuracy

The simulation accuracy was analyzed by LAI and yield. LAI was the key parameter that was used to execute assimilation by connecting WOFOST and RS data. The improvement of time series LAI accuracy that different assimilation methods can bring into the model's step crop growth simulation is a direct index to judge the effect of assimilation. From the calculation principle of the new method, the mean LAI value of the whole farm was not changed after assimilation. Hence, we analyzed the mean LAI value of Area 1 instead of the whole farm. The LAI profiles of original WOFOST model, RS calculation result, WOFOST with EnKF and the new method (Figure 6) showed that the two assimilation methods can improve the time series LAI accuracy to some extent. Due to the parameters of WOFOST were calibrated at water-limited level, the LAI simulation result of original model was larger than RS one. The EnKF method lower the LAI because this difference and included more WOFOST simulation information than RS one. The new method can take full advantage of the LAI information from both WOFOST simulation and RS calculation. In other words, the new method, which can be applied well at water-limited level, was a better method to improve the time-series LAI accuracy in this study.

**Figure 6.** LAI profiles of maize in Hongxing 2014.

The mean yield value and the correlation of the observed yield and the simulation result were selected as indexes to check the yield simulation accuracy. First, common assimilation methods cannot coexist well with the nutrient module of WOFOST, and the yield will be reduced if we use the nutrient module after applying the assimilation method. Therefore, we calculated the mean yield value of: (1) the EnKF method at the water-limited and nutrient-limited levels; (2) the original model; and (3) the new method as well as the observed yield. The results, which are listed in Table 10, showed that the original model could provide the most precise result, whereas the nutrient-limited yield of EnKF was the lowest. The new method provided a good result (only slightly worse than that of

the original model). Therefore, the results showed that we could obtain the ideal yield only at the water-limited level if we used the EnKF method.

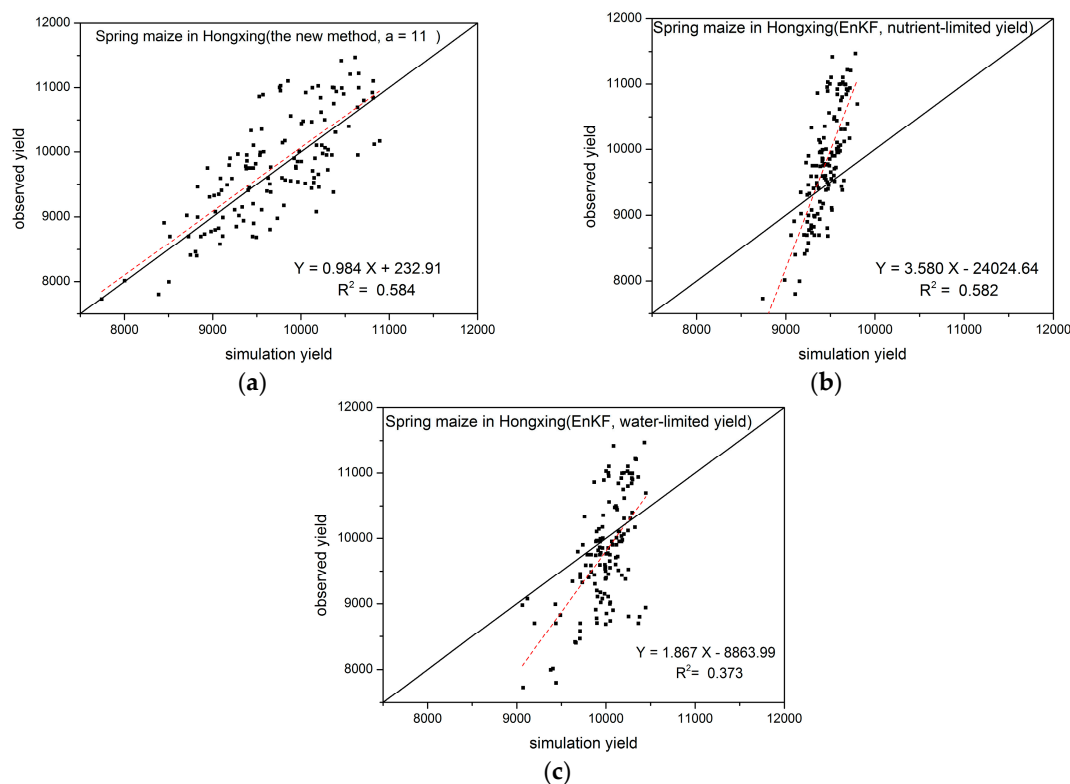
**Table 10.** Mean yield value (kg/ha) and relative error of different method.

Model	Mean Value	Relative Error
Observed yield	9808.20	0%
EnKF with nutrient module	9442.92	−3.72%
EnKF without nutrient module	9985.12	1.80%
Original model	9767.69	−0.41%
New method ( $a = 11$ )	9710.73	−0.99%

The compatibility problem can also reduce the correlation of the simulation yield and the observed yield in common assimilation methods. The  $R^2$ , F, and RMSE values were used to check this correlation. These indexes were calculated for EnKF without the nutrient module, EnKF with the nutrient module, and the new method. The results are listed in Table 11 and plotted in Figure 7. We can see that the  $R^2$ , F, and RMSE values of the new method are slightly better than those of EnKF with the nutrient module, while they are clearly better than those of EnKF without the nutrient module. Thus, we can obtain the best correlation of the simulation yield and the observed yield by using the new simulation method.

**Table 11.** Analysis results of the yields of Hongxing in 2014 (kg/ha).

Method	Quantitative Relationship	$R^2$	F	RMSE
New method	$Y = 0.984 X + 232.91$	0.584	185.11	530.81
EnKF (nutrient-limited)	$Y = 3.580 X - 24,024.64$	0.582	182.55	530.42
EnKF (water-limited)	$Y = 1.867 X - 8863.99$	0.373	79.16	652.14



**Figure 7.** Analysis results of the yields of Hongxing in 2014: (a) the new method ( $a = 11$ ); (b) EnKF method (nutrient-limited yield); (c) EnKF method (water-limited yield).

From the Tables 10 and 11 the EnKF method can obtain ideal mean yield value at water-limited level and ideal  $R^2$  at nutrient-limited level. The mean yield value becomes poor if the nutrient-limited yield is selected as simulation result and  $R^2$  become poor for water-limited result. In short, we cannot obtain ideal yield with EnKF for both  $R^2$  and mean value while the new method can. Thus, the new method is the best method for yield simulation among the three method applied in this study.

#### 4.2. Spatial Heterogeneity between Fields

We selected a variable coefficient (CV) to check the spatial heterogeneity. CV is calculated as

$$CV = \frac{SD}{Mean} \times 100\% \quad (7)$$

where  $SD$  is the standard deviation of the yield in the study area and  $Mean$  is the mean value of the yield.

Between fields and within fields were selected as different scales in order to check the spatial heterogeneity. First, the spatial heterogeneity between fields was compared using the original model, the EnKF assimilation method, and the new assimilation method. The mean yield for each field in Hongxing and their CV for different conditions were calculated. For the new method, we set the coefficients  $b$  and  $c$  to 0.5, and varied coefficient  $a$  from 1 to 20 in steps of 1.0. The results, which are listed in Table 12, showed that the spatial heterogeneity of the observed yield was much better than that of the original model and the EnKF assimilation method because the CV of the observed yield was much higher. Thus, the model's ability to simulate the spatial heterogeneity was quite poor, and common assimilation methods could not solve this problem. However, the new method could simulate the spatial difference much more effectively. Furthermore, the CV of the new method ranged from 2.95% to 6.82%, while the coefficient  $a$  ranged from 1 to 20, which was higher than that of the observed yield when  $a$  was greater than 11. We can also see that when  $a$  was greater than 10, the increase in CV slowed down, and the CV was close to that of the observed yield when  $a$  was 11. By assuming that the errors in the RS data may be amplified when the coefficient  $a$  increases,  $a = 11$  or  $a \times c = 5.5$  was selected as the ideal value to solve the spatial heterogeneity problem. For example, if we set  $c = 0.6$ , then  $a = 9$  may be the ideal value. Furthermore, coefficient  $c$  can be assigned different values according to the model simulation and RS data quality.  $a = 15$  or larger value may be selected if some key parameters were hard to be calibrated or less RS time-series images can be obtained.

Figure 8 shows similar results. The difference in the yield spatial distribution can be easily seen in the yield maps shown in Figure 8a–c. For example, in the map of the new method, the fields in the southwest had a lower yield level than others, and the observed yields supported this trend. Moreover, from Figure 8d, we can see that the amplitude of variation of the new method was larger than that of EnKF although they have similar  $R^2$  values.



Table 12. Spatial heterogeneity of yield simulation results between fields (kg/ha).

Method	Number of Fields	Summation	Mean	Variance	CV
Observed yield	134	1,314,299	9808.20	413,396.90	6.56%
Original model	134	1,308,870	9767.69	1350.03	0.39%
EnKF method	134	1,265,351	9442.92	29,253.91	1.81%
New method ( $a = 1$ )	134	1,302,662	9721.36	82,011.15	2.95%
New method ( $a = 2$ )	134	1,300,297	9703.71	141,061.91	3.87%
New method ( $a = 3$ )	134	1,299,280	9696.12	202,282.08	4.64%
New method ( $a = 4$ )	134	1,299,185	9695.41	254,321.80	5.20%
New method ( $a = 5$ )	134	1,299,440	9697.31	295,604.00	5.61%
New method ( $a = 6$ )	134	1,299,883	9700.62	327,879.85	5.90%
New method ( $a = 7$ )	134	1,300,338	9704.02	352,750.03	6.12%
New method ( $a = 8$ )	134	1,300,725	9706.90	371,514.83	6.28%
New method ( $a = 9$ )	134	1,301,017	9709.08	385,701.57	6.40%
New method ( $a = 10$ )	134	1,301,237	9710.73	396,609.74	6.49%
New method ( $a = 11$ )	134	1,301,394	9711.90	405,242.46	6.55%
New method ( $a = 12$ )	134	1,301,503	9712.71	412,132.20	6.61%
New method ( $a = 13$ )	134	1,301,575	9713.25	417,726.38	6.65%
New method ( $a = 14$ )	134	1,301,610	9713.50	422,362.01	6.69%
New method ( $a = 15$ )	134	1,301,610	9713.50	426,307.37	6.72%
New method ( $a = 16$ )	134	1,301,590	9713.36	429,652.84	6.75%
New method ( $a = 17$ )	134	1,301,552	9713.08	432,554.87	6.77%
New method ( $a = 18$ )	134	1,301,503	9712.71	435,077.34	6.79%
New method ( $a = 19$ )	134	1,301,446	9712.28	437,272.04	6.81%
New method ( $a = 20$ )	134	1,301,380	9711.79	439,230.48	6.82%

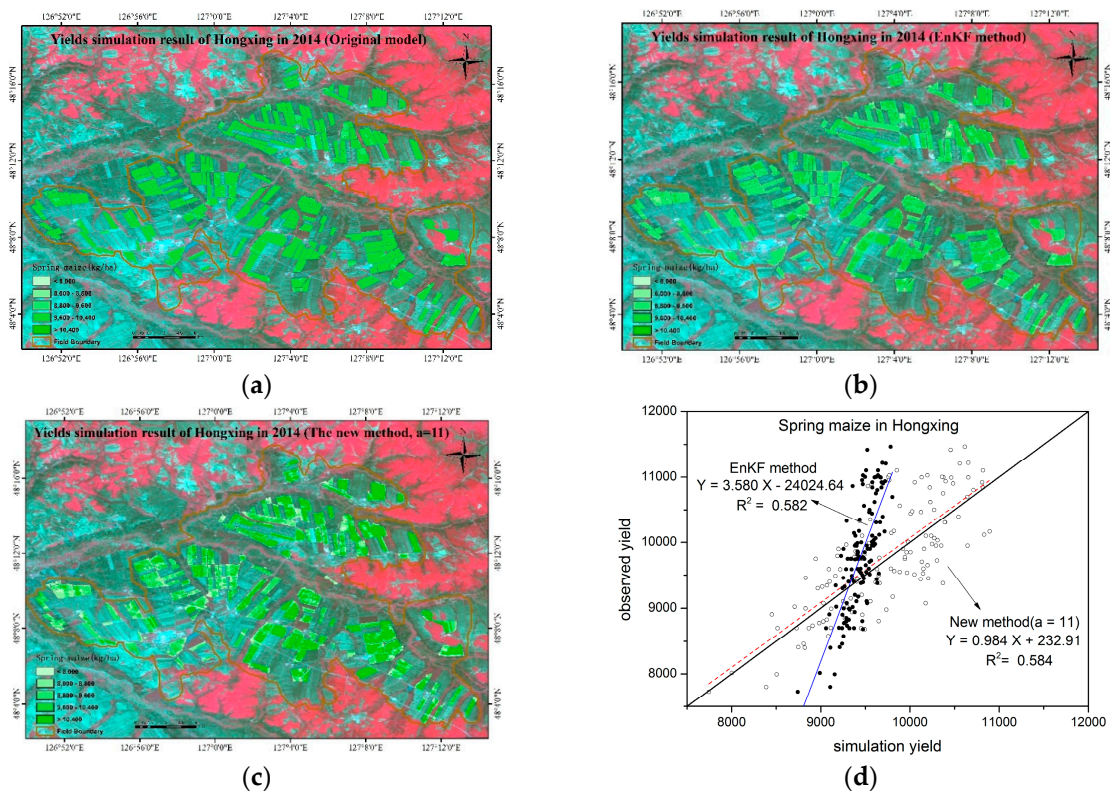
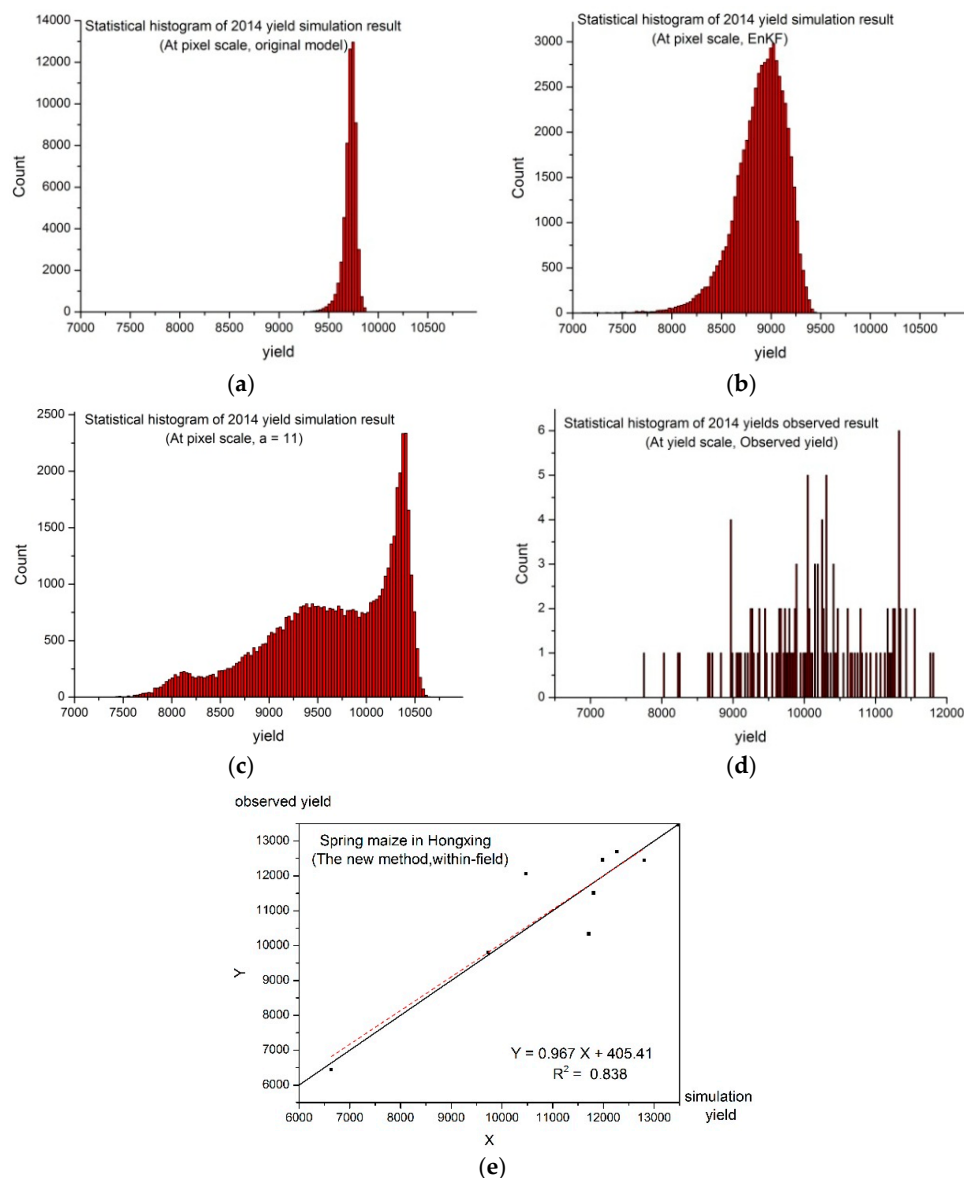


Figure 8. Yield simulation map and simulation accuracy of Hongxing in 2014: (a) simulation map of original model; (b) simulation map of EnKF; (c) simulation map of new method ( $a = 11$ ); (d) simulation accuracy result.

### 4.3. Spatial Heterogeneity within Fields

The spatial heterogeneity within fields and that between fields are equally important for guiding field management; hence, the CV for the measured points from the same or adjacent fields was calculated to check the spatial heterogeneity within fields. In this study, we selected nine points in three adjacent fields (shown in Figure 2), for which the yields for 2014 were collected. The CV and  $R^2$  (new method,  $a \times c = 5.5$ ) are shown in Table 13 and Figure 9e. The CV of new method is larger than EnKF method and original model and the  $R^2$  keep a high level. Furthermore, the histograms for all pixels in the farm were drawn, and similar information could scarcely be obtained. The histograms shown in Figure 9a–d indicate that the amplitude of variation for the new method (around 3000 kg/ha) was not only larger than that for the original method (around 600 kg/ha) but also larger than the EnKF (around 1500 kg/ha), while it was close to that for the observed yield (around 3500 kg/ha). In short, the new method can improve the model’s spatial heterogeneity simulation ability effectively at the within fields scale with high accuracy.



**Figure 9.** Statistical histograms and simulation accuracy within fields in 2014: (a) histogram of original; (b) histogram of EnKF; (c) histogram of new method ( $a = 11$ ); (d) histogram of observed yield; (e) simulation accuracy at within-field scale (new method).

**Table 13.** Spatial heterogeneity of yield simulation results within fields (kg/ha).

Method	Number of Fields	Summation	Mean	Variance	CV
Observed yield	9	99,508.55	11,056.51	3,646,241.22	2.98%
Original model	9	93,937.31	10,437.48	1,531,424.86	1.41%
EnKF method	9	92,380.36	10,264.48	1,700,142.28	1.61%
New method ( $a = 1$ )	9	94,226.02	10,469.56	1,753,887.44	1.60%
New method ( $a = 11$ )	9	101,014.58	11,223.84	3,551,981.69	2.82%
New method ( $a = 20$ )	9	100,831.12	11,203.46	4,403,890.69	3.51%

In other words, the original method could precisely obtain the mean yield but it had poor correlation with the observed yield. The EnKF could obtain high  $R^2$  and F values with the nutrient module but poor mean yield (owing to conducive growing weather in 2014, soil nutrient became the key limiting factor for yield simulation in WOFOST); the mean value could be obtained accurately but the correlation was poor if the nutrient module was not used. The correlation and mean value of the new method were better than those of the other three methods. Furthermore, the new method was the only method that could solve the spatial heterogeneity problem effectively.

#### 4.4. Simulation Time

For Hongxing, around 1.5 million pixels needed be calculated each day; thus, the simulation had to be executed more than 200 million times for the entire growing season (the growing season of spring maize is around 150 days). Therefore, the simulation time was an important factor. When the original model simulated crop growth in pixels, it took more than two hours. Considering that the model might be applied to a larger farm, the simulation time must be reduced. To improve the simulation efficiency, a different programming language was adopted: we used IDL to recode the WOFOST model; thus, the simulation could be performed with arrays. The time was reduced to 38 min by considering the LAI calculation model, whereas the time taken by the new method and EnKF method were 41 and 127 min, respectively. We can see that the new method can reduce the simulation time to less than one-third of the original, and when the model is applied to a larger farm, plenty of time will be saved. The reason for the increase in the simulation time of EnKF was that the simulation must be executed on the basis of pixels owing to the principle of EnKF, whereas the new method employs arrays.

In conclusion, the analysis results indicated that the new method could enable the WOFOST model to obtain more realistic simulation results with higher accuracy and better spatial heterogeneity without increasing the simulation time significantly.

## 5. Discussion

We developed a new assimilation method for improving spring yield simulation at field scale. The parameters of the crop model used in this study were scarce, which caused many problems for field yield simulation. In addition to the lacking input parameters, constant and similar inputs can also introduce errors into the results of this study. For example, the weather in Hongxing, to which the model is highly sensitive, may aggravate the spatial heterogeneity simulation ability of the model. Moreover, it is difficult to update all the parameters every year as some soil and crop parameters change owing to soil erosion and breeding conditions. With different assimilation methods, RS was applied to the crop model in different ways. The analysis results proved that exploiting RS data could improve the regional yield simulation; the same has been proved by many previous studies [1,3,18,34]. Some other problems appeared when applying assimilation methods to the WOFOST model for farm yield in Hongxing, and a new assimilation method was designed to resolve these problems.

One important factor is the spatial heterogeneity. The CV was calculated for the observed yields, origin model, and EnKF assimilation method. The results showed that the ability of WOFOST to simulate the spatial difference of spring maize was quiet poor, and the EnKF method played a limited role in this study. In the new method, coefficient  $a$  is introduced into the assimilation algorithms

to intensify the effect of the spatial variability obtained from the RS data. In Hongxing, the CV (6.55%,  $a = 11$ ) is close to that of the observed yield (6.56%), while it is much greater than that of the original model (0.39%) and the EnKF method (1.81%). The coefficient  $a$  can be changed depending on the application requirements for different study areas.

Another advantage of the new method is that it resolves the coexistence problem of the assimilation method and the nutrient module. This problem is caused by the structure of WOFOST: the yields are simulated at three levels ( $Y_{pt}$ ,  $Y_{wl}$ , and  $Y_{nl}$ ), and  $Y_{pt}$  or  $Y_{wl}$  must be obtained first, followed by  $Y_{nl}$ . However, RS provides the actual crop information. Common methods face problems in obtaining yields by considering both water and nutrient limits; therefore, assimilation is usually applied to obtain  $Y_{pt}$  or  $Y_{wl}$ . Further,  $Y_{nl}$  is necessary because the soil nutrient condition is a key factor affecting crop growth in Hongxing. The analysis results showed that the yield simulation accuracy is significantly influenced by the nutrient module; for the common assimilation method, the  $R^2$  value fell from 0.582 to 0.373 whereas the RMSE value increased from 530.42 to 652.14 without considering the nutrient factors. Furthermore, it is the premise for some other applications; for example, soil nutrient simulation based on the WOFOST model and RS technology needs to employ both the assimilation method and the nutrient module.

Computational efficiency is one of the main factors to be considered when designing assimilation algorithms. The new method includes three fast algorithms, one for assimilation and two for correcting bad pixels (the fast check method expressed by Equation (6) and S-G filtering). With the addition of the LAI calculation model, the simulation time is increased by around 3 min, which ensures that the crop model can be introduced to larger areas for pixel yield simulation.

In spite of these advantages, there remain certain aspects in which the new method should be improved. First, the new method requires the crop growth stages in different fields to be similar for the entire season; otherwise, uncertainty errors could be introduced into the final result. Variation in crop breed and environment might cause differences in the growth stages. Therefore, spatial scale and crop type are important factors to be considered. Although environmental differences can be addressed by the sub-area method, this method cannot be applied to the model if there are two or more breeds of crops in the same study area. Another problem is that the new method may magnify errors existing in a single RS image. Although the new method can avoid errors between images, errors within a single image may be magnified by the coefficient  $a$ . Cloudy image pixels are typical errors that can be magnified; these errors can be avoided using Equation and the S-G filter. In addition, it is necessary to avoid cloudy pixels for time-series RS data.

The new method has considerable potential in many areas such as yield calculation and nutrient simulation at field or pixel scale. It can resolve the prediction accuracy and spatial heterogeneity problems of the WOFOST model in a farm where the inputs are scarce, similar, or constant. The sub-area and cloud removal methods are key technological issues that should be resolved in future studies.

## 6. Conclusions

In this study, the WOFOST crop model was used to estimate spring maize yield at field level, and a new method was adopted to assimilate time-series HJ-CCD data into the model by means of the LAI. Although several problems need to be resolved, the new method can enhance the model's simulation accuracy and improve the model's spatial heterogeneity simulation ability. In addition, the new method can coexist with the nutrient module perfectly; thus, it is possible to apply the model to other applications such as soil nutrient simulation. The calculation principle of the new method enables us to easily select the RS time-series data without considering obstacles such as different sensors. Furthermore, the high computational efficiency of the new method ensures a relatively short simulation time for large areas. In conclusion, the new method is a simple yet effective approach for addressing the problems of the WOFOST model, especially the prediction accuracy and spatial heterogeneity.

**Acknowledgments:** This research was supported by the Science and Technology Service Network Initiative (STS) Project of CAS, No. KFJ-EW-ST-069; the National Natural Science Foundation of China, No.41171331; and the National High-Tech Research and Development Program (863 Program), No. 2013AA12A302. We are grateful to DLO Win and Staring Centre, Wageningen, for providing the technical documentation and source code of WOFOST. We also thank the China Centre for Resources Satellite Data and Application (CRESDA) for providing the HJ-1 data and Hongxing farm for providing their database to calibrate WOFOST parameters as well as the experimental field and laboratory for field campaigns.

**Author Contributions:** Jihua Meng and Zhiqiang Cheng conceived and designed the experiments; Zhiqiang Cheng contributed the new simulation method, performed the experiments and analyzed the data; Yiming Wang offered the EnKF method and collected the field data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Huang, J.; Tina, L.; Liang, S.; Ma, H.; Becker-Reshef, I.; Huang, Y.; Su, W.; Zhang, X.; Zhu, D.; Wu, W. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* **2015**, *204*, 106–121. [[CrossRef](#)]
- Fang, H.L.; Liang, S.L.; Hoogenboom, G. Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation. *Int. J. Remote Sens.* **2011**, *32*, 1039–1065. [[CrossRef](#)]
- Gerakis, A.; Ritchie, J.T. Simulation of atrazine leaching in relation to water-table management using the CERES model. *J. Environ. Manag.* **1998**, *52*, 241–258. [[CrossRef](#)]
- Soler, C.M.T.; Sentelhas, P.C.; Hoogenboom, G. Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. *Eur. J. Agron.* **2007**, *27*, 165–177. [[CrossRef](#)]
- Ma, H.; Huang, J.; Zhu, D.; Liu, J.; Su, W.; Zhang, C.; Fan, J. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST-ACRM model with Ensemble Kalman Filter. *Math. Comput. Model.* **2013**, *58*, 759–770. [[CrossRef](#)]
- Abrahamsen, P.; Hansen, S. Daisy: An open soil-crop-atmosphere system model. *Environ. Model. Softw.* **2000**, *15*, 313–330. [[CrossRef](#)]
- Akponikpè, P.B.I.; Gérard, B.; Michels, K.; Biélders, C. Use of the APSIM model in long term simulation to support decision making regarding nitrogen management for pearl millet in the Sahel. *Eur. J. Agron.* **2010**, *32*, 144–154. [[CrossRef](#)]
- Propastin, P.; Erasmi, S. A physically based approach to model LAI from MODIS 250 m data in a tropical region. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 47–59. [[CrossRef](#)]
- Boogaard, H.; Wolf, J.; Supit, I.; Niemeier, S.; van Ittersum, M. A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. *Field Crops Res.* **2013**, *143*, 130–142. [[CrossRef](#)]
- Ji, J.; Cai, H.; He, J.; Wang, H. Performance evaluation of CERES-Wheat model in Guanzhong Plain of Northwest China. *Agric. Water Manag.* **2014**, *144*, 1–10. [[CrossRef](#)]
- Boogaard, H.; van Diepen, C.A.; Rötter, R.P.; Cabrera, J.M.C.A.; Laar, H.H.V. *User's Guide for the WOFOST 7.1 Crop Growth Simulation Model and WOFOST Control Center 1.5*; DLO Winand Staring Centre: Wageningen, The Netherlands, 1998.
- Liu, H.; Yang, J.; He, P.; Bai, Y.; Jin, J.; Drury, C.F.; Zhu, Y.; Yang, X.; Li, W.; Xie, J.; Yang, J.; Hoogenboom, G. Optimizing parameters of CSM-CERES-Maize model to improve simulation performance of maize growth and nitrogen uptake in northeast China. *J. Integr. Agric.* **2012**, *11*, 1898–1913. [[CrossRef](#)]
- Mabhaudhi, T.; Modi, A.T.; Beletse, Y.G. Parameterisation and evaluation of the FAO-AquaCrop model for a South African taro (*Colocasia esculenta* L. Schott) landrace. *Agric. For. Meteorol.* **2014**, *192–193*, 132–139. [[CrossRef](#)]
- Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.W.; Brazile, J.; Singh, R.; Schaepman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 165–193. [[CrossRef](#)]
- Marie, L.; Martine, G. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric. Ecosyst. Environ.* **2005**, *111*, 321–339.

16. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [[CrossRef](#)]
17. Maas, S.J. Using satellite data to improve model estimates of crop yield. *Agron. J.* **1988**, *80*, 655–662. [[CrossRef](#)]
18. Li, R.; Li, C.-J.; Dong, Y.-Y.; Liu, F.; Wang, J.-H.; Yang, X.-D.; Pan, Y.-C. Assimilation of remote sensing and crop model for LAI estimation based on Ensemble Kalman Filter. *Agric. Sci. China* **2011**, *10*, 1595–1602. [[CrossRef](#)]
19. Li, Y.; Zhou, Q.; Zhou, J.; Zhang, G.; Chen, C.; Wang, J. Assimilating remote sensing information into a coupled hydrology-crop model to estimate regional maize yield in arid regions. *Ecol. Model.* **2014**, *291*, 15–27. [[CrossRef](#)]
20. Zhao, Y.; Chen, S.; Sheng, S. Assimilating remote sensing information with crop model using ensemble Kalman filter for improving LAI monitoring and yield estimation. *Ecol. Model.* **2013**, *270*, 30–42. [[CrossRef](#)]
21. Huang, J.; Fernando, S.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Zhang, X.; Fan, J.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to estimate regional winter wheat yield. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [[CrossRef](#)]
22. 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 to estimate winter wheat yield. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4060–4071. [[CrossRef](#)]
23. Chen, Y.; Cournède, P.-H. Data assimilation to reduce uncertainty of crop model prediction with convolution particle filtering. *Ecol. Model.* **2014**, *290*, 165–177. [[CrossRef](#)]
24. 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](#)]
25. Van Diepen, C.A.; Wolf, J.; Vankeulen, H.; Rappoldt, C. WOFOST—A simulation model of crop production. *Soil Use Manag.* **1989**, *5*, 16–24. [[CrossRef](#)]
26. Van Ittersum, M.K.; Leffelaar, P.A.; van Keulen, H.; Kropff, M.J.; Bastiaans, L.; Goudriaan, J. On approaches and applications of the Wageningen crop models. *Eur. J. Agron.* **2003**, *18*, 201–234. [[CrossRef](#)]
27. Supit, I.; van Diepen, C.A.; de Wit, A.J.W.; Kabat, P.; Baruth, B.; Ludwig, F. Recent changes in the climatic yield potential of various crops in Europe. *Agric. Syst.* **2010**, *103*, 683–694. [[CrossRef](#)]
28. Supit, I.; van Diepen, C.A.; de Wit, A.J.W.; Wolf, J.; Kabat, P.; Baruth, B.; Ludwig, F. Assessing climate change effects on European crop yields using the crop growth monitoring system and a weather generator. *Agric. For. Meteorol.* **2012**, *164*, 96–111. [[CrossRef](#)]
29. Hengsdijk, H.; Meijerink, G.W.; Mosugu, M.E. Modeling the effect of three soil and water conservation practices in Tigray, Ethiopia. *Agric. Ecosyst. Environ.* **2005**, *105*, 29–40. [[CrossRef](#)]
30. Lu, C.; Fan, L. Winter wheat yield potentials and yield gaps in the North China Plain. *Field Crops Res.* **2013**, *143*, 98–105. [[CrossRef](#)]
31. Meng, J.; Xu, J.; You, X. Optimizing soybean harvest date using HJ-1 satellite imagery. *Precis. Agric.* **2015**, *16*, 164–179. [[CrossRef](#)]
32. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604. [[CrossRef](#)]
33. Berterretche, M.; Hudak, A.T.; Cohen, W.B.; Maiersperger, T.K.; Gower, S.T.; Dungan, J. Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sens. Environ.* **2005**, *96*, 49–61. [[CrossRef](#)]
34. De Wit, A.; Duveiller, G.; Defourny, P. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agric. For. Meteorol.* **2012**, *164*, 39–52. [[CrossRef](#)]

