The Improved Winter Wheat Yield Estimation by Assimilating GLASS LAI Into a Crop Growth Model With the Proposed Bayesian Posterior-Based Ensemble Kalman Filter

Hai Huang, Jianxi Huang, Yantong Wu, Wen Zhuo, Jianjian Song, Xuecao Li, Li Li, Wei Su, Han Ma, and Shunlin Liang, *Fellow, IEEE*

Abstract— Data assimilation has been demonstrated as the potential crop yield estimation approach. Accurate quantification of model and observation errors is the key to determining the success of a data assimilation system. However, the crop growth model error is not fully taken into account in most of the previous studies. The objective of this study is to better quantify the model uncertainty in the data assimilation system. First, we calibrated a crop growth model and inferred its posterior uncertainty based on the Global LAnd Surface Satellite (GLASS) 250-m leaf area index (LAI) product, regional statistical data, station observations, and field measurements with a Markov chain Monte Carlo (MCMC) method. Second, the model posterior uncertainty was used in the ensemble Kalman filter (EnKF) algorithm to better characterize the ensemble distribution of model errors. Our results indicated that the proposed Bayesian posterior-based EnKF can improve the accuracy of winter wheat yield estimation at both the point scale (the coefficient of determination R^2 value increasing from 0.06 to 0.41, the mean absolute percentage error (MAPE) value decreasing from 12.65% to 7.82%, and the root-mean-square error (RMSE) value decreasing from 987 to 688 kg·ha^{-1}) and the regional scale (R^2 value from 0.30 to 0.57, MAPE value from 19.67% to 10.13%, and RMSE value from 1275 to 695 kg·ha^{-1}) compared with the open-loop estimation. Our analysis also indicated that the Bayesian posterior-based EnKF can perform better compared to the standard Gaussian perturbation-based EnKF. The proposed framework provides an important reference for crop yield estimation at the regional scale in similar agricultural landscapes worldwide.

Hai Huang and Yantong Wu are with the College of Land Science and Technology, China Agricultural University, Beijing 100083, China (e-mail: haihuang@cau.edu.cn; yantongwu@cau.edu.cn).

Jianxi Huang, Xuecao Li, Li Li, and Wei Su are with the College of Land Science and Technology, China Agricultural University, Beijing 100083, China, and also with the Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100083, China (e-mail: jxhuang@cau.edu.cn; xuecaoli@cau.edu.cn; lilixch@cau.edu; suwei@cau.edu.cn).

Wen Zhuo is with the State Key Laboratory of Sever Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China (e-mail: zhuowen1992@foxmail.com).

Jianjian Song is with the Changwang School of Honors, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: 3068879316@qq.com).

Han Ma and Shunlin Liang are with the Department of Geography, The University of Hong Kong, Hong Kong, China (e-mail: mahan@hku.hk; shunlin@hku.hk).

Digital Object Identifier 10.1109/TGRS.2023.3259742

I. INTRODUCTION

REGIONAL crop yield estimation and forecasting are essential for agricultural production activities and food security. Especially in recent years, with the frequent occurrence of extreme weather events, regional political turmoil and military conflicts, and the COVID-19 pandemic impact on the food system, the food crisis threats the nutrition of millions of people around the world [1]. Therefore, timely and accurate crop yield monitoring over large spatial regions can help guide farm production, food processing, transport, and logistics, thus enabling the necessary measures to address the challenge in food supply chains [2], [3], [4].

Remote sensing data can provide regional biophysical information on crop growth rapidly and frequently. The data assimilation method, combining the ability to acquire large-scale information from remote sensing and the dynamic mechanism process of crop growth models, has been a promising way to estimate regional crop yield [5]. There are three main types of data assimilation methods: variational, Kalman filtering, and Bayesian sampling-based approaches [5]. The variational method reinitializes the input parameters or initial state of the crop growth model by minimizing the difference of state variables from remote sensing and model forecasting, which are usually weighted by their inverse uncertainty [6], [7]. In variational algorithms, information from the specific assimilation temporal window needs to be used in the data assimilation system [8]. However, Kalman filters estimate a weighted average of the system state and update it at every time step from the model forecasting state and the new remote sensing observation [9], [10], [11]. The ensemble Kalman filter (EnKF) applies an ensemble of model states to represent the error statistics of the model estimate and has proven to efficiently handle strong nonlinear dynamics and large state spaces [12]. Bayesian Monte Carlo-based methods, e.g., the Markov chain Monte Carlo (MCMC) and particle filter (PF) algorithms [13], can infer the posterior distributions of the model parameters and model predictions based on sampling methods especially when the model is nonlinear and when the noises are not Gaussian.

Data assimilation has effectively improved yield estimation for different scales by integrating remote sensing information with crop growth models. Wu et al. [8] proposed a new algorithm named a variable time window and four-dimensional extension-based Ensemble Square Root Filter (VW-4-DEnSRF) combining the advantages of 4-D variational (4-DVar) and ensemble filtering methods and achieved good performances for winter wheat yield estimation at both the single point level and county level with root-mean-square error (RMSE) of 801.4 and 416.7 kg·ha^{-1}. Kang and Özdoğan [14] addressed that accurate crop model calibration is significant for data assimilation system and proposed a hierarchical data assimilation framework enabling maize yield estimation at field levels across large areas with the accuracy of 79%–91% by calibrating the crop model with MCMC method at the county scale before applying the EnKF algorithm for field/pixel level. Ziliani et al. [13] used a PF scheme to integrate the high spatiotemporal resolution satellite data into agricultural production systems simulator (APSIM) model for early season within-field maize yield prediction with high accuracy (relative RMSE of 12%). Overall, the coupling of remotely sensed information with crop growth models can be done in various ways, such as forcing the model to use remote sensing-based variables [15], [16], recalibrating the model parameters [7], [17], and updating the model state variables based on satellite observation [10], [11]. While representative studies that have been published are well-established theoretically and have shown promising results in region-specific applications, they often lack a generalized scheme for crop model calibration or error quantification. For instance, some studies rely on a subjective Gaussian noise to set the model error/uncertainty in assimilated systems. This hinders the application of these methods to different crop growth models, different remote sensing data, and other different crop growth conditions as the model parameters and uncertainties can vary widely.

The uncertainty in data assimilation systems usually arises from three aspects: model structure, model parameters, and remote sensing observations [5]. The model structure's uncertainty is difficult to be estimated quantitatively and separately. Therefore, the model parameter calibration and uncertainty analysis are a partial compensation for the uncertainty in the model structure and other model parameters that are not included in calibration [18]. The uncertainty in remote sensing observation can be estimated by comparing a limited dataset of in situ measurements with the remotely sensed estimates or provided by the satellite products established under general guidelines. In addition, optimized observation errors can be achieved via improved algorithms within the data assimilation system [19], [20], [21].

It is well-recognized that the quantitative uncertainty assessment of the crop growth model is essential for an agricultural data assimilation application [5], [22]. However, previous studies often set model uncertainty with an experience-based Gaussian perturbation based on local calibration [23], [24], leading to potential subjective errors. In addition, the uncertainties of crop growth models given in previous studies based on empirical or representative observation sets are

spatially identical, i.e., the model errors are considered to be spatially homogeneous, which may not be reasonable for model applications in large regions. Remote sensing data help quantify the uncertainty of crop growth models spatially, better express the spatial heterogeneity of model errors, and thus improve the accuracy of the data assimilation system. Theoretically, the state-of-the-art model calibration methods represented by MCMC can be perfectly combined with data assimilation theory under the Bayesian framework to provide the uncertainty distribution of model parameters and their corresponding predictions [5]. Toshichika et al. [25] calibrated a large-scale process-based model for rice simulation with an MCMC method and selected the simulated mean values from the 95% confidence interval of the posterior distribution for model validation and uncertainty assessment. Ran et al. [26] and Sexton et al. [27] showed the application framework with the AquaCrop model and demonstrated that the MCMC method helps quantify the uncertainty of the crop growth model. Although the MCMC method has been applied in the crop model calibration for parameter posterior inference [14], the uncertainty information of Bayesian posterior of model parameters has not been adequately taken into account in data assimilation algorithms. A growing number of recent studies calibrate crop growth models based on the advanced Bayesian method, e.g., various MCMC algorithms, in data assimilation systems [28]. However, they have not fully demonstrated the effective incorporation of crop growth model uncertainty into assimilation algorithms and, more importantly, its improvement in the performance of data assimilation systems.

This study introduces the Bayesian posterior uncertain information from MCMC calibration into an EnKF data assimilation framework and validates it by assimilating remote sensing leaf area index (LAI) into the WOrld FOod STudies (WOFOST) model for winter wheat yield estimation. The model uncertainty obtained by the MCMC algorithm is expressed in the form of an ensemble, which can be innately combined with the EnKF algorithm without introducing approximation error. The objective of this study is as follows:

1) to calibrate the WOFOST model and infer its posterior uncertainty based on remote sensing data, regional statistical data, station observations, and field measurements with an MCMC method;

2) couple the Bayesian posterior uncertainty with the EnKF algorithm, and then assimilate the remote sensing data into the WOFOST model at the single-point and regional scale for yield estimation;

3) to evaluate whether the proposed Bayesian posterior-based EnKF approach could improve the winter wheat yield estimation accuracy at both the scale and regional scales compared with the standard EnKF algorithm with the model ensemble generated from Gaussian perturbation.

II. DATA PREPARATION AND PREPROCESSING

A. Study Area

As one of the major grain-producing provinces, Henan produces about one-tenth of China's total grain. In particular,
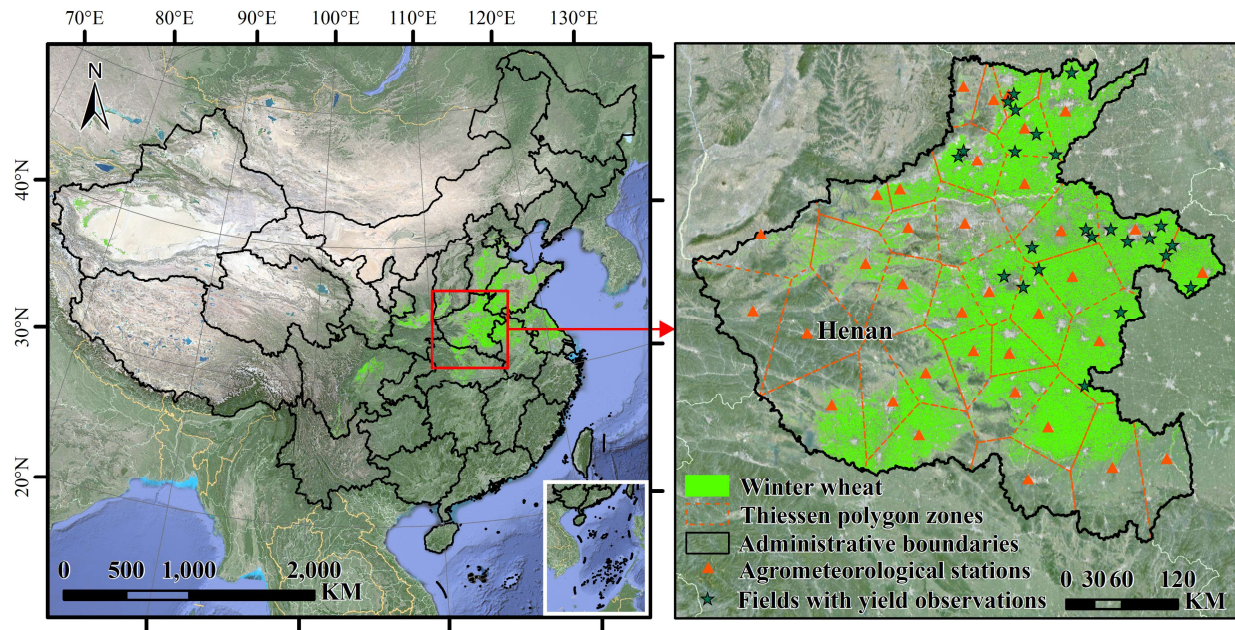
Fig. 1. Study area of Henan, China. The Thiessen polygon zones within Henan are constructed based on agrometeorological stations.

Henan is China's most crucial wheat-producing area, accounting for a quarter of the total planting area and wheat production. Henan is located in the central part of China (31.38°N–36.37°N, 110.35°E–116.65°E). The terrain is high in the West and low in the East, with mountainous areas in the West, Northwest, and Southeast, plains in the central and eastern parts, and basins in the Southwest (Fig. 1). As a whole, it belongs to the continental monsoon climate that transitions from the north subtropical zone to the warm temperate zone, with pronounced seasonal temperature differences. Winter wheat in Henan is usually sown in October, heads from late April to early May, and matures from late May to early June. The winter wheat mask data used in this study, as shown in Fig. 1, are generated by Huang et al. [29] based on the time series of temporal and spectral features of Sentinel-2 data with a spatial resolution of 10 m.

B. Remote Sensing Data

Global LAnd Surface Satellite (GLASS) 250-m LAI product (version 6) [30] was used in this study for crop growth model calibration and data assimilation. The dataset is a spatiotemporal continuous LAI product generated from MODIS data using the bidirectional long short-term memory (LSTM) deep learning model, with a temporal resolution of eight days and the spatial resolution of 250 m from 2000 to 2021. We mainly use the GLASS LAI data within the growth period and the final yield data of the previous year to calibrate the crop growth model and then use the next year's LAI product for data assimilation and yield estimation. Specifically, only a subset of time-series GLASS LAI during the growth period was used for calibration or data assimilation. We dynamically selected the previous date corresponding to the time-series GLASS LAI peak date as the starting point and set the end date to DOY 137. For most cases, six or seven LAI observations per year were used, corresponding to dates from late March or early April to late May. This is to avoid poor impacts on crop growth

model calibration or data assimilation introduced by relatively underestimated remote sensing LAI in the early growth period of winter wheat.

C. Weather Data

Agrometeorological indicators from 1979 to present derived from reanalysis (AgERA5) is a global weather dataset tailored for agriculture, and it provides daily surface meteorological data with 0.1° × 0.1° spatial resolution for the period from 1979 to the present as input for agriculture and agroecological studies [31]. In this study, 10-m wind speed (m·s^{-1}), 2-m temperature (K), precipitation flux (mm·day^{-1}), solar radiation flux (J·m^{-2}·day^{-1}), and vapor pressure (hPa) were extracted from the dataset to generate the meteorological input data for the crop growth model.

D. Field Observations

Observations of the phenological development of winter wheat from 2017 to 2021 were collected from 35 agrometeorological stations in Henan (Fig. 1), including the date of emergence, anthesis, and maturity. This data were used for calibrating the parameters related to the development stage (DVS) simulation.

During the maturity period of winter wheat in 2020 and 2021, 24 fields in northeastern Henan were selected for two consecutive years of yield measurement (Fig. 1). Five quadrats were arranged in each field (Fig. 2). For each quadrat, the average plant density and the average number of ears per plant were first measured. Then, 20 wheat ears were removed for threshing, followed by counting the number of grains, weighing, and measuring moisture content. Finally, the yield was estimated according to the following equation:

$$Y = N_{\text{ear}} \times N_{\text{grain}} \times \frac{M_{1000}}{1000} \times (1 - C) \times 0.85 \qquad (1)$$

where Y is the estimated yield without moisture, N_{ear} is the number of ears per unit area calculated from plant density
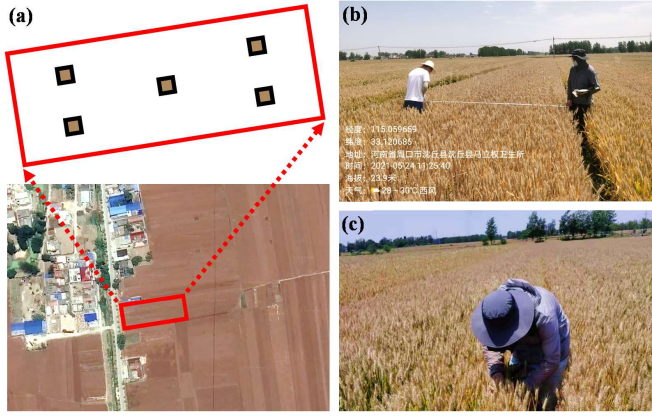
Fig. 2. Field-level winter wheat yield measurements: (a) layout of five quadrats in the field; (b) plant density measurements; and (c) picking winter wheat ears.

and the number of ears per plant, N_{grain} is the number of grains per ear, M_{1000} is the 1000-grain weight, C is the water content in weight, and 0.85 is an empirical ratio of actual harvestable yield to theoretical yield. In this study, field yield was considered the median of the estimated yields of the five quadrats to avoid the influence of extreme values due to measurement errors.

E. County Statistics

The county statistical yield of Henan from 2007 to 2020 was from Henan Survey Yearbook (http://hazd.stats.gov.cn/dcsj/list-226.html, accessed on May 1, 2022). The county-level yield was estimated by dividing the total production by planting area, and all data were calculated in units of kg·ha^{-1}. To be consistent with the simulation results, all yield data were converted to dry weight by the standard moisture content of 12.5%.

III. METHODOLOGY

We assimilate the GLASS LAI data into the WOFOST model with the Bayesian posterior-based EnKF framework at field-level point and county-level regional scales, respectively. For the point scale, field-measured yields and the field-corresponded GLASS LAI data of 2020 were used for calibration and model uncertainty assessment based on an MCMC method. Then, the corresponding GLASS LAI data of 2021 were assimilated into the calibrated WOFOST model for yield estimation. For the county-level regional scale, first, the GLASS LAI data were calculated zonally at county boundaries and the 10-km meteorological grids to obtain the mean and standard deviation, respectively. Then, the county-level statistical yield and the calculated county-level GLASS LAI information for 2017–2019 were used for calibration and model uncertainty assessment based on an MCMC method. Finally, the calculated 10-km GLASS LAI information of 2018–2020 was assimilated into the calibrated WOFOST model for yield estimation (Fig. 3).

A. WOFOST Model

The WOFOST model is a mechanistic process-based model that describes crop growth considering crop phenology development, leaf development, light interception, CO_2 assimilation, root growth, transpiration, respiration, partitioning of

assimilates to the various organs, and dry matter formation [32]. Winter wheat yield is expressed as the total-level dry weight of storage organs (TWSO) in the WOFOST model. TWSO represents the weight of grains with zero moisture content. This study implements the WOFOST model in the Python environment based on the python crop simulation environment (PCSE) package (https://github.com/ajwdewit/pcse.git, accessed on March 22, 2021).

In this study, ten parameters [emergence date (IDEM), the thermal time from emergence to anthesis (TSUM1), the thermal time from anthesis to maturity (TSUM2), initial total crop dry weight (TDWI), the life span of leaves growing at an average temperature of 35 °C (SPAN), specific leaf area as a function of the development stage (SLATB), maximum CO_2 assimilation rate as a function of the development stage (AMAXTB), the fraction of total dry matter to leaves as a function of the development stage (FLTB), the fraction of total dry matter to storage organs as a function of the development stage (FOTB), and the fraction of total dry matter to stems as a function of the development stage (FSTB)] in the WOFOST model are selected for calibration. Since the value of some parameters is in the form of a list, we use nine single-valued variables ($\beta\_$IDEM, $\alpha\_$TSUM1, $\alpha\_$TSUM2, $\alpha\_$TDWI, $\alpha\_$SPAN, $\alpha\_$SLATB, $\alpha\_$AMAXTB, $\alpha\_$v, and $\beta\_$DVS) to control these ten parameters for relative consistency of unit's digit. The details are listed in Table I.

B. MCMC Method

MCMC is a family of algorithms for generating samples from an unknown posterior distribution through an equilibrium Markov chain [33]. MCMC makes it tractable for parameter calibration and uncertainty analysis of complex process-based models based on their prior information and observational likelihood within the Bayesian framework.

The DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm is one of the state-of-the-art MCMC methods. It runs multiple chains in parallel and uses a self-adaptive differential evolution sampling algorithm [34]. In this study, we use the DREAM algorithm for the parameter calibration of the WOFOST model under the Python environment with SPOTPY (https://github.com/thouska/spotpy.git, accessed on August 1, 2022). The algorithm proceeds as follows [35], [36].

1) Specify the prior distribution of parameters $p()$ and likelihood distribution $L()$ of observations to approximate the target posterior distribution $\pi()$.
2) The randomly selected initial point for each chain from the prior distribution of parameters, denoted as X_0, $X_0 = \{x_0^1, x_0^2, \ldots, x_0^N\}$. For x_t^i in X_0, the subscript indicates the tth iteration, and the superscript indicates the ith chain. N is the total number of chains.
3) Continue to generate candidate point of parameters space with N chains in parallel, and generate a proposal $X_{t+1,p} = \{x_{t+1,p}^1, x_{t+1,p}^2, \ldots, x_{t+1,p}^N\}$ based on $X_t = \{x_t^1, x_t^2, \ldots, x_t^N\}$. The DREAM algorithm dynamically uses subspace sampling with multiple chains. Subspace sampling means that only selected dimensions of x_t^i will be updated randomly each time a proposal is generated.
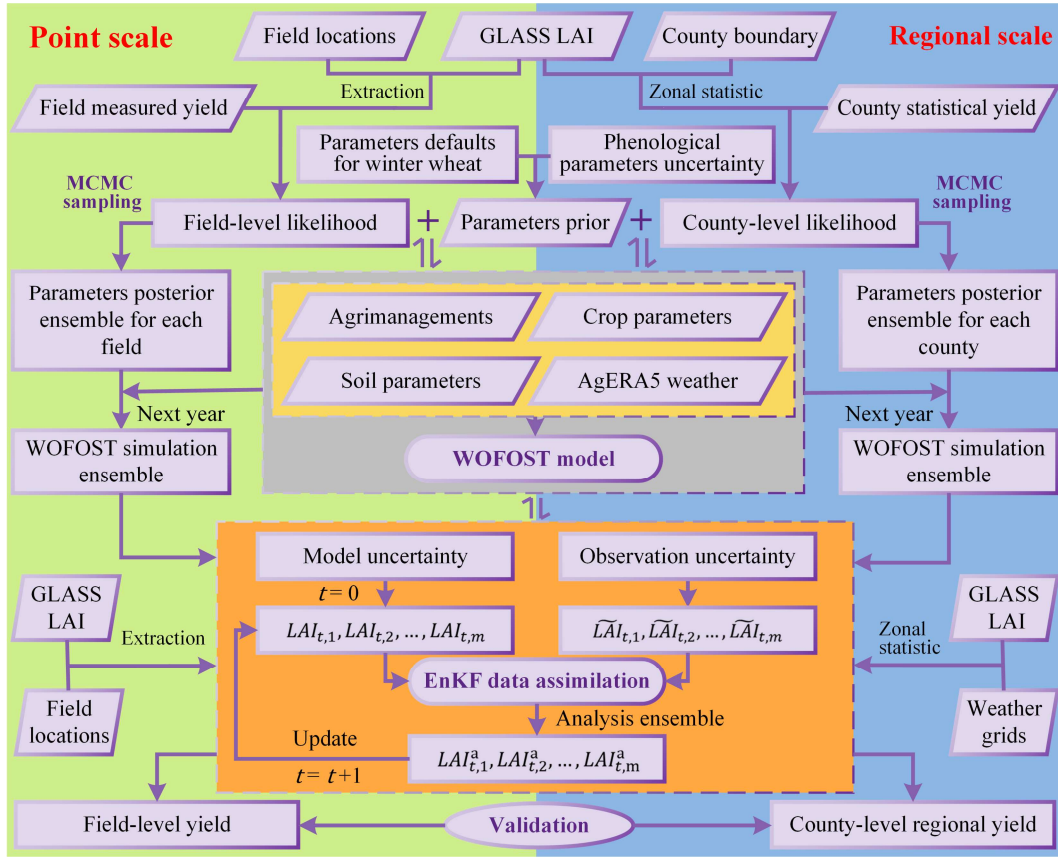
Fig. 3.   Flowchart for assimilating the GLASS LAI into the WOFOST model with the Bayesian posterior-based EnKF framework.

Let d be the dimension of $\boldsymbol{x}_t^i$, the d^* be the length of the selected dimensions. For each chain, update the parameters in the following steps sequentially.

a) Select the d^*-dimensions. First, generate a vector $\boldsymbol{CR} = [1/\text{ncr}, 2/\text{ncr}, \ldots, \text{ncr}/\text{ncr}]$, ncr is a parameter determining the number of crossover proposals. If ncr $= 1$, all dimensions of $\boldsymbol{x}_t^i$ will be updated jointly, i.e., $d^* = d$. Then, generate a random value z from a uniform distribution $U(0, 1)$ and generate a random value cr from $\boldsymbol{CR}$. Finally, repeat the last step d times independently; for each time, if $z \leq \text{cr}$, the dimension of $\boldsymbol{x}_t^i$ corresponding to this time is selected.

b) Propose a candidate point. The $\boldsymbol{x}_{t+1,p}^i$ is proposed by a jump $\Delta \boldsymbol{x}_t^i$, that is, $\boldsymbol{x}_{t+1,p}^i = \boldsymbol{x}_t^i + \Delta \boldsymbol{x}_t^i$. Let $\boldsymbol{A}$ be a subset of the selected d^*-dimensions, the $\Delta \boldsymbol{x}_t^i$ can be calculated from the collection of chains with differential evolution as follows:

$$\begin{cases} \Delta \boldsymbol{x}_{t,A}^i = \left[\boldsymbol{\lambda}_t \gamma_{(\delta_t, d_t^*)} \sum_{j=1}^{\delta_t} \left(\boldsymbol{x}_{t,A}^{a_j} - \boldsymbol{x}_{t,A}^{b_j} \right) \right] \\ \qquad + \boldsymbol{\zeta}_t, \, \boldsymbol{a}_j \neq \boldsymbol{b}_j \neq i \\ \Delta \boldsymbol{x}_{t, \neq A}^i = 0 \end{cases} \quad (2)$$

where $\boldsymbol{\lambda}_t$ is a d^*-dimensional diagonal matrix, $\boldsymbol{\lambda}_t = \text{diag}(1+\lambda_{t,1}, 1+\lambda_{t,2}, \ldots, 1+\lambda_{t,d^*})$, and the values of $\lambda_{t,1}, \lambda_{t,2}, \ldots, \lambda_{t,d^*}$ are sampled independently of a uniform distribution $U(-0.1, 0.1)$. $\boldsymbol{\zeta}_t$ is a d^*-dimensional vector, the values of $\boldsymbol{\zeta}_t$ are sampled

independently of a norm distribution $N(0, 10^{-6})$. $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors consisting of d integers drawn without replacement from $[1, 2, \ldots, N]$, and the subscript j represents the jth element of the vector. $\gamma_{(\delta_t, d_t^*)}$ is the jump rate, and its value is related to a random number between 0 and 1, denoted as $z_t = \text{random}(0, 1)$. Then,

$$\gamma_{(\delta_t, d_t^*)} = \begin{cases} \dfrac{2.38}{\sqrt{2\delta_t d_t^*}}, & z_t \geq 0.2 \\ 1, & z_t < 0.2 \end{cases} \quad (3)$$

where δ_t is the number of chain pairs used to generate the jump, δ_t is a random integer sampled from $[1, 2, \ldots, \delta]$, the default value of δ is three, and $2\delta + 1 \leq N$. d_t^* is the length of the selected dimensions for the tth iteration.

c) Accept the proposal with probability. For each chain, the probability of $\boldsymbol{x}_{t+1}^i = \boldsymbol{x}_{t+1,p}^i$ is calculated as α according to the metropolis rule (the following equation):

$$\alpha = \min \left(1, \dfrac{\pi \left(\boldsymbol{x}_{t+1,p}^i \right)}{\pi \left(\boldsymbol{x}_t^i \right)} \right)$$

$$= \min \left(1, \dfrac{p \left(\boldsymbol{x}_{t+1,p}^i \right) L \left(M \left(\boldsymbol{x}_{t+1,p}^i \right) \right)}{p \left(\boldsymbol{x}_t^i \right) L \left(M \left(\boldsymbol{x}_t^i \right) \right)} \right) \quad (4)$$

where M is a model that transforms parameters into simulations of the same dimension as

TABLE I
PARAMETER CONFIGURATION FOR MODEL CALIBRATION

Parameters	Description	Initial Values[a]	Optimized Values and Implementation
IDEM	Emergence date (day of the year)	Obs_IDEM	Obs_IDEM + β_IDEM
TSUM1	The thermal time from emergence to anthesis (°C·d^{-1})	Prior_TSUM1	Prior_TSUM1 × α_TSUM1
TSUM2	The thermal time from anthesis to maturity (°C·d^{-1})	Prior_TSUM2	Prior_TSUM2 × α_TSUM2
TDWI	Initial total crop dry weight (kg·ha^{-1})	50.0	50.0 × α_TDWI
SPAN	The life span of leaves growing at an average temperature of 35 °C (days)	31.3	31.3 × α_SPAN
SLATB	Specific leaf area as a function of development stage (ha·kg^{-1};-)	[0.00, 0.00212, 0.50, 0.00212, 2.00, 0.00212]	[0.00, 0.00212 × α_SLATB, 0.50, 0.00212 × α_SLATB, 2.00, 0.00212 × α_SLATB]
AMAXTB	Maximum CO^2 assimilation rate as a function of the development stage of the crop (kg·ha^{-1}·hr^{-1};-)	[0.00, 35.83, 1.00, 35.83, 1.30, 35.83, 2.00, 4.48]	[0.00, 35.83 × α_AMAXTB, 1.00, 35.83 × α_AMAXTB, 1.30, 35.83 × α_AMAXTB, 2.00, 4.48 × α_AMAXTB]
FLTB	Fraction of total dry matter to leaves as a function of DVS (kg·kg^{-1};-)	[0.000, 0.650, 0.100, 0.650, 0.250, 0.700, 0.500, 0.500, 0.646, 0.300, 0.950, 0.000, 2.000, 0.000]	[0.000, 0.650 × α_v, 0.100, 0.650 × α_v, 0.250, 0.700 × α_v, 0.500, 0.500 × α_v, 0.646, 0.300 × α_v, 0.950 + β_DVS, 0.000, 2.000, 0.000]
FOTB	Fraction of total dry matter to storage organs as a function of DVS (kg·kg^{-1};-)	[0.000, 0.000, 0.950, 0.000, 1.000, 1.000, 2.000, 1.000]	[0.000, 0.000, 0.950 + β_DVS, 0.000, 1.000 + β_DVS, 1.000, 2.000, 1.000]
FSTB	Fraction of total dry matter to stems as a function of DVS (kg·kg^{-1};-)	[0.000, 0.350, 0.100, 0.350, 0.250, 0.300, 0.500, 0.500, 0.646, 0.700, 0.950, 1.000, 1.000, 0.000, 2.000, 0.000]	[0.000, 1-0.650 × α_v, 0.100, 1-0.650 × α_v, 0.250, 1- 0.700 × α_v, 0.500, 1-0.500 × α_v, 0.646, 1-0.700 × α_v, 0.950 + β_DVS, 1.000, 1.000 + β_DVS, 0.000, 2.000, 0.000]

[a] The initial value of these parameters was taken from the default sets for the winter wheat variety "Winter_wheat_105" (https://github.com/ajwdewit/WOFOST_crop_parameters, accessed on 22 March 2021). The values of IDEM, TSUM1 and TSUM2 are observation-based.

observations. In practice, the x^i_{t+1} is determined using a random number z_a sampled each time independently of $U(0, 1)$

$$x^i_{t+1} = \begin{cases} x^i_{t+1,p}, & z_a \leq \alpha \\ x^i_t, & z_a > \alpha. \end{cases} \quad (5)$$

d) Assess the convergence of chains by the $\hat{R}$-statistic [37]. The samples from the collection of chains X_t after convergence are essentially equivalent to sampling from the target posterior distribution $\pi()$.

C. Calibration of the WOFOST Model

The calibration of the WOFOST model can generally be divided into two steps, the calibration of phenological parameters and the calibration of other crop parameters. For the WOFOST model, the relative accuracy of phenology is the

basis for crop simulation, but this process is a less-mechanistic module mainly driven by effective thermal time. Therefore, in this study, we first established a Thiessen polygon based on the observation data of agrometeorological stations to calibrate the model phenology zonally, and the posterior distribution of the calibrated parameters was used as the prior information of the following crop parameters calibration along with other selected parameters. In this study, the DREAM algorithm performed the above two calibration steps. As long as the prior and likelihood distribution is given, the posterior of the parameters can be obtained by iterative sampling.

1) Preliminary Calibration and 1 Uncertainty Assessment of Phenological Parameters: We first use the date of emergence from each agrometeorological station to set the value of IDEM (Obs_IDEM in Table I) and then the dates of anthesis and maturity to calibrate the parameters TSUM1 and TSUM2 of WOFOST (Prior_ TSUM1 and Prior_ TSUM2 in Table I). For simulation grids without agrometeorological stations, the parameter values are set to be the same as the calibration values of the nearest station, which can be done by constructing a Thiessen polygon (Fig. 1).

The priors of TSUM1 and TSUM2 are set to obey the uniform distribution $U(200, 1000)$ and $U(400, 1500)$, respectively. We assumed that the observed errors of anthesis and maturity dates were normally distributed with zero mean and standard deviation of 1.5 days. Therefore, the likelihood function for the phenological DVS is set as follows:

$$L(\boldsymbol{D}) = \frac{1}{(2\pi)^{\frac{k_D}{2}}} \cdot \frac{1}{|\boldsymbol{\Sigma}_{\mathrm{D}}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{D}-\boldsymbol{D}_{\mathrm{obs}})^{\top}\boldsymbol{\Sigma}_{\mathrm{D}}^{-1}(\boldsymbol{D}-\boldsymbol{D}_{\mathrm{obs}})} \quad (6)$$

where $\boldsymbol{D}$ is a 2-D vector of WOFOST simulated DVS dates, i.e., anthesis and maturity dates simulated based on given TSUM1 and TSUM2. Similarly, $\boldsymbol{D}_{\mathrm{obs}}$ is the corresponding observations of $\boldsymbol{D}$. k_D is the dimension of $\boldsymbol{D}$, and $k_D = 2$ for this study. $\boldsymbol{\Sigma}_{\mathrm{D}}$ is a 2-D diagonal matrix, $\boldsymbol{\Sigma}_{\mathrm{D}} = \mathrm{diag}(1.5, 1.5)$. In practice, the logarithm likelihood is used more generally

$$\log L(\boldsymbol{D}) = -\frac{1}{2}(\boldsymbol{D} - \boldsymbol{D}_{\mathrm{obs}})^{\top}\boldsymbol{\Sigma}_{\mathrm{D}}^{-1}(\boldsymbol{D} - \boldsymbol{D}_{\mathrm{obs}})$$
$$- \frac{k_D}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_{\mathrm{D}}|). \quad (7)$$

2) Crop Parameters Calibration: Crop parameters listed in Table I including the preliminarily calibrated IDEM, TSUM1, and TSUM2 are calibrated together based on the observation of remotely sensed time-series LAI and the field-measured or county statistical yield. The priors of these parameters are assumed to be normally distributed (Table II).

For the time-series LAI data and single-valued yield data, we assume that their errors conform to a multivariate Gaussian distribution and a univariate Gaussian distribution with zero mean, respectively. The log-likelihood function is established as follows:

$$\log L(\boldsymbol{LAI}) = -\frac{1}{2}(\boldsymbol{LAI} - \boldsymbol{LAI}_{\mathrm{obs}})^{\top}\boldsymbol{\Sigma}_{\mathrm{LAI}}^{-1}(\boldsymbol{LAI} - \boldsymbol{LAI}_{\mathrm{obs}})$$
$$- \frac{k_{\mathrm{LAI}}}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_{\mathrm{LAI}}|) \quad (8)$$

$$\log L(Y) = -\frac{1}{2}\left(\frac{Y - Y_{\mathrm{obs}}}{\sigma}\right)^2 - \frac{1}{2}\log(2\pi) - \log\sigma \quad (9)$$

TABLE II
PRIOR DISTRIBUTION OF CROP PARAMETERS

Parameters	Mean	Standard deviation[a]
β_IDEM	0	1.5
α_TSUM1	1	std_ TSUM1
α_TSUM2	1	std_ TSUM2
α_TDWI	1	0.3
α_SPAN	1	0.05
α_SLATB	1	0.05
α_AMAXTB	1	0.05
α_v	1	0.05
β_DVS	0	0.08

[a] The standard deviation for α_TSUM1 and α_TSUM2 are fitted from the posterior samples.

$$\log L = \log L(\boldsymbol{LAI}) + \log L(Y) \quad (10)$$

where $L(\boldsymbol{LAI})$ and $L(Y)$ are the likelihood for LAI and yield, and $\boldsymbol{LAI}$ and Y are WOFOST simulated LAI and yield corresponding to the observed LAI and yield. $\boldsymbol{\Sigma}_{\mathrm{LAI}}$ is the covariance matrix for LAI. k_{LAI} is the dimension of $\boldsymbol{LAI}$. σ is the standard deviation of yield.

In this study, we perform the calibration at two scales. For the point scale, we calibrate the WOFOST model with the measured yield of the fields and the GLASS LAI data from the corresponding pixels. The standard deviation of the LAI for each period is assumed as 10% of the remote sensing observation, i.e., $\boldsymbol{\Sigma}_{\mathrm{LAI}} = \mathrm{diag}((\mathrm{LAI}_1 \times 10\%)^2, (\mathrm{LAI}_2 \times 10\%)^2, \ldots, (\mathrm{LAI}_{k_{\mathrm{LAI}}} \times 10\%)^2)$, where $\mathrm{LAI}_1, \mathrm{LAI}_2, \ldots, \mathrm{LAI}_{k_{\mathrm{LAI}}}$ are the elements in vector $\boldsymbol{LAI}$. The field yield standard deviation is 10% of the measurement. For the county scale, we perform zonal statistics on GLASS LAI based on the county boundary to obtain the frequency distribution of the crop-masked pixel values. Then, the $\boldsymbol{LAI}_{\mathrm{obs}}$ and $\boldsymbol{\Sigma}_{\mathrm{LAI}}$ can be obtained by fitting the frequency distribution of the GLASS LAI value to a normal distribution. And the standard deviation of county yield is set as 10% of the statistical value.

D. Bayesian Posterior-Based EnKF

EnKF is developed based on the Kalman filter by storing, propagating, and updating an ensemble of vectors to approximate the state distribution. It is a sequential filtering method used successfully in many nonlinear and non-Gaussian data-assimilation applications [11], [37], [38], [39]. Usually, the methods for the forecast ensemble initialization can be divided into two categories. The first method generates the initial ensemble by adding Gaussian perturbation to the model input parameter, and the other directly adds Gaussian perturbation to the model simulated state variable [6], [23]. However, these two methods for perturbation are often experience-based and

can lead to potential subjective errors. This study generates the initial forecast ensemble from the model's posterior predictions. Therefore, the corresponding steps of the Bayesian posterior-based EnKF for assimilating GLASS LAI into the WOFOST model can be described explicitly as follows.

1) Generate 500 ensembles of parameter posterior samples after the MCMC chains achieve convergence in WOFOST model calibration. Thin the chains by only using every tenth sample to decrease autocorrelation [41]. Then, the 50 ensembles of posterior predictions, i.e., the model simulated state variables corresponding to the posterior parameters, are set as the initial forecast ensemble $WS_0 = [W_{0,1}, W_{0,2}, \ldots, W_{0,m}]$, $m = 50$ in this study.

2) Move the ensemble forward in time, i.e., from WS_0 to WS_t, $t = 1, 2, 3, \ldots$, until observations are available. For time t, retrieve the target states LAI_t from the ensemble WS_t, $LAI_t = [\text{LAI}_{t,1}, \text{LAI}_{t,2}, \ldots, \text{LAI}_{t,m}]$, and the covariance matrix of the target states can be estimated as P_t.

3) A 50-D ensemble of observations, denoted as $\widetilde{LAI}_t = [\widetilde{LAI}_{t,1}, \widetilde{LAI}_{t,2}, \ldots, \widetilde{LAI}_{t,m}]$, can be generated by adding a Gaussian perturbation with zero mean and a covariance matrix of R_t to the observations. In this study, we perform data assimilation at spatial resolutions of 250 m and 10 km, corresponding to the two different calibration scales. Therefore, the determination of the variance can also be divided into two cases. For data assimilation of 250 m, i.e., the same resolution of GLASS LAI, the standard deviation of the GLASS LAI is set as 2.5% of the remote sensing observation. For that of 10 km, i.e., the same resolution of AgERA5 weather, we perform zonal statistics on GLASS LAI based on the 10-km grid to obtain the frequency distribution of the crop-masked pixel values within it. Then, the Gaussian distribution of LAI observation can be fit.

4) The forecast state LAI_t in WS_t can be replaced by the analyzed state denoted as LAI_t^a as follows:

$$K_t = P_t \mathcal{H}^T \left(\mathcal{H} P_t \mathcal{H}^T + R_t \right)^{-1} \tag{11}$$

$$LAI_t^a = LAI_t + K_t \left(\widetilde{LAI}_t - \mathcal{H} LAI_t \right) \tag{12}$$

where K_t is the Kalman gain, and H is the measurement operator which specifies the relationship between the simulated state and the observed quantity. Since only one state variable, i.e., LAI is used in this study, both the covariance matrices P_t and R_t for forecast and observation states are 1-D, that is, variance scalar P_t and R_t. And H is an identity matrix because LAI is observed directly. Hence, (11) and (12) can be rewritten as follows:

$$K_t = \frac{P_t}{P_t + R_t} \tag{13}$$

$$LAI_t^a = LAI_t + K_t \left(\widetilde{LAI}_t - LAI_t \right) \tag{14}$$

where K_t is the Kalman gain, a scalar corresponding to P_t and R_t. The model forecast ensemble WS_t is updated after the above replacement. Note that some internal states in WOFOST

related to the LAI must be changed during this process to keep a balance of material and energy.

5) Repeat steps 2)–4) for the next analysis step until the end of the growing season.

The Python scripts that implement the above process are publicly available at https://github.com/paperoses/Bayesian-posterior-based-EnKF.

IV. RESULTS AND ANALYSIS

A. Uncertainty Assessment of Phenological Parameters

The phenological parameters of the WOFOST model were calibrated with the MCMC method for each agrometeorological station. In this study, the DREAM algorithm with five parallel chains approximated the posterior distribution of TSUM1 and TSUM2. The $\hat{R}$-statistic for the convergence assessment was set as 1.01. As shown in Fig. 4, the likelihood value quickly reaches a large value range and remains relatively stable, and the convergence diagnostic is close to one quickly with the MCMC chains running. This indicates that the calibration of TSUM1 and TSUM2 is relatively easy to achieve convergence. It can be seen that the two parameters are close to the normal distribution, and the maximum likelihood value falls in the high probability density area of the 2-D posterior distribution as the prior is uniformly distributed. In addition, from the perspective of the quantile, the uncertainty of TSUM2 is about twice that of TSUM1, which may be due to the cumulative effect of the parameter uncertainty. Since TSUM1 and TSUM2 were calibrated simultaneously, but TSUM2 is estimated based on TSUM1, the uncertainty of TSUM1 could therefore be accumulated to that of TSUM2. The calibrated results of each station were then applied to the corresponding Thiessen polygon zones (Fig. 1) as the prior of crop phenology and further jointly calibrated with other crop parameters.

B. Assimilating the GLASS LAI Into the WOFOST Model at the Point Scale

In this section, we further calibrated the WOFOST model for 24 fields with the field-measured yield and corresponding GLASS LAI data of 2020 based on the phenological parameters calibration results. The GLASS LAI data of 2021 during the growth period were then assimilated into the calibrated WOFOST model. We compared the estimated yield from the standard EnKF method and the Bayesian posterior-based EnKF method by validation with the field-measured yield of 2021. The fields used for calibration and validation are from the exact locations but only in different years. In addition, the percentages of 10-m winter wheat pixels within the 250-m GLASS LAI pixels corresponding to the fields have been verified to be greater than 50%, which ensures the reliability of the remote sensing pixels for the representation of winter wheat growth information.

1) Bayesian Posterior-Based EnKF: Nine parameters listed in Table II were calibrated using the DREAM algorithm with eight parallel chains and $\hat{R}$-statistic convergence diagnostic as 1.2. The posterior distribution of the selected parameters is less regular (Fig. 5) compared with that of phenological parameters calibration (Fig. 4). The reason for this is mainly
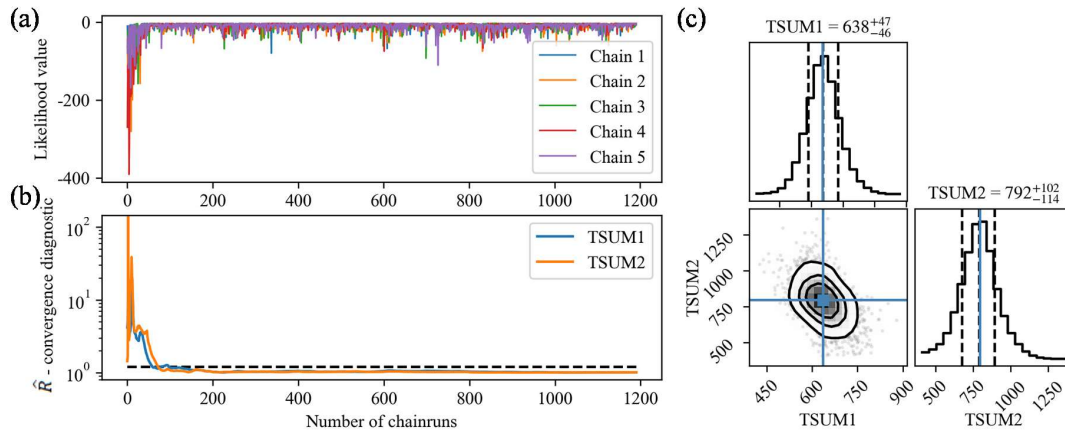
Fig. 4. Representative result for phenological parameters calibration: (a) change of likelihood value with the number of iterations; (b) change of $\hat{R}$-statistic with the number of iterations; and (c) corner plot of the parameters' posterior samples. The blue solid line indicates the maximum likelihood value. The black dashed line to the left and right of the blue line indicates the 16% and 84% quantile corresponding to the label at the top of the subplots.
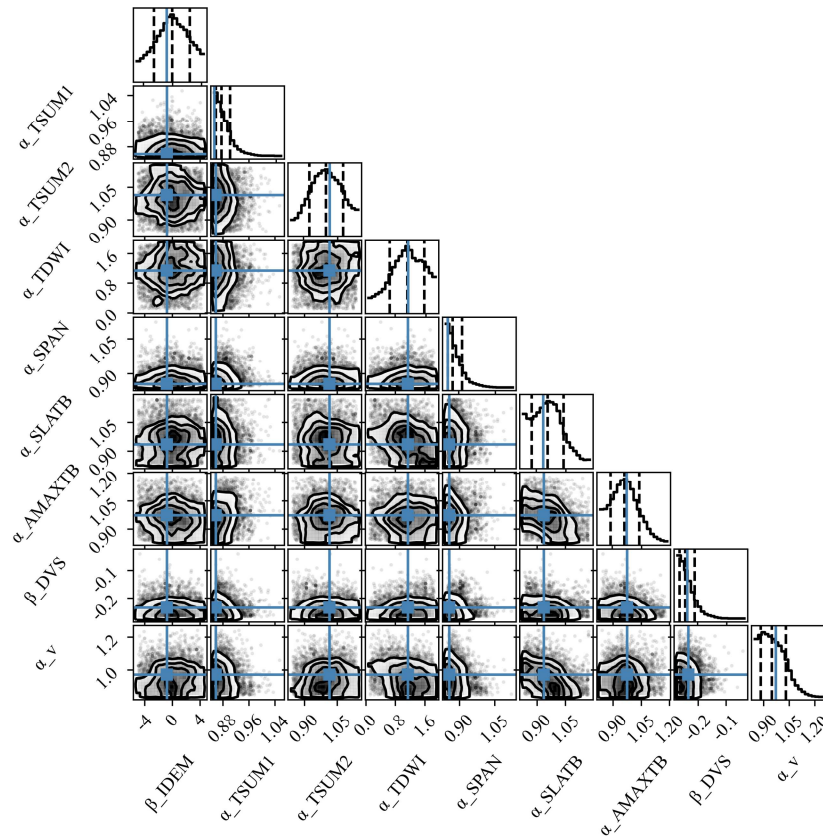

Fig. 5. Representative result of the parameters' posterior distribution from MCMC calibration at point scale (35.38°N, 113.66°E, Huixian, Henan). The blue solid line indicates the maximum likelihood value. The black dashed line from left to right indicates the 16% quantile, mean value, and 84% quantile of the posterior samples.

two aspects. On the one hand, these crop parameters or parameter combinations have alternative roles for specific effects in the WOFOST model. On the other hand, as the dimension of the parameter space increases, the convergence of MCMC sampling will be more difficult. Therefore, we set a looser threshold. The correlation between some parameters can be reflected in the 2-D distribution subgraph in Fig. 5; for example, there is a negative correlation trend between α_SLATB and α_v. This is because they have a certain degree

of equivalence to the LAI simulation, i.e., the α_SLATB has a positive correlation with the specific leaf area, and α_v has a positive effect on the fraction of total dry matter to leaves (Table I). Therefore, these two parameters have a positive regulatory effect on the LAI, calculated as the dry matter of leaves multiplied by the specific leaf area in the WOFOST model.

There can be several perspectives on the posterior uncertainty results in Fig. 5. From the view of conventional point
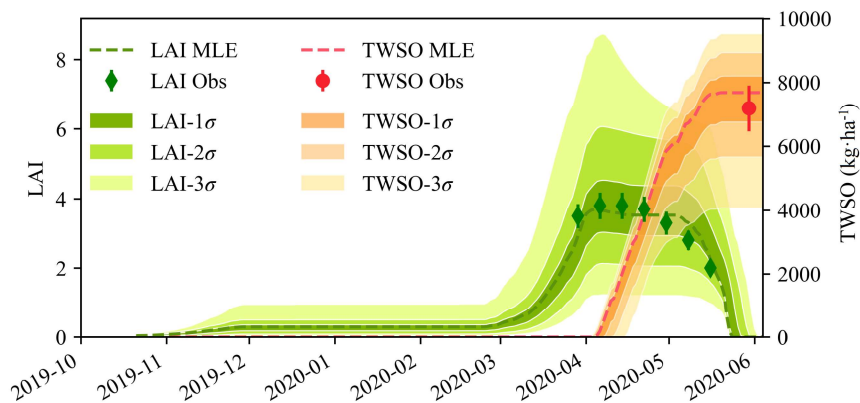
Fig. 6. Representative result of the WOFOST model's uncertainty after MCMC calibration at point scale (35.38°N, 113.66°E, Huixian, Henan). MLE and Obs are the acronyms for maximum likelihood estimate and observation, respectively.
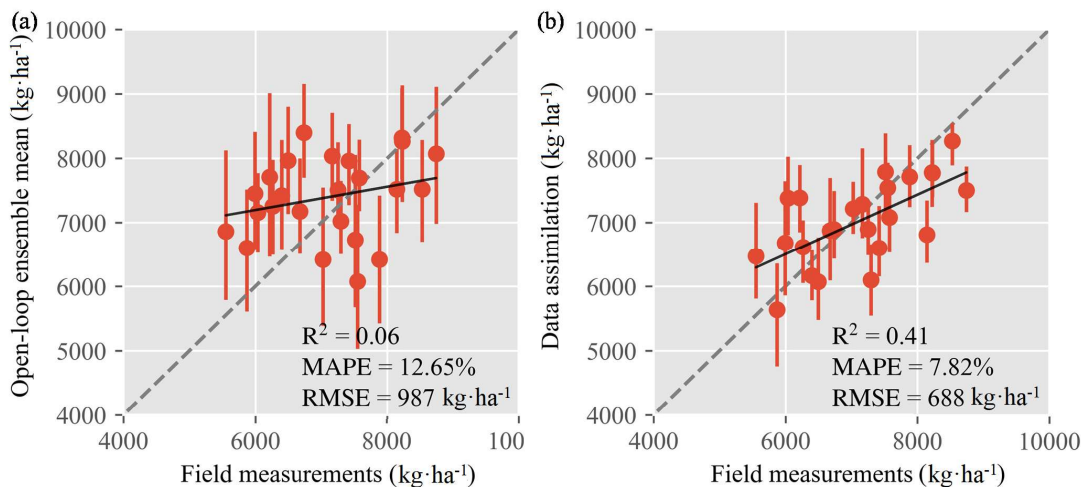

Fig. 7. Estimated winter wheat yield accuracy at point scale: (a) open loop and (b) data assimilation. The bottom and top of the error bar correspond to the 16% and 84% quantiles.

estimation, we can obtain the samples' mean value, maximum likelihood estimation, or maximum posterior estimation as the optimal-calibrated parameters set. Meanwhile, we can calculate the quantiles of the parameters individually to infer their distribution structure. The abovementioned treatments can be helpful and effective for some applications of crop growth models, but these views ignore the intrinsic correlation of parameters. They only use the 1-D projection information of the multidimensional parameter space, which is undoubtedly a less effective use of MCMC posterior information. It is reasonable to consider all posterior samples to obtain a corresponding model ensemble of simulations or predictions. For example, as shown in Fig. 6, the simulation ensemble corresponding to all samples after convergence can be calculated and characterize the uncertainty of the calibrated model, where, 1σ, 2σ, and 3σ cover 68%, 95%, and 99.7% of the ensemble intervals, respectively. In the realm of data assimilation, quantitative posterior uncertainty information is essential. For example, these posterior samples can directly calculate the parameter covariance matrix in the widely used 4-DVar data assimilation system [7], [17]. Another popular data assimilation method, EnKF, considers the uncertainty of the model in the form of

an ensemble of state variables. And the proposed Bayesian posterior-based EnKF in this study is designed to combine the MCMC posterior samples with the EnKF algorithm.

The Bayesian posterior-based EnKF method has dramatically improved the yield estimation accuracy of the corresponding 24 plots in 2021 as shown in Fig. 7. The open-loop run was based on the same 50 samples ensemble as the proposed Bayesian posterior-based EnKF method. The result shows that when the MCMC calibrated crop growth model parameters ensemble is directly used for the yield forecast, i.e., in this study, the parameters calibrated from the dataset of 2020 are used for predicting the crop yield of 2021, relatively large errors could remain. However, assimilating the within-season remote sensing observations could improve crop model prediction accuracy while reducing uncertainty (Fig. 7).

We further compare the relationship between the WOFOST ensemble simulations and observations for all fields. In general, they can be divided into four cases, i.e., the model LAI ensemble is generally higher than the remote sensing LAI, and the corresponding yield ensemble mean is higher than the field observation (shorted as "high–high"); the model LAI ensemble is generally lower than the remote sensing
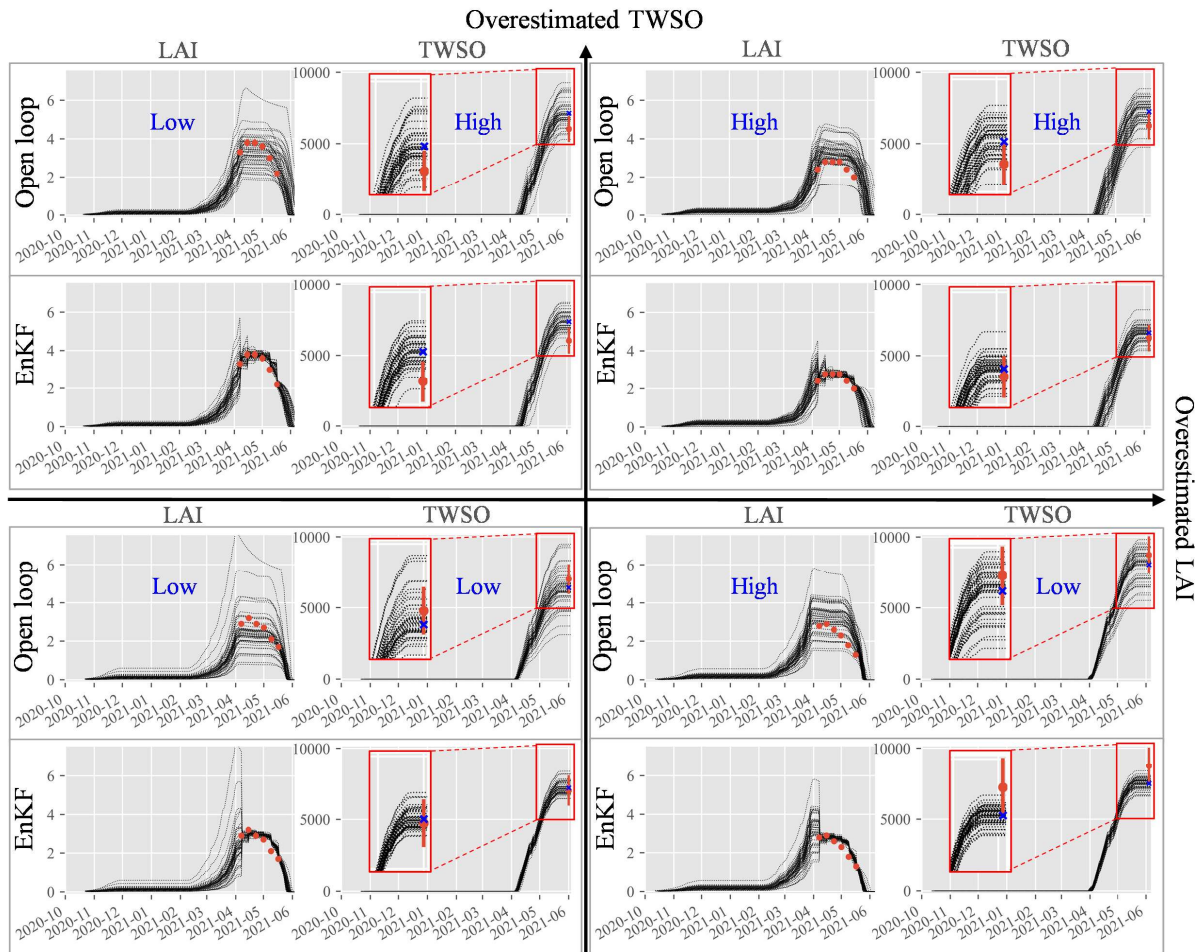
Fig. 8. Four relationship types between model and observation and their corresponding results of EnKF data assimilation. The division of the four quadrants is based on the relative relationship between the open-loop model ensemble and the observation. For example, the first quadrant in the top right corner indicates that the model LAI ensemble is higher than the remote sensing LAI as a whole and the yield ensemble mean (the blue cross) is higher than the field observation. We refer to it as "high–high" for short. Similarly, the second, third, and fourth quadrants correspond to the types "low–high," "low–low," and "high–low.".

LAI, and the corresponding yield ensemble mean is higher than the field observation (shorted as "low–high"); the model LAI ensemble is generally lower than the remote sensing LAI, and the corresponding yield ensemble mean is lower than the field observation (shorted as "low–low"); and the model LAI ensemble is generally higher than the remote sensing LAI, and the corresponding yield ensemble mean is lower than the field observation (shorted as "high–low"), as shown in Fig. 8. Although there is not necessarily a positive correlation between yield and LAI from the perspective of the WOFOST model mechanism, the results show that the yield accuracy after EnKF data assimilation will be reduced when the direction of the simulated LAI and yield deviate from the corresponding observed value is inconsistent, i.e., the cases of "low–high" and "high–low." On the contrary, for the cases "high–high" and "low–low," most of the estimated yield accuracy after data assimilation is improved, and the few decreases in accuracy because of the over-adjustment of the yield estimation after data assimilation, e.g., a slight overestimation becomes a relative serious underestimation. Specifically, 83% of the fields can fall into the first or third

quadrants in Fig. 8, 75% of which increases in accuracy, so the overall accuracy of all fields shown in Fig. 7 can be improved after data assimilation.

2) Comparison With Standard EnKF: The posterior sample of MCMC can provide lots of point estimates for model calibration. Therefore, we can generate the model ensemble by a manual Gaussian perturbation, apply the standard Gaussian perturbation-based EnKF data assimilation framework [6], [23], [42], and compare it with the Bayesian posterior-based EnKF. In this study, based on the posterior mean estimate, maximum likelihood estimate, and maximum posterior estimate, we added five kinds of perturbations of 1%, 5%, 10%, 15%, and 20% to the parameters TDWI and SPAN, respectively. The accuracy of all different treatments shown in Table III is lower than that of the proposed Bayesian posterior-based EnKF method with R^2 equal to 0.41 and mean absolute percentage error (MAPE) equal to 7.82% (Fig. 7). In addition, the results indicate that the posterior mean estimate is the optimal point estimate for data assimilation with Gaussian perturbation-based EnKF. The treatment based on the posterior mean estimate and a Gaussian perturbation of

Point estimates	Perturbations	R^2	MAPE (%)	RMSE (kg·ha^{-1})
Posterior mean	1%	0.08	13.41	1066
	5%	0.35	8.49	719
	10%	0.37	8.20	707
	15%	0.38	8.20	708
	20%	0.38	8.66	761
Maximum likelihood	1%	0.03	14.55	1207
	5%	0.10	11.82	1022
	10%	0.18	10.19	933
	15%	0.22	9.95	928
	20%	0.22	10.04	955
Maximum posterior	1%	0.02	13.67	1094
	5%	0.13	11.02	880
	10%	0.20	10.01	822
	15%	0.23	9.39	829
	20%	0.24	9.46	817

10%–15% achieves the highest accuracy with the coefficient of determination R^2 from 0.37 to 0.38 and the MAPE of about 8.20% (Table III). The relatively better performance with the posterior mean estimation may be because it contains more information about the Bayesian high probability space of the calibrated parameters. The apparent disadvantage of the standard Gaussian perturbation-based EnKF is that the model uncertainty of all data assimilation units is subjective and invariant. However, the Bayesian posterior provides an objective description of the model uncertainty individually, which makes the data assimilation system more reasonable and reliable.

C. Assimilating the GLASS LAI Into the WOFOST Model at the Regional Scale

In this section, we further calibrated the model for the counties in Henan with the county-level statistical yield and corresponding GLASS LAI data within the county boundary from 2017 to 2019 based on the phenological parameters calibration results. The GLASS LAI data during the growth period from 2018 to 2020 were then assimilated into the calibrated WOFOST model at the spatial resolution of 10 km, i.e., the same resolution as the AgERA5 weather data. Therefore, we calculated the percentage of 10-m winter wheat pixels within each 250-m GLASS LAI grid and fit the GLASS LAI pixels with a purity of more than 50% to a Gaussian distribution within each 10-km grid.

1) Bayesian Posterior-Based EnKF: The county-level posterior of MCMC calibration is similar to that of the point scale. However, the posterior parameter ensemble is used for all 10-km winter wheat grids within the corresponding county. The open-loop run of the WOFOST model is driven by the current year's weather data based on the crop parameters calibrated by the previous year's data. Limited by the potential insufficient representativeness of model calibration for different growth conditions, the open-loop simulations [Fig. 9(a), (d), and (g)] can have large uncertainties because of the differences in meteorological environment and management between the two years before and after. The data assimilation recalibrated the state variables of the crop model and improved yield simulations based on remote sensing observations within the growth period. The simulations after data assimilation [Fig. 9(b), (e), and (h)] have more consistent spatial trends and additional details compared with the county statistics [Fig. 9(c), (f), and (i)] even though the open-loop simulations can have a large error. We aggregated the simulation results at a 10-km resolution to the county scale, and as shown in Fig. 10, simulated county-level yields after data assimilation were in good agreement with statistical values. Simulated county-level yield after data assimilation for 2018, 2019, and 2020 achieved R^2 values of 0.59, 0.67, and 0.42, and RMSE values of 605, 658, and 812 kg·ha^{-1}, respectively. The MAPE value decreased from 13.24%, 30.15%, and 14.87% of the open loop to 8.85%, 9.58%, and 12.06% of the data assimilation, respectively. The overall results with the Bayesian posterior-based EnKF for 2018–2020 show a 9.54% improvement in accuracy based on the MAPE value compared with the open-loop results. Considering that yield observations corresponding to 10-km resolution are not available and only winter wheat in high cropping density areas was selected, the 10-km yield estimates cannot be directly and adequately validated with the county statistics, but the improved accuracy shown in Fig. 10 can still demonstrate to some extent the validity of the proposed data assimilation method for regional yield estimation.

2) Comparison With Standard EnKF: We compare the proposed Bayesian posterior-based EnKF method with the Gaussian perturbation-based EnKF method. Table IV shows the results of the different Gaussian perturbations based on the posterior mean estimate, which is the overall optimal point estimate and contains the treatment with the highest accuracy. The results show that our proposed method is overall higher than the different comparison treatments, except for the treatment of 10%–15% Gaussian perturbation based on the posterior mean estimate in 2018, which achieves R^2 values of 0.58–0.61, MAPE values of 9.09–8.63, and RMSE values of 594–596 kg·ha^{-1}, and is slightly higher than the accuracy of our proposed method with R^2 values of 0.59, MAPE values of 8.85%, and RMSE values of 605 kg·ha^{-1}. It should be noted that the Gaussian perturbation corresponding to the highest accuracy is not the same for different years (Table IV), so in practice, it is difficult to determine a fixed optimal perturbation applicable to different years or areas. Our proposed method provides adaptive model uncertainty based on the MCMC
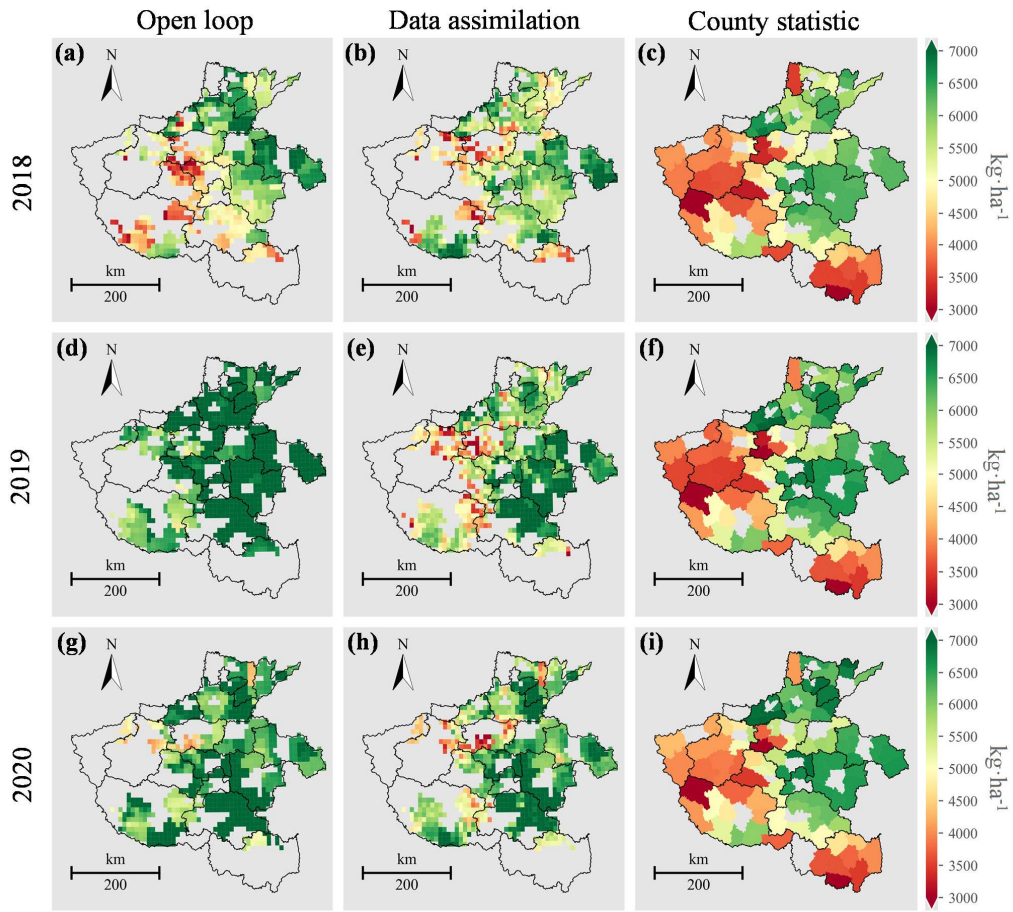
Fig. 9.   Estimated winter wheat yield map. (a)–(c) Represent the yield map of open loop, data assimilation, and county statistic for the year 2018. (d)–(f) Represent the yield map of open loop, data assimilation, and county statistic for the year 2019. (g)–(i) Represent the yield map of open loop, data assimilation, and county statistic for the year 2020. All yields are expressed in kg·ha^{-1} dry weight.
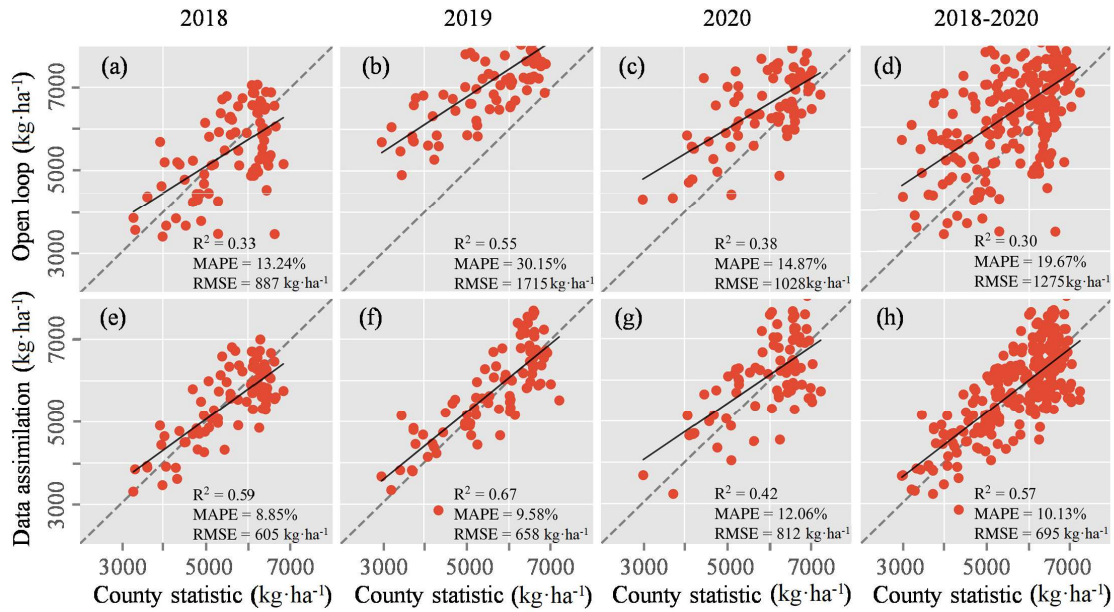

Fig. 10.   Estimated winter wheat yield accuracy at county-level scale: (a)–(d) represent the accuracy of open-loop estimated yields for year-by-year and all years from 2018 to 2020, respectively, and (e)–(h) are the corresponding accuracy of data assimilation.

algorithm and is, therefore, more universally applicable. From the total results for 2018–2020 (Fig. 10 and Table IV), the Bayesian posterior-based EnKF improved R^2 by 0.06–0.24, MAPE by 0.94%–11.00%, and RMSE by 46–653 kg·ha^{-1} compared with five different sets of Gaussian perturbation-based EnKF based on the optimal point estimate.

TABLE IV
Comparison With Standard EnKF Framework Based on
the Posterior Mean Estimate and Five Different
Gaussian Perturbations

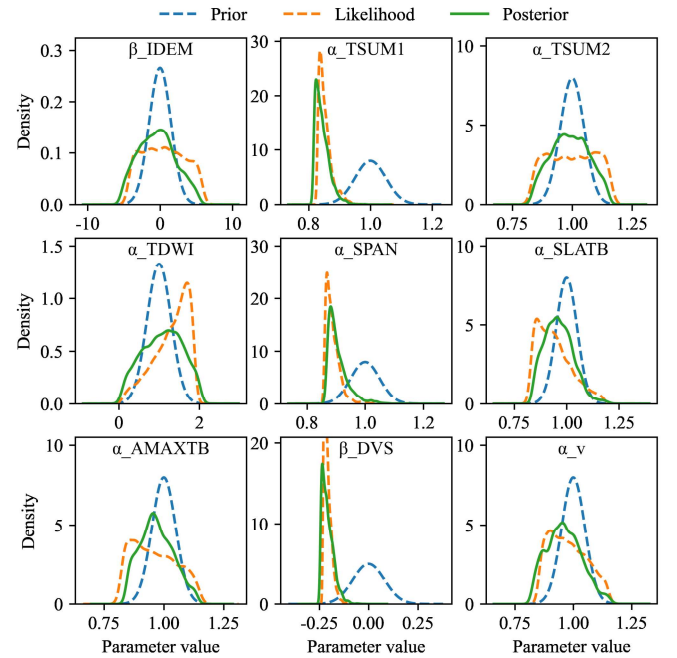| Year	Gaussian perturbations	R^2	MAPE (%)	RMSE (kg·ha^{-1})
2018	1%	0.41	13.19	821
	5%	0.48	11.40	712
	10%	0.58	9.09	594
	15%	0.61	8.63	596
	20%	0.59	9.16	661
2019	1%	0.50	32.21	1825
	5%	0.63	27.23	1556
	10%	0.69	19.20	1134
	15%	0.67	12.78	793
	20%	0.63	10.13	672
2020	1%	0.41	17.20	1157
	5%	0.43	14.25	985
	10%	0.41	12.00	822
	15%	0.35	12.55	816
	20%	0.27	14.09	947
2018-2020	1%	0.33	21.13	1348
	5%	0.40	17.86	1151
	10%	0.48	13.56	885
	15%	0.51	11.30	741
	20%	0.49	11.07	767

Fig. 11. Comparison of prior, likelihood, and posterior distributions of parameters in the MCMC calibration process at point scale (35.38°N, 113.66°E, Huixian, Henan). Note that the likelihood function is constructed based on crop model outputs such as LAI and yield, and the likelihood of the model parameters cannot be calculated directly. Therefore, the likelihood in the figure is obtained by setting the model parameters to a uniform distribution and then obtaining the parameter posterior based on the observation likelihood and MCMC sampling.

V. Discussion

A. Advantages of MCMC Method for Crop Growth Model Calibration

Calibration of crop growth models involves estimating model parameters according to observations, including phenological DVSs, leaf, biomass dynamics, and final yield. Standard practices are often based on field measurements or agronomic trial data under different control treatments. However, at least part of such a dataset is often rarely available for common model users. In contrast, the remotely sensed LAI products and yield statistics used as the observational dataset for parameter inference in this study are more readily available. Fig. 11 shows the change of model parameters and their uncertainty from the prior to the posterior under the action of the observational likelihood. The posterior of the parameters provides a comprehensive way to quantify model uncertainty. But one problem that may arise is that there is often an underestimation of crop LAI because remote sensing LAI products are often not produced for specific crops. This can cause the calibration values of parameters closely related to LAI in the crop model to deviate to some extent from the

reasonable interval (e.g., the α_TSUM1, α_SPAN, and β_DVS in Fig. 11). In addition, an increasing number of datasets on crop growth and development [43], [44] can be used for crop growth model calibration. In practice, ideal parameter calibration results can be obtained based on different methods for cases with more and better data. However, in many cases, the datasets used for calibration are inadequate and uncertain, and complex process-based models usually have the equifinality problem, i.e., different parameter combinations have equal performance for a given set of observations [45]. Therefore, widely used approaches such as the search for the parameters that minimize the sum of squared errors and maximum likelihood function tend to be difficult to obtain reasonable parameter estimates. The Bayesian theory-based MCMC method can infer the posterior distribution by combining the prior information of the parameters and the observation likelihood. Model calibration with the MCMC method can avoid parameter estimates beyond the explicit physiological and physical interpretation by setting a strongly information-bounded prior distribution, such as the Gaussian distribution in this study. This is particularly important when the observation is insufficient (Fig. 11). Most importantly, the MCMC method provides not only point estimates of the parameters but also uncertainty information, including the covariance and quantile of the parameters. In addition, the Bayesian posterior mean is shown to be optimal as a point estimate in the standard Gaussian perturbation-based EnKF framework, which is used to compare with our proposed method. But in reality, less than 10% of model users calibrate crop growth models based
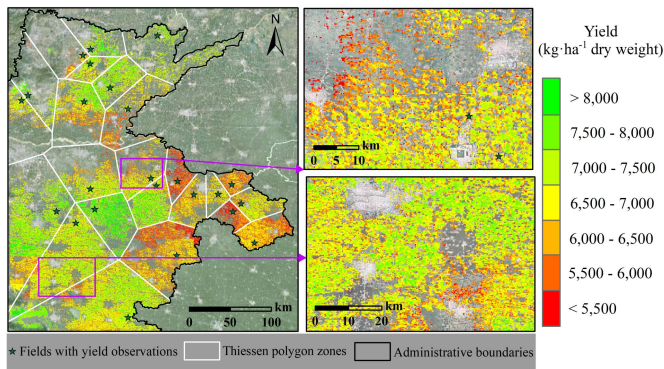
Fig. 12. Estimated 250-m regional winter wheat yield map of 2021.

on Bayesian approaches [45], so in practice, the standard Gaussian perturbation-based EnKF may perform worse than that shown in this study. This study focuses on obtaining posterior samples of model parameters based on the MCMC algorithm to improve the quantified estimation of model errors in the EnKF algorithm. The results show that this allows for improved data assimilation accuracy compared to the empirical model error setting regionally. The method is also helpful for model error quantification in other data assimilation algorithms, e.g., the posterior samples of model parameters can be used to calculate the covariance matrix of model parameters in the 4-DVar algorithm.

B. Relationship Between the Point and Regional Scales

Theoretically, the calibration of the point scale in this study can be used for regional yield estimation based on the spatialization of parameters. For example, Thiessen polygons can be generated from the sample points with field yield observations, so that any location inside the polygon shares the same calibrated parameters posterior as the field-level point. We then assimilated the GLASS LAI into the WOFOST model with the proposed framework for 250-m regional yield estimation (Fig. 12). To a certain extent, spatial pattern in yield variability in Fig. 12 is reasonable, for example, areas with higher planting densities tend to have higher yields. However, the sparse distribution or insufficient representation of sampling points can cause large errors, especially the significant differences between adjacent Thiessen polygon zones. In this study, the point scale is used to verify the validity of the proposed framework, as this does not involve scale differences in yield. We estimated the regional yield by data assimilation at 10-km grids based on county-level calibration with only statistical yield and aggregated GLASS LAI data considering the representativeness of the calibration and the consistency of uncertainty quantification. The primary challenge for the regional application of our proposed framework is the quantitative assessment of uncertainty in remote sensing observations. We addressed this challenge by approximating regional uncertainty through zonal statistics in both county calibration and 10-km data assimilation. If regional pixel-level uncertainty can be accurately quantified, our framework can be used to estimate regional yield at the remote sensing resolution [14].

C. Model Simulation and Data Assimilation Under Stress Conditions

As presented in Fig. 8, the data assimilation with remote sensing LAI in this study improves the yield accuracy only for the cases where the open-loop modeled yield errors and LAI errors are in the same direction. An important reason is that the simulated yield is closely related to multiple state variables, yet adjusting the model state by LAI alone may lead to over-adjustment or larger errors, especially when crop growth is stressed by water or disaster, for example. Previous studies have demonstrated the data assimilation of other variables, including soil moisture [46], evapotranspiration [47], biomass [48], and gross primary production [11], into crop growth models for improving yield estimation. Further, simultaneous data assimilation of multiple state variables into the crop growth model tends to provide more promising yield accuracy [49], [50]. Therefore, assimilating remote sensing LAI and other model state variables under water-limited or nutrient-limited modes should be considered in future studies.

D. Filter Divergence Problem of EnKF

A common problem with the EnKF algorithm is the poor representation of the covariance functions and the error covariance matrix caused by the limited-size ensemble, leading to under-estimated prediction uncertainty and possible filter divergence, i.e., the Kalman gain is getting smaller and the system tends to reject the observation in the later stages of data assimilation [9]. The inflation approach is often used to avoid this problem by offsetting excessive variance reduction in the update [51]. To a certain extent, the inflation for the forecast ensemble is a way to compensate for unknown model errors. However, in this study, the model ensemble is generated by different parameter sets sampled from the parameter posterior in each ensemble member. The forward move of forecast ensemble after each update contains information about the model errors from the parameter posterior. Therefore, it is no longer necessary to add additional inflation factors to manually increase the prediction uncertainty. Specifically, filter divergence is still a concern when the model ensemble is generated by adding Gaussian perturbations directly to the state variables. In future work, we still need to develop more robust methods for data assimilation, such as hybrid algorithms where model parameters and state variables can be updated simultaneously [22].

VI. CONCLUSION

The accurate quantification of model uncertainty is significant for a successful data assimilation system. The widely used methods including EnKF and 4-DVar algorithms assume that the errors of the models are Gaussian distributed and most studies can only give the errors based on subjective experience. The proposed Bayesian posterior-based EnKF framework in this study quantifies the model uncertainty in data assimilation system with the satellite product and Bayesian method. The proposed framework was applied with GLASS 250-m LAI product and the WOFOST model at point and regional scales in Henan. We first obtained the

posterior prediction ensemble of the WOFOST model based on the time-series GLASS LAI data and the field-measured yield. We carried out data assimilation on the point-level 250-m grid using the GLASS LAI data within the critical growth period in the following year. The results showed that the accuracy of yield estimation with data assimilation was significantly improved compared with open-loop simulation, with R^2 increasing from 0.06 to 0.41, MAPE decreasing from 12.65% to 7.82%, and RMSE decreasing from 987 to 688 kg·ha^{-1}. Similarly, we estimated winter wheat yield at 10-km spatial resolution based on county statistical yield and zonal statistical values of GLASS LAI. The results show that the simulated yields were in better agreement with county statistical values with R^2 values of 0.57, MAPE values of 10.13%, and RMSE values of 695 kg·ha^{-1} compared with the open-loop simulation. Moreover, the Bayesian posterior-based EnKF applied to both scales performs better overall than the standard EnKF method, with the model error estimated based on the Gaussian perturbation. This study shows that crop yield estimation can be significantly improved by better quantifying the crop growth model errors in the data assimilation system.

REFERENCES

[1] A. R. Bentley et al., "Near-to long-term measures to stabilize global wheat supplies and food security," *Nature Food*, vol. 3, no. 7, pp. 483–486, Jul. 2022.

[2] S. Fritz et al., "A comparison of global agricultural monitoring systems and current gaps," *Agricult. Syst.*, vol. 168, pp. 258–272, Jan. 2019.

[3] H. Huang et al., "A dataset of winter wheat aboveground biomass in China during 2007–2015 based on data assimilation," *Sci. Data*, vol. 9, no. 1, p. 200, May 2022.

[4] X. Wang, J. Huang, Q. Feng, and D. Yin, "Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches," *Remote Sens.*, vol. 12, no. 11, p. 1744, May 2020.

[5] J. Huang et al., "Assimilation of remote sensing into crop growth models: Current status and perspectives," *Agricult. Forest Meteorol.*, vols. 276–277, Oct. 2019, Art. no. 107609.

[6] Y. Curnel, A. J. W. De Wit, G. Duveiller, and P. Defourny, "Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS experiment," *Agricult. Forest Meteorol.*, vol. 151, no. 12, pp. 1843–1855, Dec. 2011.

[7] J. Huang et al., "Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model," *Agricult. Forest Meteorol.*, vol. 204, pp. 106–121, May 2015.

[8] S. Wu, P. Yang, J. Ren, Z. Chen, and H. Li, "Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm," *Remote Sens. Environ.*, vol. 255, Mar. 2021, Art. no. 112276.

[9] G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Cham, Switzerland: Springer, 2022.

[10] A. V. M. Ines, N. N. Das, J. W. Hansen, and E. G. Njoku, "Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction," *Remote Sens. Environ.*, vol. 138, pp. 149–164, Nov. 2013.

[11] W. Zhuo et al., "Assimilating remote sensing-based VPM GPP into the WOFOST model for improving regional winter wheat yield estimation," *Eur. J. Agronomy*, vol. 139, Sep. 2022, Art. no. 126556.

[12] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dyn.*, vol. 53, no. 4, pp. 343–367, Nov. 2003.

[13] M. G. Ziliani et al., "Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model," *Agricult. Forest Meteorol.*, vol. 313, Jan. 2022, Art. no. 108736.

[14] Y. Kang and M. Özdogan, "Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach," *Remote Sens. Environ.*, vol. 228, pp. 144–163, Jul. 2019.

[15] V. Bandaru et al., "Geo-CropSim: A geo-spatial crop simulation modeling framework for regional scale crop yield and water use assessment," *ISPRS J. Photogramm. Remote Sens.*, vol. 183, pp. 34–53, Jan. 2022.

[16] S. J. Maas, "Use of remotely-sensed information in agricultural crop growth models," *Ecological Model.*, vol. 41, nos. 3–4, pp. 247–268, Jun. 1988.

[17] L. Dente, G. Satalino, F. Mattia, and M. Rinaldi, "Assimilation of leaf area index derived from ASAR and MERIS data into CERES-wheat model to map wheat yield," *Remote Sens. Environ.*, vol. 112, no. 4, pp. 1395–1407, Apr. 2008.

[18] D. Wallach, "Crop model calibration: A statistical perspective," *Agronomy J.*, vol. 103, no. 4, pp. 1144–1151, Jul. 2011.

[19] S. Cheng and M. Qiu, "Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks," *Neural Comput. Appl.*, vol. 34, no. 16, pp. 13149–13167, Aug. 2022.

[20] A. Gruber, G. De Lannoy, and W. Crow, "A Monte Carlo based adaptive Kalman filtering framework for soil moisture data assimilation," *Remote Sens. Environ.*, vol. 228, pp. 105–114, Jul. 2019.

[21] J. Tian et al., "Improving surface soil moisture retrievals through a novel assimilation algorithm to estimate both model and observation errors," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112802.

[22] A. S. Nair, R. Mangla, and J. Indu, "Remote sensing data assimilation," *Hydrological Sci. J.*, vol. 102, pp. 1–33, Jun. 2020.

[23] J. Huang et al., "Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation," *Agricult. Forest Meteorol.*, vol. 216, pp. 188–202, Jan. 2016.

[24] H. Pan, Z. Chen, A. de Wit, and J. Ren, "Joint assimilation of leaf area index and soil moisture from Sentinel-1 and Sentinel-2 data into the WOFOST model for winter wheat yield estimation," *Sensors*, vol. 19, no. 14, p. 3161, Jul. 2019.

[25] T. Iizumi, M. Yokozawa, and M. Nishimori, "Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a Bayesian approach," *Agricult. Forest Meteorol.*, vol. 149, no. 2, pp. 333–348, Feb. 2009.

[26] H. Ran et al., "A framework to quantify uncertainty of crop model parameters and its application in arid Northwest China," *Agricult. Forest Meteorol.*, vol. 316, Apr. 2022, Art. no. 108844.

[27] J. Sexton, Y. Everingham, and G. Inman-Bamber, "A theoretical and real world evaluation of two Bayesian techniques for the calibration of variety parameters in a sugarcane crop model," *Environ. Model. Softw.*, vol. 83, pp. 126–142, Sep. 2016.

[28] Y. Wu, W. Xu, H. Huang, and J. Huang, "Bayesian posterior-based winter wheat yield estimation at the field scale through assimilation of Sentinel-2 data into WOFOST model," *Remote Sens.*, vol. 14, no. 15, p. 3727, Aug. 2022.

[29] X. Huang, J. Huang, X. Li, Q. Shen, and Z. Chen, "Early mapping of winter wheat in Henan province of China using time series of Sentinel-2 data," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1534–1549, Dec. 2022.

[30] H. Ma and S. Liang, "Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model," *Remote Sens. Environ.*, vol. 273, May 2022, Art. no. 112985.

[31] H. Boogaard, J. Schubert, A. De Wit, J. Lazebnik, R. Hutjes, and G. Van der Grijn, "Agrometeorological indicators from 1979 to present derived from reanalysis, version 1.0," Copernicus Climate Change Service (C3S), Climate Data Store (CDS), 2020, doi: 10.24381/cds.6c68c9bb.

[32] A. de Wit et al., "25 years of the WOFOST cropping systems model," *Agricult. Syst.*, vol. 168, pp. 154–167, Jan. 2019.

[33] S. Brooks, A. Gelman, G. Jones, and X. Meng, *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL, USA: CRC Press, 2011.

[34] J. A. Vrugt, C. J. F. T. Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon, "Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling," *Int. J. Nonlinear Sci. Numer. Simul.*, vol. 10, no. 3, pp. 273–290, Jan. 2009.

[35] T. Houska, P. Kraft, A. Chamorro-Chavez, and L. Breuer, "SPOTting model parameters using a ready-made Python package," *PLoS ONE*, vol. 10, no. 12, Dec. 2015, Art. no. e0145180.

[36] J. A. Vrugt, "Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation," *Environ. Model. Softw.*, vol. 75, pp. 273–316, Jan. 2016.

[37] A. Gelman and D. Rubin, "A single series from the Gibbs sampler provides a false sense of security," *Bayesian Statisics*, vol. 4, pp. 625–631, Jul. 1992.

[38] A. J. W. de Wit and C. A. van Diepen, "Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts," *Agricult. Forest Meteorol.*, vol. 146, nos. 1–2, pp. 38–56, Sep. 2007.

[39] M. Katzfuss, J. R. Stroud, and C. K. Wikle, "Understanding the ensemble Kalman filter," *Amer. Statist.*, vol. 70, no. 4, pp. 350–357, 2016.

[40] X. Li, X. Liu, and P. Gong, "Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 5, pp. 762–785, May 2015.

[41] W. A. Link and M. J. Eaton, "On thinning of chains in MCMC," *Methods Ecology Evol.*, vol. 3, no. 1, pp. 112–115, Feb. 2012.

[42] F. Ji, J. Meng, Z. Cheng, H. Fang, and Y. Wang, "Crop yield estimation at field scales by assimilating time series of Sentinel-2 data into a modified CASA-WOFOST coupled model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4400914.

[43] Y. Luo, Z. Zhang, Y. Chen, Z. Li, and F. Tao, "ChinaCropPhen1km: A high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products," *Earth Syst. Sci. Data*, vol. 12, no. 1, pp. 197–214, Jan. 2020.

[44] Q. Niu et al., "A 30 m annual maize phenology dataset from 1985 to 2020 in China," *Earth Syst. Sci. Data*, vol. 14, no. 6, pp. 2851–2864, Jun. 2022.

[45] S. J. Seidel, T. Palosuo, P. Thorburn, and D. Wallach, "Towards improved calibration of crop models—Where are we now and where should we go?" *Eur. J. Agronomy*, vol. 94, pp. 25–35, Mar. 2018.

[46] G. S. Nearing, W. T. Crow, K. R. Thorp, M. S. Moran, R. H. Reichle, and H. V. Gupta, "Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment," *Water Resour. Res.*, vol. 48, no. 5, pp. 1–13, May 2012.

[47] A. V. M. Ines, K. Honda, A. D. Gupta, P. Droogers, and R. S. Clemente, "Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture," *Agricult. Water Manag.*, vol. 83, no. 3, pp. 221–232, Jun. 2006.

[48] X. Jin et al., "Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm," *ISPRS J. Photogram. Remote Sens.*, vol. 126, pp. 24–37, Apr. 2017.

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

[50] A. K. Mishra et al., "Anatomy of a local-scale drought: Application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study," *J. Hydrol.*, vol. 526, pp. 15–29, Jul. 2015.

[51] J. L. Anderson and S. L. Anderson, "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," *Monthly Weather Rev.*, vol. 127, no. 12, pp. 2741–2758, Dec. 1999.

Jianxi Huang received the B.Sc. degree from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1999, the M.Sc. degree from Wuhan University, Wuhan, in 2002, and the Ph.D. degree in agricultural remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2006.

He is currently a Professor with the College of Land Science and Technology, China Agricultural University, Beijing. He has authored more than 90 articles in international journals. His research interests include data assimilation applications in agriculture, crop yield estimation and forecasting, and agricultural disaster monitoring and assessment.

Yantong Wu received the bachelor's degree from the College of Management Science, Chengdu University of Technology, Chengdu, China, in 2015, and the Ph.D. degree from the College of Resources and Environment, University of Electronic Science and Technology, Chengdu, in 2022.

She is currently a Research Assistant with the College of Land Science and Technology, China Agricultural University, Beijing, China. Her research interests include remote sensing, crop growth models, and data assimilation application in agriculture.

Wen Zhuo received the B.Sc. degree in geographic information system from Huazhong Agricultural University, Wuhan, China, in 2013, and the M.Sc. degree in geographic information system and the Ph.D. degree in agricultural information technology from China Agricultural University, Beijing, China, in 2015 and 2020, respectively.

He is currently a Post-Doctoral Fellow with the Chinese Academy of Meteorological Sciences, Beijing. His research interests include remote sensing data assimilation with crop growth models, drought monitoring, and climate change impacts on agriculture using remote sensing.

Jianjian Song is currently pursuing the bachelor's degree with the Changwang School of Honors, Nanjing University of Information Science and Technology, Nanjing, China.

His major research interests include remote sensing, crop growth models, and data assimilation-based crop yield estimation.

Hai Huang received the bachelor's degree from the Agricultural College, Yangzhou University, Yangzhou, China, in 2017, and the master's degree from the College of Land Science and Technology, China Agricultural University, Beijing, China, in 2019, where he is currently pursuing the Ph.D. degree.

His major research interests include remote sensing, crop growth models, and data assimilation-based crop yield estimation.

Xuecao Li received the B.Sc. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2012, and the Ph.D. degree in ecology from Tsinghua University, Beijing, China, in 2016.

He is currently a Professor with the College of Land Science and Technology, China Agricultural University, Beijing. He has authored more than 90 articles in international journals. His research interests include land use/cover change and modeling, global environmental change, and sustainable development.

Li Li received the B.Sc. degree in physics and the M.Sc. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 1999 and 2002, respectively, and the Ph.D. degree in communication and information system from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2005.

She is currently an Associate Professor with the College of Land Science and Technology, China Agricultural University, Beijing. She has authored more than 50 articles in the research area in recent years. Her research focuses on microwave remote sensing and its application in agriculture, land information monitoring, and agricultural disaster monitoring.

Han Ma received the B.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2010, and the M.S. and Ph.D. degrees from Beijing Normal University, Beijing, China, in 2013 and 2018, respectively.

She was a Joint Ph.D. Student with the Department of Geographical Sciences, University of Maryland, College Park, MD, USA, in 2017. She was working with Wuhan University, Wuhan, China, from 2018 to 2022. She is currently a Research Assistant Professor with the Department of Geography, The University of Hong Kong, Hong Kong, China. Her main research interests include radiative transfer modeling, development of land surface and atmosphere variables retrieval algorithms using data assimilation, and machine learning techniques, to support global environmental change applications and sustainable development.

Shunlin Liang (Fellow, IEEE) received the Ph.D. degree from Boston University, Boston, MA, USA, in 1993.

He was a Faculty Member with the Department of Geographical Sciences, University of Maryland, College Park, MD, USA, for nearly 30 years. He is currently a Chair Professor with the Department of Geography, The University of Hong Kong, Hong Kong. He has authored over 440 SCI indexed peer-reviewed journal articles, 42 book chapters, and nine special issues of different journals. He has authored/edited eight books and five of which were translated in Chinese, such as *Quantitative Remote Sensing of Land Surfaces* (Wiley, 2004), *Advances in Land Remote Sensing: System, Modeling, Inversion and Application* (Springer, 2008), *Advanced Remote Sensing: Terrestrial Information Extraction and Applications* (Academic Press, 2012, 2019), *Global LAnd Surface Satellite (GLASS) Products: Algorithms, Validation and Analysis* (Springer, 2013), *Land Surface Observation, Modeling, Data Assimilation* (World Scientific, 2013), *Earth's Energy Budget* (Elsevier, 2017), and the *Geography of Contemporary China* (Spring, 2022). He has led to develop the multiple satellite products, including the GLASS products suite that is freely available (www.glass.umd.edu and www.geodata.cn) and has been widely used Worldwide. His main research interests focus on estimating essential environmental variables from satellite data, global high-level satellite product generation, Earth's energy balance, and environmental changes.

Dr. Liang was the Editor-in-Chief of the nine-volume books entitled *Comprehensive Remote Sensing* (Elsevier, 2017). He was an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and is currently the Editor-in-Chief of *Science of Remote Sensing*.
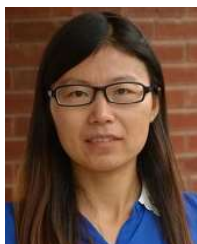
Wei Su received the B.Sc. degree in management and the M.Sc. degree in agronomy from Shandong Agricultural University, Taian, China, in 2001 and 2004, respectively, and the Ph.D. degree in science from the College of Resources, Beijing Normal University, Beijing, China, in 2007.

She is currently a Professor with the College of Land Science and Technology, China Agricultural University, Beijing. She has authored more than 50 articles in the research area in recent years. Her research focuses on LIDAR remote sensing and its application in agriculture, land information monitoring, and quantitative remote sensing inversion for agriculture.