European Journal of Agronomy 139 (2022) 126556

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Assimilating remote sensing-based VPM GPP into the WOFOST model for improving regional winter wheat yield estimation

Wen Zhuo [a], Jianxi Huang [a,b,*], Xiangming Xiao [c,**], Hai Huang [a], Rajen Bajgain [c], Xiaocui Wu [c], Xinran Gao [a], Jie Wang [c], Xuecao Li [a,b], Pradeep Wagle [d]

[a] College of Land Science and Technology, China Agricultural University, Beijing 100083, China
[b] Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100083, China
[c] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA
[d] USDA, Agricultural Research Service, Grazinglands Research Laboratory, El Reno, OK 73036, USA

ARTICLE INFO

Keywords:
Gross primary production
Vegetation photosynthesis model
WOFOST
Data assimilation
Winter wheat yield

ABSTRACT

Crop growth models are powerful tools for predicting crop growth and yield. Gross primary production (GPP) is a major photosynthetic flux that is directly linked to crop grain yield. To better understand the potential of GPP for regional crop yield estimation, in this study, a novel crop data-model assimilation (CDMA) framework was proposed that assimilates accumulative GPP estimates from the satellite-based vegetation photosynthesis model (VPM) into the WOrld FOod STudies (WOFOST) model using the ensemble Kalman filter (EnKF) algorithm to estimate winter wheat GPP and grain yield. Results showed that the WOFOST simulated GPP agreed with the GPP_{EC} derived from eddy flux tower (R^2 = 0.74 and 0.47 in 2015 and 2016, respectively). Assimilating GPP_{VPM} into the WOFOST model improved site-scale GPP estimation (R^2 = 0.87 and 0.67 in 2015 and 2016, respectively), and also improved regional-scale winter wheat yield estimates (R^2 = 0.36 and 0.29; RMSE= 479 and 572 kg/ha in 2015 and 2016, respectively) compared with the open loop simulations (R^2 = 0.14 and 0.10; RMSE= 801 and 788 kg/ha in 2015 and 2016, respectively). Our study demonstrated that assimilation of remotely sensed GPP optimized the results of carbon simulation in the WOFOST model and highlighted the potential of GPP for regional winter wheat yield estimation using a data assimilation framework.

1. Introduction

Winter wheat is the most widely cultivated food crop globally, and one of the most important cereal crops traded in international markets (Becker-Reshef et al., 2010; Fao, 2017; Zhu et al., 2022). Winter wheat production is directly related to winter wheat sale and trade price and plays an important role for food security. Therefore, accurate monitoring of regional winter wheat growth and estimation of winter wheat yield are becoming increasingly crucial for the world's food security and sustainable crop production.

Many previous studies used crop growth model (CGM) to simulate the growth status and grain yields of crops, and the CGM based on a series of input data for relevant physiological characteristics of the crops, as well as other factors, such as soil properties, crop management strategies, and weather conditions throughout the growing season (Huang et al., 2015b). A number of crop models have been developed, and most have been successfully applied to simulate the growth and yield of winter wheat, such as APSIM (Xiao and Tao, 2014), AquaCrop (Iqbal et al., 2014), EPIC (Lu and Fan, 2013), DSSAT (Attia et al., 2016), STICS (Palosuo et al., 2011), and WOFOST (Boogaard et al., 2013). However, crop models were widely used at the point (site) scale (de Wit et al., 2012), making it difficult to extrapolate from those estimates to a regional scale.

This extrapolation to a large area becomes possible using data assimilation methods that integrate regional-scale information from remote sensing observations into process-based crop growth models. This approach offers powerful tools to simulate the physiological development, growth, and yield of a crop, and have been demonstrated to be one of the promising approaches for crop growth monitoring and yield estimation at the regional scale (de Wit and van Diepen, 2007; de Wit et al., 2012; Huang et al., 2015a, 2015b, 2016; Ines et al., 2013; Nearing et al., 2012; Xie et al., 2017). Moreover, important biophysical

* Correspondence to: College of Land Science and Technology, China Agriculture University, No. 17 Qinghua East Road, Haidian District, Beijing 100083, China.
** Corresponding author.
E-mail addresses: jxhuang@cau.edu.cn (J. Huang), xiangming.xiao@ou.edu (X. Xiao).

https://doi.org/10.1016/j.eja.2022.126556
Received 28 November 2021; Received in revised form 22 February 2022; Accepted 24 May 2022
Available online 3 June 2022

parameters, including the leaf area index (LAI), canopy cover and soil moisture can be retrieved from remotely sensed data, and can help in fine-tuning input parameters and state variables by reducing the uncertainty with data assimilation approaches (Weiss et al., 2020).

Researchers have accounted for various state variables into different crop models using data assimilation methods to estimate winter wheat yield at a regional scale, such as LAI (Chen et al., 2018; Dong et al., 2016; Huang et al., 2015b; Silvestro et al., 2017; Xie et al., 2017; Zhuo et al., 2022), soil moisture (SM) (de Wit and van Diepen, 2007; Zhuo et al., 2019), evapotranspiration (ET) (Huang et al., 2015a; Ines et al., 2006; Vazifedoust et al., 2009; Zhuo et al., 2022), above ground biomass (AGB) (Dumont et al., 2014; Jin et al., 2017), and canopy reflectance (Huang et al., 2019b). They have achieved promising results and have improved crop yield estimation. However, how to combine carbon fluxes (e.g., gross primary production (GPP), net primary production (NPP)) within crop data-model assimilation (CDMA) scheme for crop yield estimation need more advanced research. Migliavacca et al. (2009) assimilated MODIS NDVI into a process-based model (BIOME-BGC) by constructing a cost function to simulate GPP, and obtained good accuracy for daily and annual GPP predictions. Combe et al. (2017) used the WOFOST model coupled with a soil respiration function to obtain the net surface CO_2 exchange between croplands and the atmosphere. Their method generated satisfactory daily to multiannual hindcasts of cropland GPP and net ecosystem exchange (NEE) under normal to mild water-stress conditions in Europe. Wang et al. (2013) used simulated annealing algorithm to calibrate the WOFOST model based on in situ observation data (e.g., GPP, LAI and yield) from one sample site. Their study demonstrated successful assimilation of these state variables into the WOFOST model at a local scale. They found a marked enhancement of the accuracy for these state variables and better performance after assimilation. In summary, these studies only assimilated variables (e.g., LAI) into a process-based model to generate CO_2-assimilation related variables (e.g., NEE, GPP) or assimilated carbon flux variables (GPP) at a local scale for crop yield estimation, but such efforts have not been performed at the regional scale for winter wheat yield estimation and the potential of assimilating GPP into crop models has not been fully exploited.

Several remote sensing-based GPP data products have been proposed, such as the MODIS GPP product (Running et al., 2004; Zhao and Running, 2010), vegetation photosynthesis model (VPM) (Xiao et al., 2004a; Zhang et al., 2017b), FLUXCOM (Jung et al., 2017), Support vector regression (SVR) (Kondo et al., 2015), Photosynthesis-respiration (PR) (Keenan et al., 2016), Breathing Earth System Simulator (BESS) (Jiang and Ryu, 2016) and Eddy Covariance Light Use Efficiency (EC-LUE) (Yuan et al., 2019). In the present study, VPM was chosen as the observational GPP simulation model to be coupled with the WOFOST model to construct an improved CDMA framework. One advantage of VPM is its ability to account for the climate variables that are used for light use efficiency (LUE) calculation (Wu et al., 2010). In addition, VPM has shown promising results for GPP estimates for various land cover types, including forest (Xiao et al., 2004a, 2004b, 2005), grasslands (Li et al., 2007; Wagle et al., 2014; Wu et al., 2008), savannas (Jin et al., 2013), upland crops such as maize, winter wheat and soybean (Doughty et al., 2018; Kalfas et al., 2011; Wang et al., 2010), paddy rice (Xin et al., 2017) and inland freshwater wetlands (Kang et al., 2014).

In summary, previous research has not yet fully addressed several key points. First, we'd like to test whether GPP simulated by the WOFOST model can reflect carbon sequestration of crops especially winter wheat, because most crop growth models are designed to simulate crop growth, and focus primarily on crop yield (Wang et al., 2013). Second, previous research examined the most commonly used state variables (e.g., LAI, SM, ET and AGB) in a CDMA scheme, but did not test whether GPP was an effective variable to improve the performance of winter wheat yield estimation at the regional scale. As a consequence, we designed the present study to provide a deeper understanding of these questions by developing a novel CDMA framework that assimilates remotely sensed GPP data derived from VPM into the WOFOST model for improving winter wheat yield estimation. Specifically, we aimed to answer the following three key questions: (1) evaluating the performance of GPP_{WOFOST} in winter wheat fields; (2) at the local scale, assessing the feasibility of assimilating GPP_{VPM} into the WOFOST model and validating the GPP simulation results with and without data assimilation; (3) at the regional scale, estimating regional-scale winter wheat yield, and validate how much accuracy can it improve by assimilating GPP_{VPM} into the WOFOST model.

2. Materials and methods

2.1. Study area

The state of Oklahoma is located in the south-central United States, where it covers an area of 181,000 km^2. We used this state as a case study of our data-assimilation method. Oklahoma has a temperate climate, and experiences the occasional extremes of temperature and precipitation that are typical of a continental climate. Oklahoma's main winter wheat production areas are located in the west, which has a semi-arid climate (Fig. 1). In the western areas, the average annual temperature is approximately 14 ℃, with an annual rainfall of approximately 500 mm (Illston et al., 2004). An eddy covariance flux tower site (35.56850° N, 98.05580° W) was established at the United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Grazinglands Research Laboratory (GRL) in El Reno, Oklahoma (Fig. 1). The study site was a single-crop winter wheat site and the collected flux data were used to calculate the site-scale GPP. In Fig. 1, the 30 m winter wheat map came from Cropland Data Layer (CDL, https://www.nass.usda.gov), and the 500 m winter wheat map, which was corresponding to the CDMA assimilation unit, was resampled from the CDL and have the same spatial reference with MODIS data.

Winter wheat is a cool season crop and is the primary agricultural crop grown in Oklahoma. Winter wheat in this study area was planted from early September to October. About 90% of the winter wheat emerges by mid-November (USDA-NASS, 2014). Jointing typically occurs from early March to mid-April, followed by the flowering until late April. Grain-filling begins from late April to May, and harvest usually occurs in early June (Zhang et al., 2017a).

2.2. Data

2.2.1. Eddy covariance flux tower data from the winter wheat site

An eddy covariance (EC) tower was installed to measure the CO_2, H_2O and energy fluxes (Bajgain et al., 2018). The tower uses a three-dimensional sonic anemometer (CSAT3) and an open path infrared gas analyzer (LI-COR 7500) to measure NEE for CO_2 between the crop, the soil, and the atmosphere. The measured NEE was gap filled and then partitioned into ecosystem respiration (ER) and GPP based on the short-term temperature sensitivity of the ER (Lloyd and Taylor, 1994; Reichstein et al., 2005). The R package "REddyProc" tool was used for calculation of the tower scale GPP (Moffat et al., 2007; Reichstein et al., 2005). More details regarding this method can be found at https://www.bgcjena.mpg.de/bgi/index.php/Services/REddyProcWeb-RPackage.

2.2.2. Oklahoma agricultural statistical data

The Oklahoma agricultural data included statistical data from the United States Department of Agriculture (USDA), the National Agricultural Statistics Service (NASS) and a cropland area dataset from Cropland Data Layer (CDL). The annual county level statistics of winter wheat planted area, harvest area, winter wheat yield in 2015 and 2016 were collected from the USDA-NASS (https://quickstats.nass.usda.gov/). The annual CDL data at a 30-m resolution in 2015 and 2016 were also from the USDA (https://www.nass.usda.gov). CDL is a remote

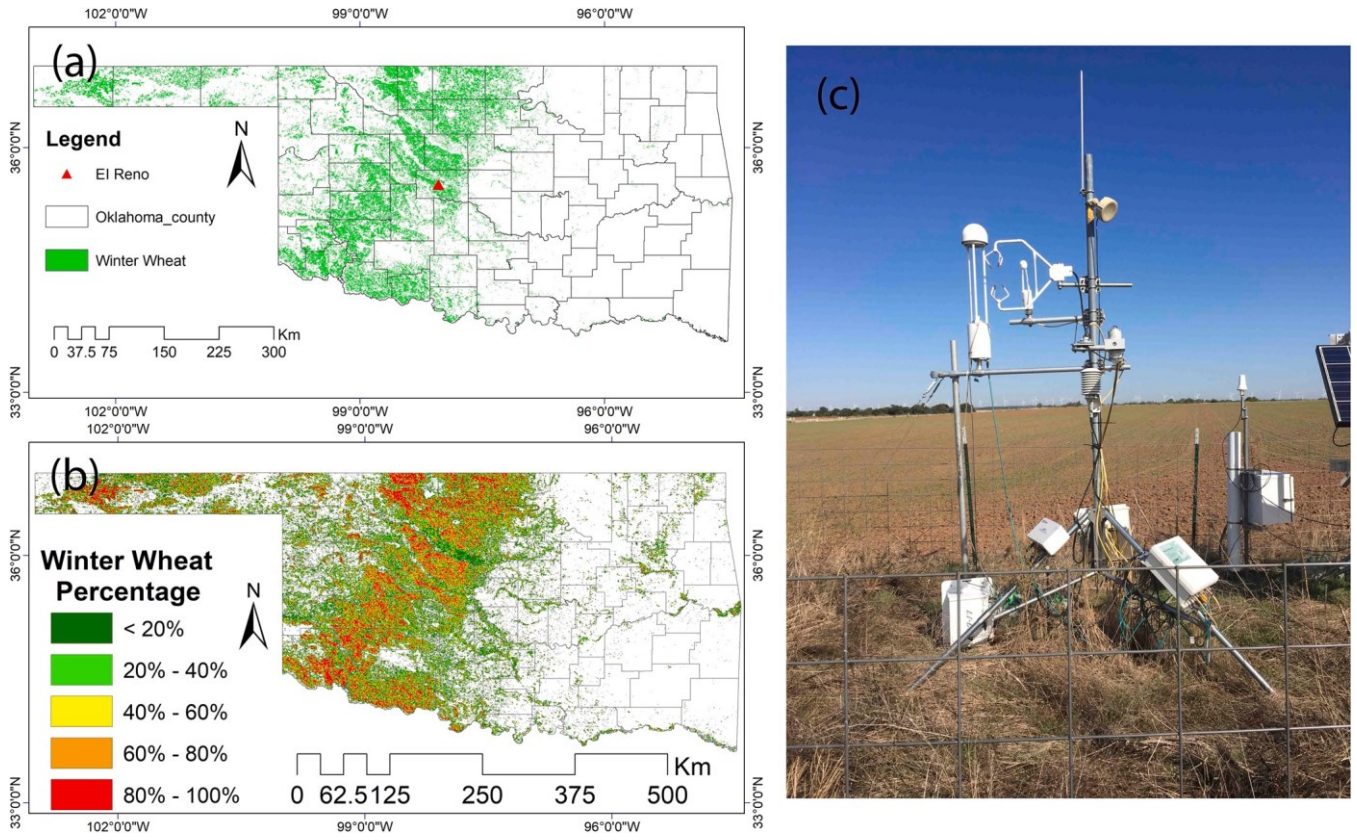

Fig. 1. Maps of the study area. (a) winter wheat distribution (at 30-m spatial resolution); (b) winter wheat percentage of pixel area map within individual MODIS pixels (at 500-m spatial resolution) in 2015; (c) Eddy covariance flux tower at El Reno site.

sensing-based land cover product and has more than 100 types of crops with classification accuracy greater than 90% for major crops such as maize, soybean and winter wheat (Boryan et al., 2011). The Landcover map of Oklahoma State generated from CDL is provided in the Supplementary.

2.2.3. Gross primary production data from VPM, the MOD17A2 and MOD15A2 product

The VPM model was employed to estimate the 8-day GPP with 500 m spatial resolution. It is constructed upon the concept that a vegetation canopy is composed of non-photosynthetic vegetation and chlorophyll (Xiao et al., 2004a, 2005). The primary input datasets for the VPM included MODIS (MOD09A1, MYD11A2 and MCD12Q1), NCEP-reanalysis II (temperature and radiation) (https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html), Earth Stat (major crop type distribution), and ISLSCP II (C4 vegetation percentage map). These datasets were used to calculate the temperature, LUE, vegetation indices and photosynthetically active radiation (PAR), and then generate the GPP time series. The global GPP_{VPM} dataset is available at https://doi.org/10.1594/PANGAEA.879560 (Zhang et al., 2017b).

The 8-day MOD17A2H version 6 GPP product (GPP_{MOD17}) and MOD15A2H version6 LAI product (LAI_{MODIS}) with 500 m spatial resolution from 1 January 2015–31 December 2016 was also used in this study. GPP_{MOD17} was used to provide a comparison with the different GPP results. This product also uses a LUE model to estimate GPP, one of the major differences between GPP_{VPM} and GPP_{MOD17} is that VPM uses the fraction of PAR absorbed by chlorophyll (Xiao et al., 2004a), whereas the MOD17A2H product uses the fraction of PAR absorbed by the canopy (Running et al., 2004). LAI_{MODIS} was used to be assimilated into the WOFSOT model for comparing with the GPP CDMA simulation results.

2.3. The vegetation photosynthesis model (VPM)

The VPM model is a light (energy) use efficiency model based on satellite remote sensing data and flux observations. It was developed to estimate GPP during the photosynthetically active period of vegetation as the product of PAR absorbed by chlorophyll ($APAR_{chl}$) and light use efficiency (Xiao et al., 2004a). VPM can be described as follows:

$$GPP = \varepsilon_g \times APAR_{chl} \quad (1)$$

$$APAR_{chl} = FPAR_{chl} \times PAR \quad (2)$$

$$\varepsilon_g = \varepsilon_0 \times T_{scalar} \times W_{scalar} \quad (3)$$

Where ε_g is the light use efficiency ($\mu molCO_2/\mu mol$, PPFD), PAR is photosynthetically active radiation (μmol photosynthetic photon flux density, PPFD), $APAR_{chl}$ is the amount of PAR absorbed by chlorophyll, $FPAR_{chl}$ is the fraction of PAR absorbed by chlorophyll, which is estimated as a linear function of Enhanced Vegetation Index (EVI). The light use efficiency (ε_g) is affected by temperature and water, and it can be calculated by Eq. (3). Where ε_0 is the apparent quantum yield or maximum light use efficiency, and T_{scalar} and W_{scalar} are the scalars for the effects of temperature and water, respectively.

2.4. WOFOST model

The WOFOST (WOrld FOod STudies) model (de Wit, 1965; Boogaard et al., 2013) estimates crop growth given a set of crop, soil and meteorological parameters. The major processes are phenological development, CO_2-assimilation, transpiration, respiration, partitioning of assimilates among various organs, and dry matter formation. The model can run in a potential mode (with no yield limitations caused by water and nutrient stress), a water-limited mode (with water stress), or a

nutrient-limited mode (with nutrient stress). We used the water-limited mode because the winter wheat in the study area is not irrigated. The general structure of the WOFOST model and potential daily gross CO_2 assimilation rate calculation in the WOFOST model are provided in the Supplementary.

The input data for the WOFOST model included climate, crop and soil parameters. The daily climate data used for the point scale WOFOST model in this study were obtained from the Oklahoma Mesonet (https://www.mesonet.org), which consists of 121 meteorological stations across Oklahoma. The stations measure ten variables every 15 min including soil moisture, humidity, soil temperature, pressure, solar radiation, rainfall, temperature, leaf wetness, wind speed and direction (Brock et al., 1995). For the regional scale WOFOST model, we interpolated data for the six required daily-scale WOFOST weather variables (maximum temperature, minimum temperature, wind speed, solar radiation, precipitation and pressure) to create a 0.1° × 0.1° grid. In this study, the WOFOST input data were calibrated by field measurement, published values and the WOFOST default values. In addition, we assumed that Oklahoma state only planted one dominant cultivar. Ma et al. (2013) and Huang et al. (2015b) describe details of the parameterization and calibration of WOFOST for winter wheat. The water-limited mode of the WOFOST model was used in this study, and the primary crop and soil parameters are shown in the supplementary file. Table 1 shows the spatial and temporal resolution of different input parameters used in this study.

2.5. The Ensemble Kalman Filter

Evensen (1994) developed the Ensemble Kalman Filter (EnKF) method, which is based on forecasting the error statistics using Monte Carlo methods. Burgers et al. (1998) showed that the addition of random perturbations to the measurements was essential for calculation of the analyzed ensemble, because it would have too low of a variance unless the random perturbations were added to the observations. In the present study, we assumed that the observation errors followed a Gaussian distribution. The core of the EnKF method is the Kalman filter. If the observations are related to the true data at time t x_t,

$$y_t = Hx_t + \varepsilon_t \tag{4}$$

$$x^m_t = Ax^m_{t-1} \tag{5}$$

$$x_t = x^f_t + v_t \tag{6}$$

Table 1
The spatial and temporal resolution of different input data/parameters.

Model	Input data	Spatial resolution	Temporal resolution
WOFOST model (**Output:** daily outputs at point scale)	Crop parameters	Point scale	
	Soil parameters	Point scale	
	Weather data	10 km (interpolated from weather station)	Daily
	Management data	Point scale	
VPM model (**Output:** 8-day GPP with 500 m resolution)	MOD09A1	500 m	8-day
	MCD12Q1	500 m	Annual
	MYD11A2	1 km	8-day
	NCEP-reanalysis II	~1.875° × 2°	Daily
	Earth Stat	0.083°	invariant
	ISLSCP II	1° × 1°	invariant
CDL map		30 m	Annual
Winter wheat pixel (resampled from CDL)		500 m	Annual

Where y is the observation vector (which can be regarded as GPP_{VPM} in this study), and H is the observation operator that relates to y (which can be taken as an identity matrix in this study), A represents a state-transition model that links x_t and x_{t-1} (which can be regarded as WOFOST model in the CDMA framework), ε and ν are Gaussian random error vector with a mean of zero (which can be regarded as error of VPM and WOFOST model in this study). x^m_t is represent the model simulation result at the time t (which can be regarded as GPP_{WOFOST}).

So, in this study, the optimal estimation $x^{EnKF}_{t=k}$ of x_t at $t = k$ can be calculated using Eqs. (7) and (8):

$$x^{EnKF}_{t=k} = (I - KH)x^m_{t=k} + Ky_{t=k} \tag{7}$$

$$P^{EnKF}_{t=k} = (I - KH)P^m_{t=k} \tag{8}$$

Where $P^m_{t=k}$ is the WOFOST model state covariance at time k, I and H is the identity matrix, and K is the Kalman gain matrix, which is defined as follows:

$$K = P^m_{t=k} H^T (HP^m_{t=k} H^T + R_t)^{-1} \tag{9}$$

Where R_t is the covariance of VPM at the time t. When solving the Kalman gain equation, Houtekamer and Mitchell (2001) suggest calculating $P^m_{t=k}H^T$ and $HP^m_{t=k}H^T$ directly from the ensemble members, rather than calculating each element of Eq. (9):

$$P^m H^T = (N_e - 1)^{-1} \sum_{n=1} (x^m_n - \bar{x}^m)(Hx^m_n - \overline{Hx^m})^T \tag{10}$$

$$HP^m H^T = (N_e - 1)^{-1} \sum_{n=1} (Hx^m_n - \overline{Hx^m})(Hx^m_n - \overline{Hx^m})^T \tag{11}$$

$$R_t = (N_e - 1)^{-1} \sum_{n=1} (o_n - \bar{o})(o_n - \bar{o})^T \tag{12}$$

Where N_e is the number of ensemble members, n is a running index for ensemble member, and $\bar{x}^m$ represents the GPP_{WOFOST} ensemble mean calculated as (13), $\bar{o}$ represents the GPP_{VPM} ensemble mean calculated as (15).

$$\bar{x}^m = N_e^{-1} \sum_{n=1} x^m_n \tag{13}$$

$$H\bar{x}^m = N_e^{-1} \sum_{n=1} Hx^m_n \tag{14}$$

$$\bar{o} = N_e^{-1} \sum_{n=1} o_n \tag{15}$$

2.6. General framework

The CDMA framework was developed to assimilate GPP into WOFOST using the EnKF algorithm (Fig. 2). The framework consists of three parts: VPM, WOFOST and the data assimilation using EnKF. In the VPM component, ISLSCP II, Earth Stat, NCEP reanalysis II and MODIS data were used to calculate the light use efficiency and the absorbed PAR by chlorophyll. Then a time series of the GPP was generated.

In the WOFOST component, the model was driven by crop, soil and meteorological calibrated parameters, which is calibrated by MCMC method, and model ensembles were generated. We used field measurements of LAI and AGB observations to parameterize the likelihood function that represented the mismatch between the WOFOST output and the measurements. The observations were assumed to be corrupted by noise that followed a normal distribution with a mean of zero and a 10% variance. We used a Markov-chain Monte Carlo (MCMC) method

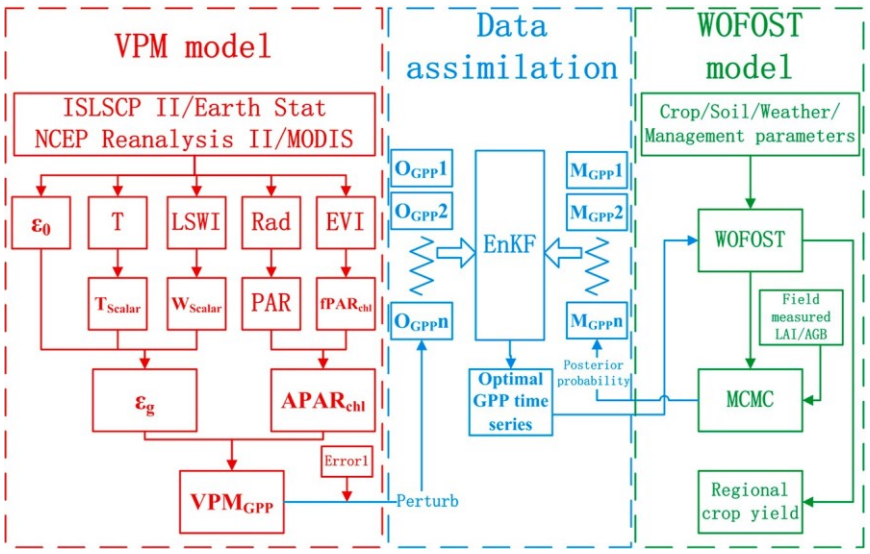

Fig. 2. Flowchart for the assimilation of GPP_{VPM} into the WOFOST model using the EnKF-based assimilation algorithm. Notes: ISLSCP II (International Satellite Land Surface Climatology Project, Initiative II); Earth Stat (global major crop types distribution); NCEP (national centers for environmental prediction); MODIS (moderate resolution imaging spectroradiometer); ε_0 (maximum light use efficiency); T (temperature); LSWI (land surface water index); Rad (radiation); EVI (enhanced vegetation index); T_{Scalar} (temperature limitation for photosynthesis); W_{Scalar} (water limitation for photosynthesis); PAR (photosynthetically active radiation); $fPAR_{chl}$ (fraction of PAR absorbed by chlorophyll); ε_g (light use efficiency); $APAR_{chl}$ (absorbed PAR by chlorophyll); Error1 (GPP error in VPM model); $O_{GPP}1..n$ (GPP_{VPM} ensemble); $M_{GPP}1..n$ (GPP_{WOFOST} ensemble); Tmax (maximum temperature); SM (soil moisture).

(specifically, the DiffeRential Evolution Adaptive Metropolis algorithm, [DREAM]; Vrugt et al., 2009) to update three crop parameters (TSUM1 (the sum of the effective temperatures from emergence to anthesis), SPAN (the lifespan of leaves growing at 35 ℃, in days) and TDWI (the total initial dry weight of the crop)) of WOFOST from a uniform prior to a posterior distribution that is based on the likelihood function. First, we ran the DREAM sampler algorithm with 4 parallel chains until we achieved a stationary distribution, in which the convergence statistic R hat (Brooks and Gelman, 1998) was less than 1.05 or after reaching the maximum 50,000 repetitions. Then another 50 repetitions were run, which essentially represented a sample from the posterior distribution. Finally, the 50 corresponding WOFOST simulated GPP formed the ensemble.

In this study, the ensemble number was set as 50 and the uncertainty of the GPP_{VPM} was set to 8% based on the results of Zhang et al. (2017b). The WOFOST model ran with a daily time step until a new observation (8-day GPP_{VPM} from Jan 1st to May 10th) became available and the run was interrupted; the EnKF (the data assimilation component) was then used to update the target model state variable (GPP); and the WOFOST model was then run using the updated states (optimal GPP). Data assimilation was conducted for as long as observation data was available. Because the daily GPP is a rate variable rather than a state variable, we used accumulated GPP (GPP_A) instead of daily GPP (GPP_D) in the CDMA framework. GPP_A can be calculated by Eq. (16):

$$GPP_A = \sum_{D=1}^{D=N} GPP_D \tag{16}$$

Where GPP_A is accumulated GPP, GPP_D is daily GPP, N is the day.

3. Results

3.1. WOFOST simulated LAI and AGB at the eddy covariance flux tower site

Fig. 3 shows the simulation and validation with the calibrated WOFOST model using the field measured LAI and AGB at the El-Reno site. The LAI of winter wheat increased slowly from emergence to wintering, and almost stopped growing during the wintering stage. It then began to grow rapidly during the green-up stage and reached its maximum value prior to flowering stage. Then, it decreased gradually from the flowering stage to maturity stage, and decreased rapidly after maturity stage. The time series for LAI and AGB simulated by WOFOST agree well with the measured LAI and AGB. As shown in Fig. 3, the coefficient of determination (R^2) of the simulated and observed LAI in 2015 and 2016 were 0.82 and 0.77, respectively, with RMSE of 0.78 and 0.77 m^2/m^2, respectively; the R^2 of AGB was 0.94 and 0.93, respectively, with RMSE of 121.25 and 81.06 g/m^2, respectively. Results indicated that the WOFOST model was well calibrated and could simulate the winter wheat growth status at the study area properly.

Fig. 4-1(a) and (b) compare the different GPP products. The GPP trends followed a similar overall pattern. During the winter wheat growing season in both 2015 and 2016, GPP started to increase in mid-March and reached a peak in early April. In addition, EVI, which can indicate the health, density and growth of vegetation, also had good consistency with the GPP results. This reflects the authenticity of the different GPP products. Among the four GPP products, GPP_{MOD17} had the lowest value during most days of the growing season, whereas GPP_{WOFOST} was slightly higher than GPP_{VPM} and GPP_{EC} during the peak growing season (from late March to late April). The scatterplots in Fig. 4-1 show good relationship between GPP_{WFOSOT} and GPP_{EC} with R^2 of 0.74 and 0.47 in 2015 and 2016, respectively; the corresponding RMSE were 3.65 and 4.12 g $Cm^{-2}day^{-1}$ in 2015 and 2016, respectively. The accuracy of the WOFOST simulated GPP was better in 2015 than that in 2016. As shown in Fig. 4-1(b), GPP_{VPM} and GPP_{MOD17} had a peak during late April to early May in 2016 which were inconsistent with GPP_{EC} and GPP_{WOFOST}. This might be due to the impact of pasture in the pixel where eddy covariance flux tower located.

Before early March in 2015 and late February in 2016, the value of GPP_{WOFOST} was consistently 0, which was generally because of the wintering stage of winter wheat. During this stage, the average temperature was low (Fig. 4-2(a) and (b)), and winter wheat will have little or no photosynthesis. In 2016, GPP_{WOFOST} during two time periods (approximately early March and late March) also had a value of 0. This was due to the effect of low temperature limitations in the WOFOST model, in which an average minimum temperature for the previous seven days lower than 0 ℃ causes the model to assume that photosynthesis does not occur. These time periods are indicated in gray Fig. 4-2 (b).

3.2. Assimilation of GPP_{VPM} into WOFOST at the eddy tower site

The performance of the WOFOST simulated GPP after assimilation (GPP_{EnKF}) was compared with GPP_{EC}, GPP_{VPM} and GPP_{WOFOST} during the winter wheat growing season at the point scale. To eliminate the value of 0 for GPP_{WOFOST} in the EnKF data assimilation, we assigned these values of GPP_{VPM} to those GPP_{WOFOST} which value equal to 0. In contrast, if the GPP_{WOFOST} was greater than 0, the GPP_{EnKF} was

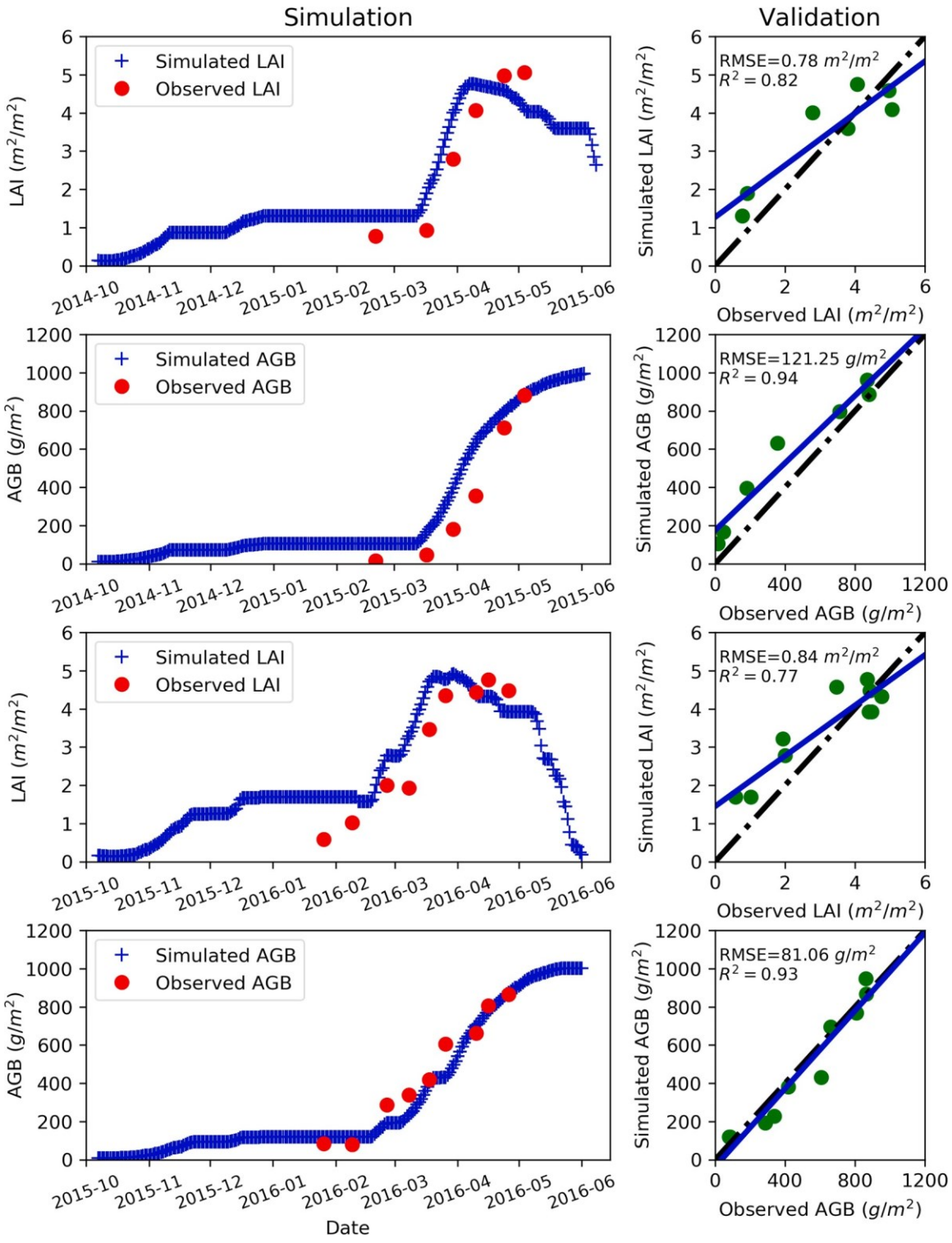

Fig. 3. Validation of the simulated WOFOST LAI and AGB with field measured data in 2015 and 2016.

calculated from GPP_{WOFOST} and GPP_{VPM} using EnKF algorithm. Fig. 5 compares the time series for the four GPP products. The accumulated GPP increased slowly before early March, followed by an increase during March and late April. From early May, the increase of accumulated GPP gradually decreased and finally became stable. Fig. 6 shows the validation of GPP_{VPM}, GPP_{EnKF} and GPP_{WOFOST} compared with GPP_{EC}. The winter wheat GPP before DOY 60 (purple circle) and after DOY 140 (dark green circle) was generally lower than 10, whereas the main period of GPP simulation is during DOY 60–140, when is corresponding to jointing, flowering and grain filling stages of winter wheat. Compared with GPP_{WOFOST}, GPP_{VPM} performed better at the El-Reno site both in 2015 (R^2 = 0.85, RMSE = 2.01 g $Cm^{-2}day^{-1}$) and 2016 (R^2 = 0.56, RMSE = 3.86 g $Cm^{-2}day^{-1}$). Assimilation GPP_{VPM} improved the GPP estimates (GPP_{EnKF} in 2015: R^2 = 0.87, RMSE = 2.26 g $Cm^{-2}day^{-1}$; GPP_{EnKF} in 2016: R^2 = 0.67, RMSE =3.25 g $Cm^{-2}day^{-1}$) compared with the values estimated without data assimilation (GPP_{WOFOST} in 2015: R^2 = 0.74, RMSE = 3.65 g $Cm^{-2}day^{-1}$; GPP_{WOFOST} in 2016: R^2 = 0.47, RMSE = 4.12 g $Cm^{-2}day^{-1}$). Furthermore, the R^2 of GPP_{EnKF} improved

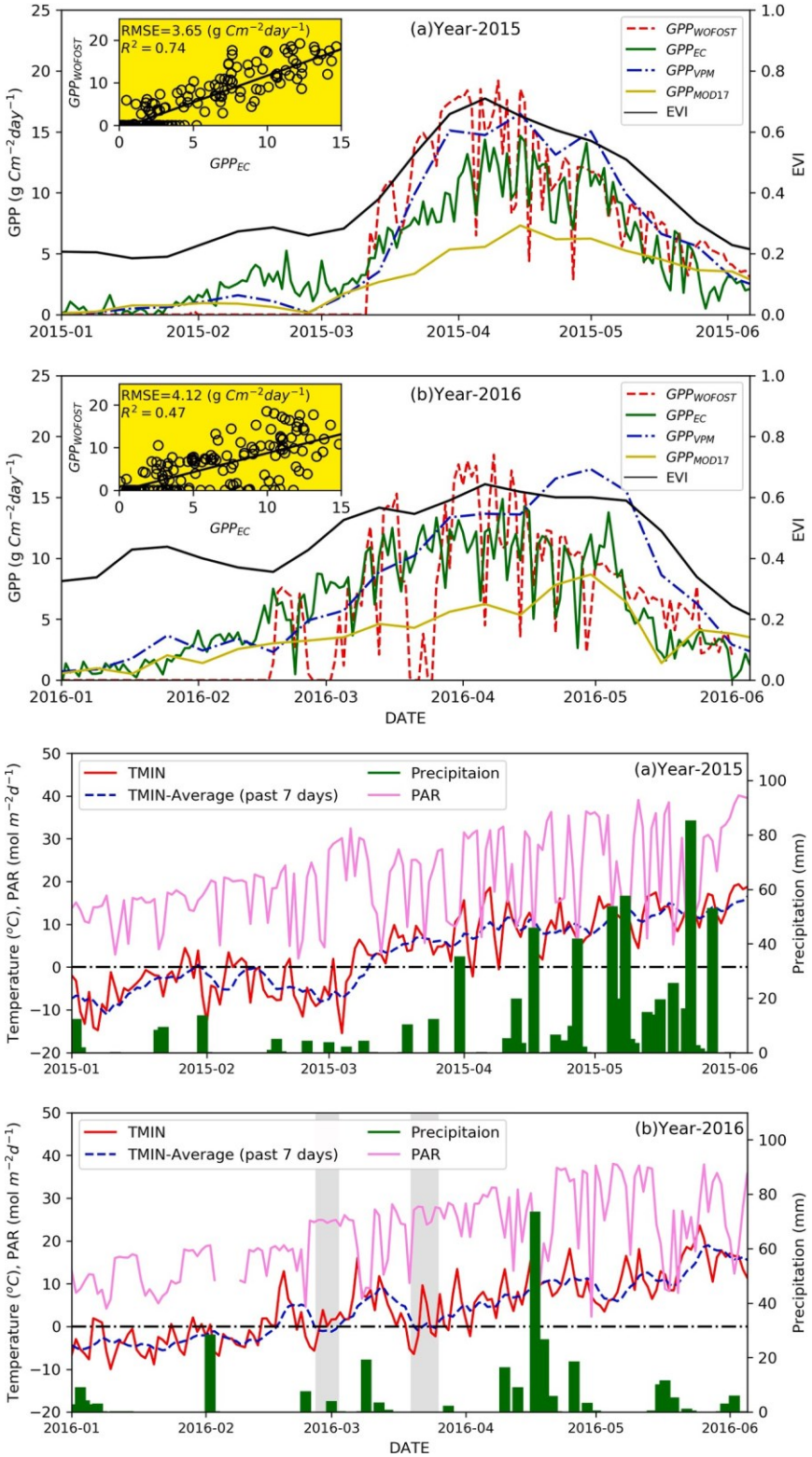

Fig. 4. -1. Comparison among different daily GPP time series products at the El-Reno flux tower site in (a) 2015 and (b) 2016. **4-2**. Daily minimum temperature (TMIN) and averaged minimum temperature (TMIN-Average) at El-Reno flux tower site in (a) 2015 and (b) 2016. (Gray areas represent TMIN-Average lower than 0 ℃ after the wintering stage).

compared with GPP_{VPM} both in 2015 and 2016, while RMSE was increased slightly in 2015 and decreased in 2016. Results showed that assimilating GPP_{VPM} into the WOFOST model have potential for improving the GPP simulation at the site level.

The accumulated GPP and yield results at the El Reno site were also compared. Fig. 7 shows that GPP_{MOD17} had the lowest values during winter wheat growing season in both years, whereas GPP_{VPM} had the highest accumulated GPP in both years. The accumulated GPP_{EC} and GPP_{WOFOST} were close to each other. For the winter wheat yield, $Yield_{EnKF}$ was higher than $Yield_{WOFOST}$ in both years, and was closer to

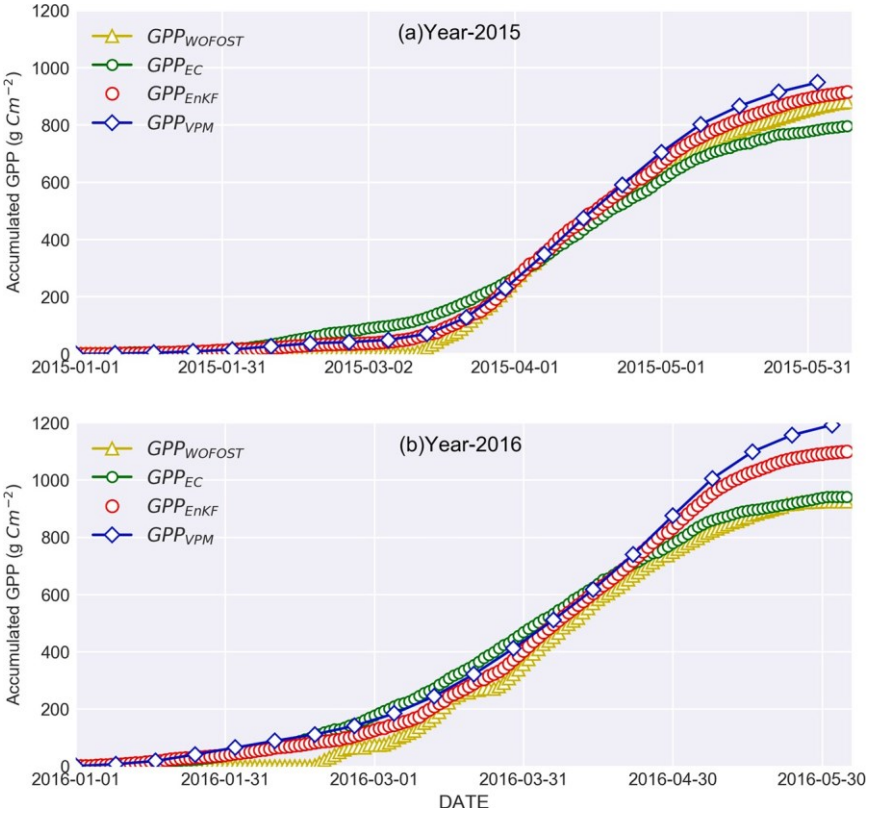

Fig. 5. Comparison of accumulated GPP time series simulated by WOFOST with and without EnKF-based assimilation method during winter wheat growing season (from Jan. 1st to Jun. 5th) at El-Reno flux tower site.

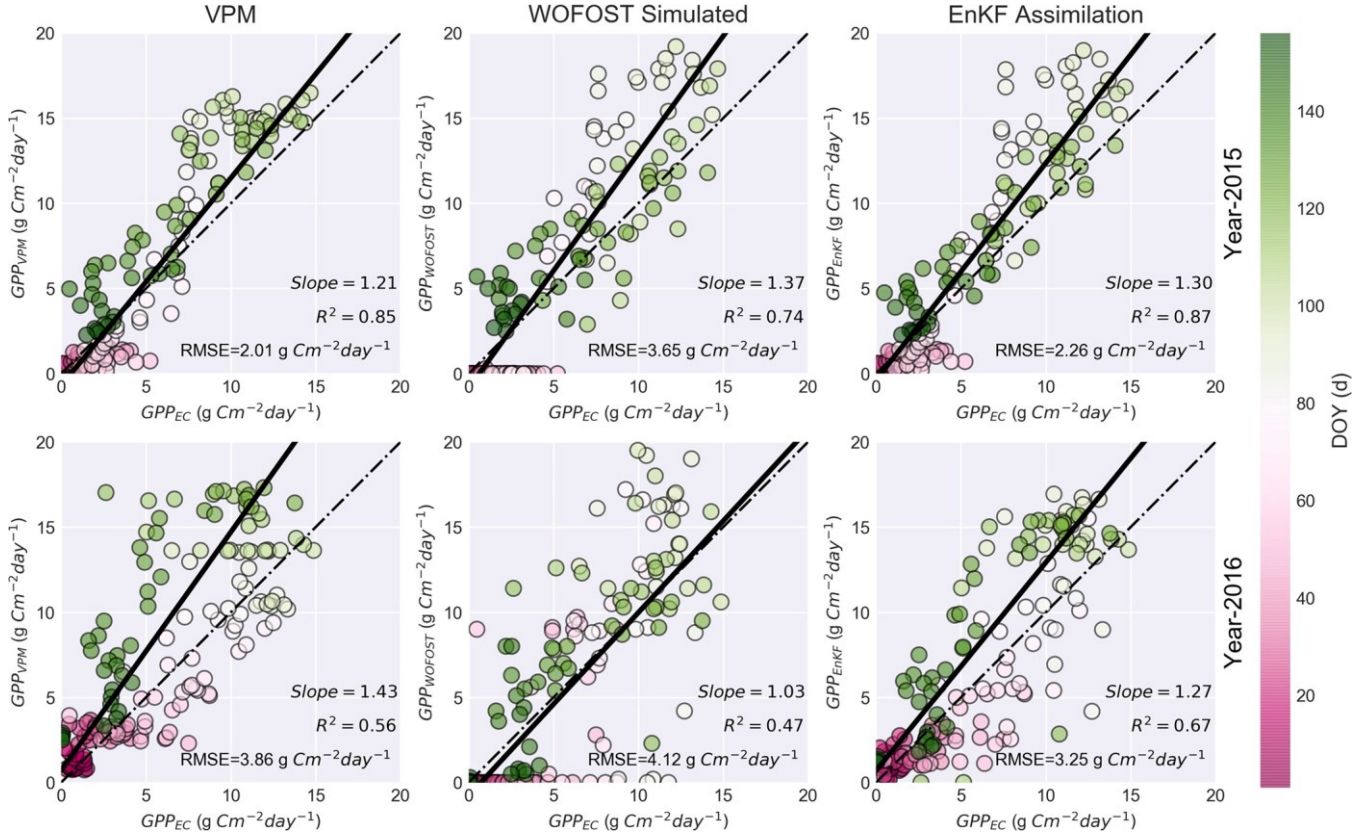

Fig. 6. Validation of daily GPP simulated by WOFOST with and without EnKF-based assimilation method and VPM compared with the GPP_{EC} at El-Reno flux tower site.

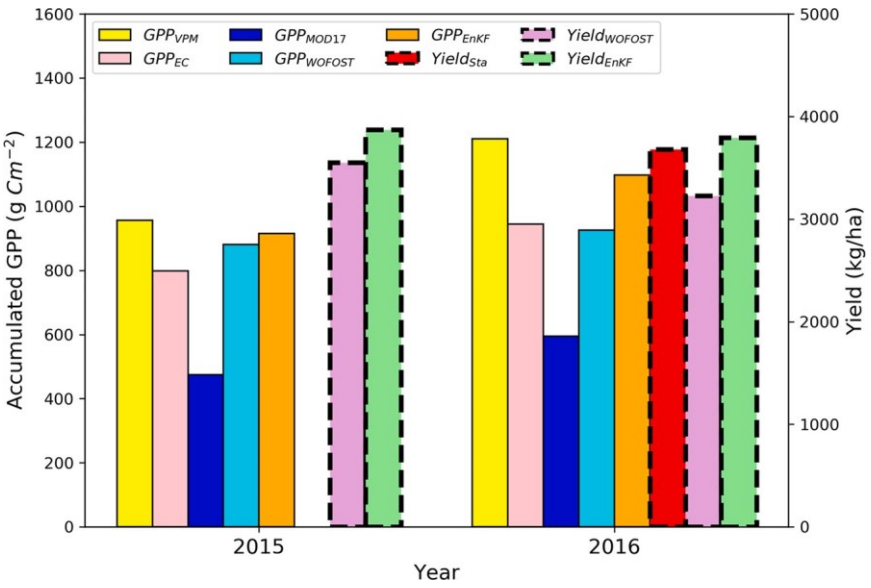

Fig. 7. Comparison of accumulated GPP during winter wheat growing season (from Jan. 1st to Jun. 5th) and yield at El-Reno flux tower site.

$Yield_{meas}$ in 2016. (Yield measurements from the EI Reno site were not available in 2015.).

3.3. Assimilation of GPP_{VPM} into WOFOST at the regional scale

The regional scale data assimilation was conducted in 500 m grid cells in which the spatial resolution coincides with GPP_{VPM}. Fig. 8 shows the resulting regional winter wheat yields in both years based on both government statistics (Fig. 8(a), (b)) and our simulations (Fig. 8(c-j)). From Fig. 8(c) and (d) can be seen that the spatial variation in the un-assimilated winter wheat yield maps was small, especially in 2016, and yield mostly ranged from 1500 to 2500 kg/ha. This was mainly due to the coarse spatial resolution of the meteorological parameters. The yield maps with assimilation (Fig. 8(g), (h)) showed more spatial variance, and the spatial pattern agreed well with official yield statistics at the county level (Fig. 8(i) and (j)). Moreover, we also found a mismatch in winter wheat yield between 2015 and 2016. The county-level statistical data showed that the winter wheat yield of most counties was higher in 2016 than 2015 (Fig. 8(a), (b)), whereas the open loop simulated yield (Fig. 8(e) and (f)) showed the opposite result especially in central and eastern Oklahoma. One reason might be that we only used one set of crop parameters to simulate all of Oklahoma, and Fig. 7 showed slightly higher WOFOST simulated yield at the flux tower site in 2015 than in 2016. After we assimilated GPP_{VPM} into WOFOST, the simulated winter wheat yield results (Fig. 8(i) and (j)) agreed better than open loop results with the official statistics. This demonstrates that the assimilation of GPP_{VPM} into WOFOST improved the estimation of winter wheat yield.

Fig. 9 shows the validation results based on a comparison of the simulated values with the official regional statistical data. The circle size represents the variance of the pixel yield in the county, the larger the circle the greater the variance of pixel yield in the county. Assimilation of GPP improved the winter wheat yield estimation in both years. R^2 increased from 0.14 and 0.10 without assimilation in 2015 and 2016, respectively, to 0.36 and 0.29 with assimilation, and the RMSE decreased from 801 and 788 kg/ha to 479 and 572 kg/ha in 2015 and 2016, respectively. However, the county statistical winter wheat yields were mostly lower than the WOFOST simulated yields and were lower than that in North China Plain, where the official winter wheat yield ranges from 4000 to 8000 kg/ha (Huang et al., 2015b, 2016, 2019a). One possible reason is that the winter wheat in Oklahoma is mostly rainfed, whereas in Northern China, winter wheat is irrigated 2–4 times during the growing season. This is important, since winter wheat is more sensitive to water stress during the jointing and grain-filling stage (Huang et al., 2018), which means that winter wheat in Oklahoma is more likely to suffer from water stress that can easily cause lower yield.

4. Discussion

4.1. Assimilation of remotely sensed GPP into process-based crop model

GPP is widely used to understand an ecosystem's carbon cycle and its relationship with the climate system. However, it has not often been used for crop yield estimation in a CDMA framework; instead, studies have generally relied on LAI or SM. One of the primary reasons for this is that GPP and crop yield are secondary variables that are only indirectly accessible from remotely sensed observations. In contrast, LAI or SM are considered to be primary variables directly involved in the radiative process (Weiss et al., 2020).

In this study, we developed a GPP CDMA framework to improve winter wheat yield estimation at a regional scale. Our results suggested that the WOFOST simulated GPP time series achieved high accuracy of simulating the actual carbon sequestration processes in this winter wheat field, and it also agreed well with eddy covariance tower simulated GPP (Fig. 4(a) and (b)). Moreover, assimilating GPP_{VPM} into the WOFOST model using the EnKF algorithm eliminated zero values from GPP_{WOFOST}, which resulted from WOFOST's assumption that photosynthesis will not occur when the air temperature is below 0 ℃. Particularly, the GPP CDMA framework improved GPP estimation at the point scale (Fig. 6) and improved winter wheat yield estimation compared to WOFOST simulation without assimilation at the regional scale (Fig. 9). These results demonstrated the high potential to use GPP as a state variable in the CDMA framework.

To compare the performance of assimilating LAI (the most commonly used state variable in CDMA) and GPP for winter wheat yield estimation, a same CDMA experiment, which using the EnKF assimilation method, was conducted by using MODIS LAI product, and the comparison results were shown in Fig. 10. Assimilating GPP_{VPM} performed better than LAI_{MODIS} both in 2015 and 2016 (R^2 = 0.36 and 0.29; RMSE= 479 and 572 kg/ha in 2015 and 2016, respectively, for assimilating GPP. R^2 = 0.25 and 0.19; RMSE= 511 and 770 kg/ha in 2015 and 2016, respectively, for assimilating LAI). The yield estimation results obtained by assimilating GPP_{VPM} were generally higher than assimilating LAI_{MODIS}, and assimilating GPP_{VPM} can get a higher average yield value (Fig. 10(c)). The main reason is that LAI_{MODIS} always

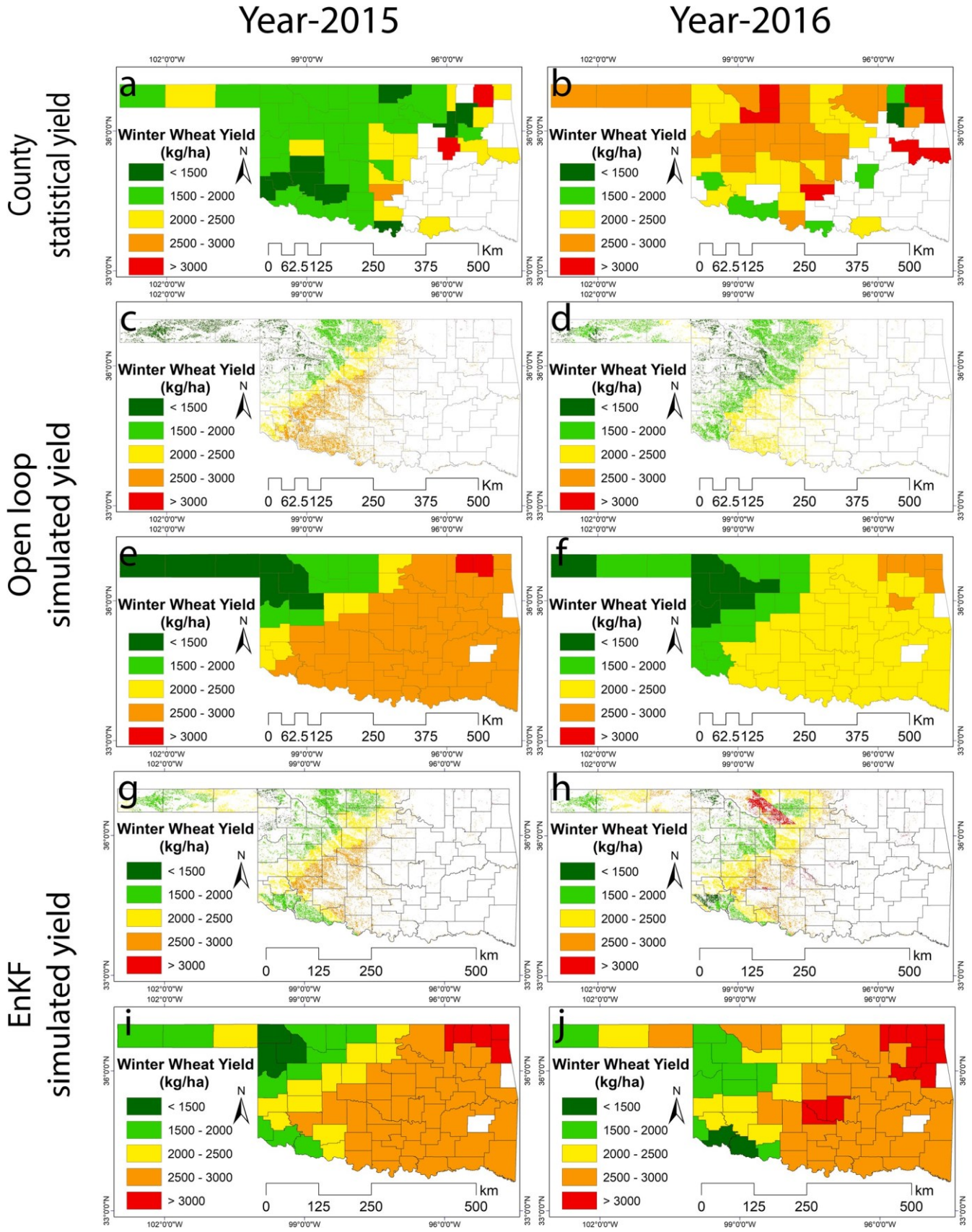

Fig. 8. Regional winter wheat yield estimation without and with EnKF assimilation method. Note: (a) and (b) represent county official statistical yield in 2015 and 2016, respectively; (c) and (d) represent WOFOST model simulated yield without assimilation in 2015 and 2016, respectively; (e) and (f) represent zonal mean statistics at a county level based on (c) and (d), respectively; (g) and (h) represent WOFOST model simulated yield with EnKF assimilation in 2015 and 2016, respectively; (i) and (j) represent zonal mean statistics at a county level based on (g) and (h), respectively.

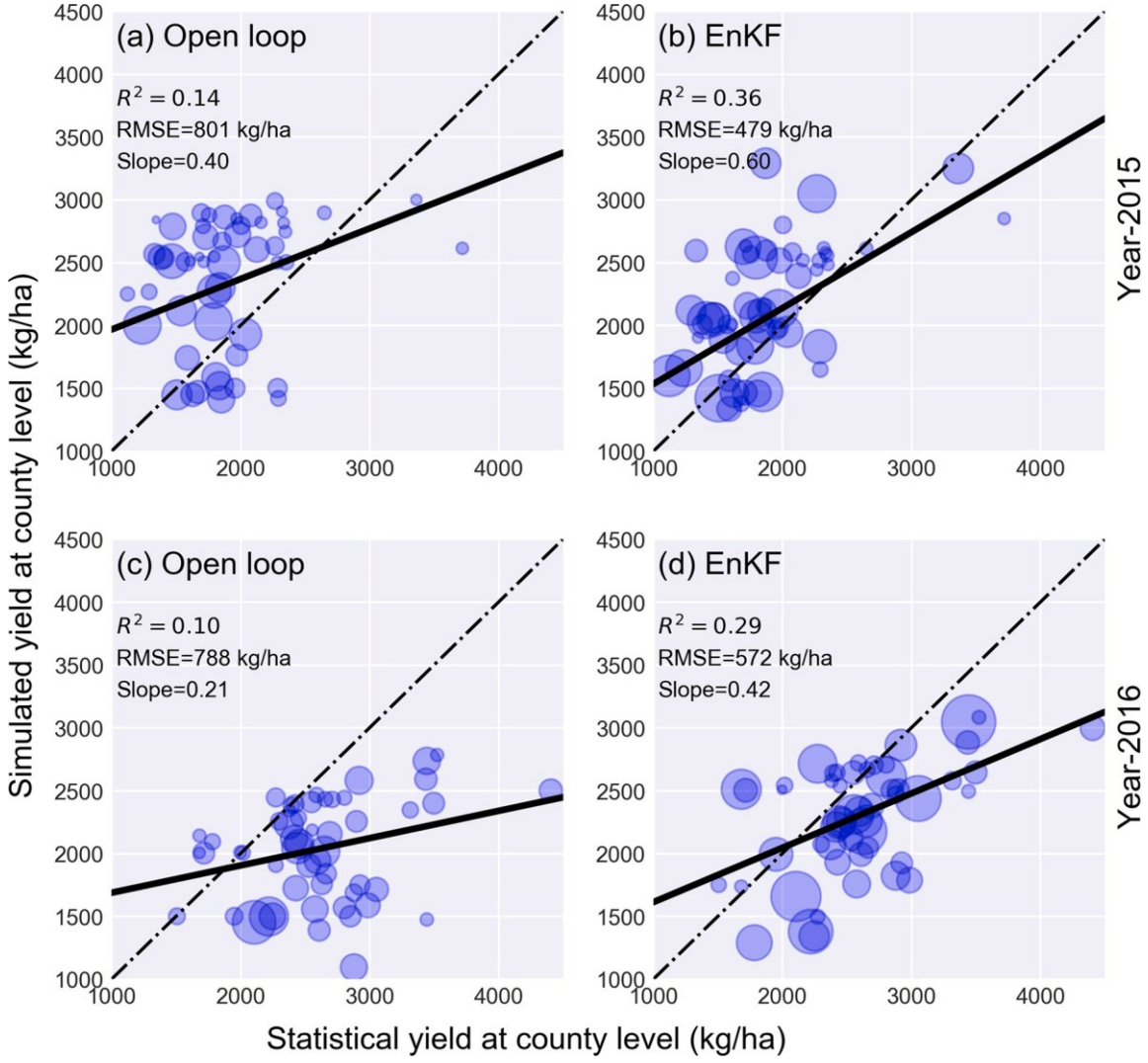

Fig. 9. Comparison of the estimated winter wheat yield with statistical data at a county level. Comparison results are for the open loop in (a) 2015 and (c) 2016, and using EnKF assimilation in (b) 2015 and (d) 2016.

underestimates winter wheat LAI, whose highest value ranges from 5 to 7 m^2/m^2 in actual growth state while LAI_{MODIS} ranges from 2 to 3 m^2/m^2 (Fig. 11). Due to the low value of LAI_{MODIS}, LAI_{EnKF} (red line) was generally lower than LAI_{WOFOST} (yellow line) (Fig. 11). Several previous studies proven that assimilating LAI_{MODIS} into crop model directly may reduce the crop yield estimation results, due to the lower LAI values caused by scaling effect (Dente et al., 2008; Huang et al., 2015b, 2016, 2022; Ma et al., 2013;). At present, the commonly used strategies of assimilating LAI_{MODIS} are (1) to stretch the LAI_{MODIS} value to the normal range by data fusion methods (Huang et al., 2016) or (2) assimilating the trend information of LAI_{MODIS} using variational method (Zhuo et al., 2020). In summary, assimilating GPP_{VPM} is able to obtain higher accuracy than LAI_{MODIS} using EnKF method for winter wheat yield estimation at the regional scale.

4.2. Comparison with the linear regression model using harvest index for winter wheat yield estimation

Harvest index (HI) is calculated as the ratio between crop grain yield and crop aboveground biomass (AGB), or net primary production (NPP) or gross primary production (GPP) (Wu et al., 2021a). In this study, we used the accumulated GPP (GPP_A) over winter wheat growing season as a dependent variable for winter wheat yield estimation using HI. Numerous researches reported the HI values range from 0.2 to 0.6 for different crops (Guan et al., 2016; Lobell et al., 2002; Monfreda et al., 2008; Wu et al., 2021a, 2021b), and we followed results concluded by Wu et al. (2021b) that set the HI as 0.25. The validation results of winter wheat yield and production at the county level were presented in Fig. 12.

The validation results showed that winter wheat yield estimated by CDMA ($Yield_{CDMA}$) had higher accuracy (R^2 = 0.36 and 0.29; RMSE= 479 and 572 kg/ha in 2015 and 2016, respectively) than by HI ($Yield_{HI}$) (R^2 = 0.28 and 0.19; RMSE= 423 and 682 kg/ha in 2015 and 2016, respectively), and larger spatial differences were found for the $Yield_{CDMA}$ ranging 1000 – 3500 kg/ha (Fig. 12(a), (b)). The winter wheat production validation results were performed well for both $Production_{CDMA}$ and $Production_{HI}$, in which R^2 were generally bigger than 0.8 and RMSE were generally lower than 55 kton, due to the high precision winter wheat mapping results (CDL, https://www.nass.usda.gov). HI method is feasible and easy to operate, which only costs about a few minutes for Oklahoma while CDMA will cost more than 10 h for calculating. However, HI is also affected by many factors, such as crop varieties, crop management, and environmental conditions (Wu et al., 2021b), and a constant HI could cause large uncertainties for crop yield estimation at the regional scale. In general, CDMA, which comprehensively considers crop, soil, management and weather conditions, is more

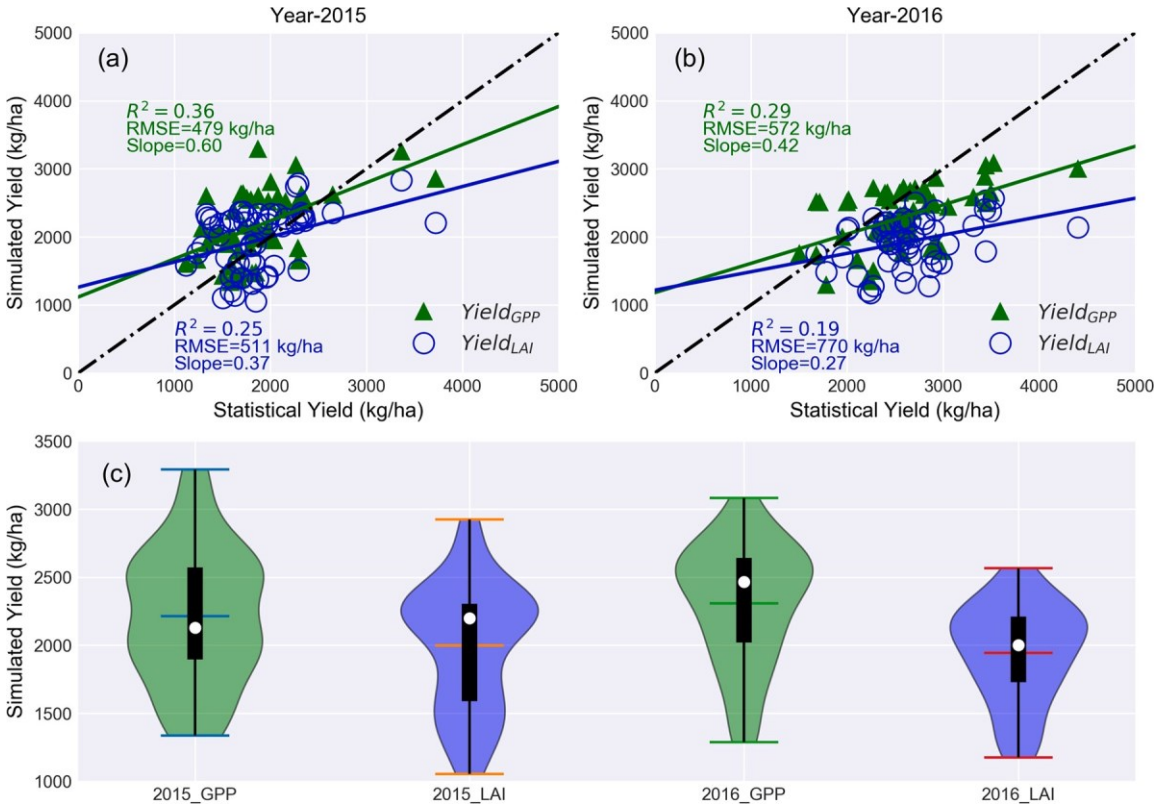

Fig. 10. Comparison of the estimated county-level winter wheat yield by assimilating GPP_{VPM} and LAI_{MODIS} in (a) 2015 and (b) 2016, and comparison of yield distribution (c).

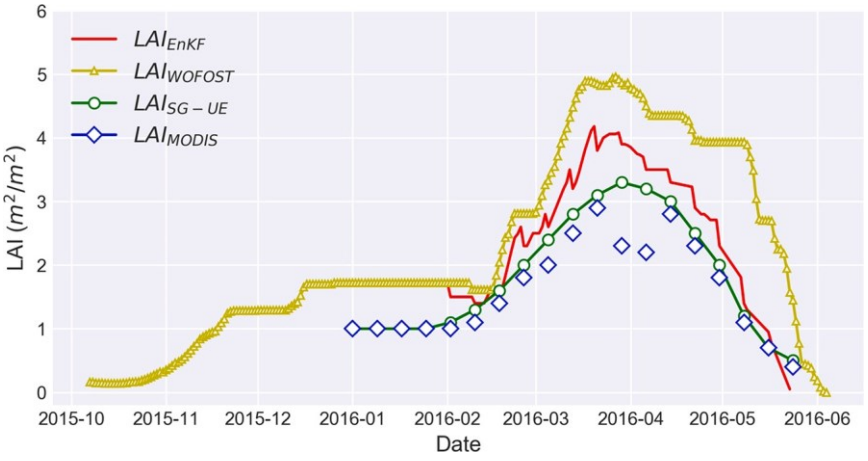

Fig. 11. Comparison of LAI simulated by WOFOST with and without EnKF-based assimilation method during winter wheat growing season at the site level. (LAI_{EnKF} represent the LAI value simulated by WOFOST with EnKF method; LAI_{WOFOST} represent the LAI value simulated by WOFOST; LAI_{MODIS} represent MODIS LAI product; LAI_{SG-UE} represent the LAI_{MODIS} filtered by upper enveloped SG filter method).

suitable for winter wheat yield estimation, and can obtain higher winter wheat yield estimation accuracy than HI method.

4.3. Uncertainty of the GPP CDMA framework and future development

Uncertainty of GPP observations is a potentially important source of uncertainty in GPP CDMA framework. This is because the estimation of GPP is still highly uncertain by now, although the related theories have made great progress in recent years (Ryu et al., 2019). In the present study, the error related to GPP_{VPM} came from the following sources. First, VPM integrates coarse spatial resolution data into its calculation, such as the NCEP reanalysis II meteorological data, which has an original resolution of approximately 1.875° × 2° and is primary forcing data for the estimation of LUE and PAR. This approach is feasible at the global scale, but it may lead to large errors when it is applied at local and regional scales. Second, estimation of the maximum LUE can contain significant errors. The maximum LUE values at the ecosystem level vary among vegetation types and may also differ between regions for a given ecosystem type (Xiao et al., 2011). In this study, the maximum LUE was estimated by using the look up table method, and the LUE value of winter wheat was set to 0.42 g C/mol $PPFD^{-1}$ (Zhang et al., 2017b). Thus, if the maximum LUE values for different crops can be specified in

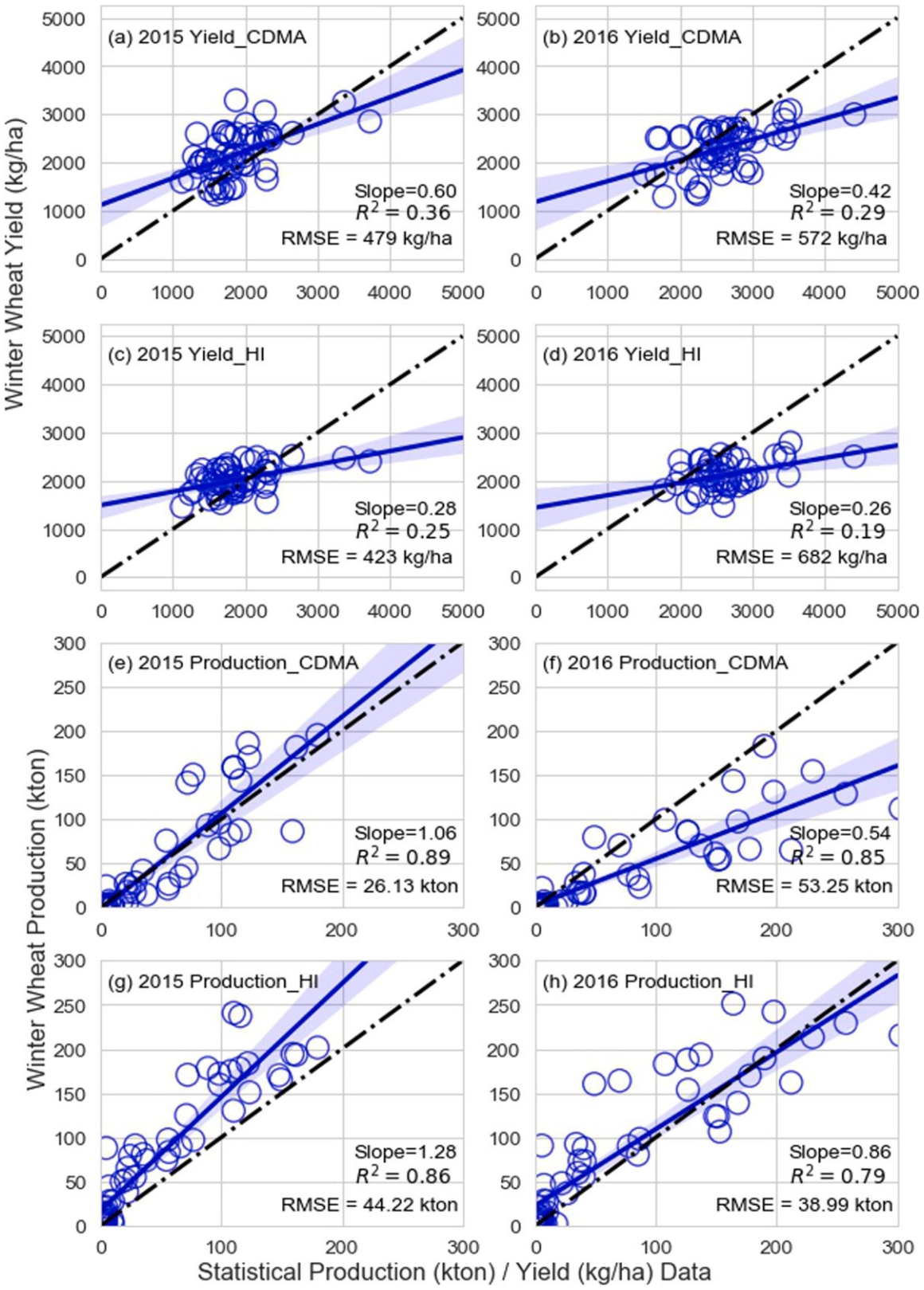

Fig. 12. Comparison of the estimated winter wheat yield/production with statistical data at a county level. (a) winter wheat yield estimation using GPP CDMA method in 2015, (b) winter wheat yield estimation using GPP CDMA method in 2016, (c) winter wheat yield estimation HI method in 2015, (d) winter wheat yield estimation using HI method in 2016, (e) winter wheat production estimation using GPP CDMA method in 2015, (f) winter wheat production estimation using GPP CDMA method in 2016, (g) winter wheat production estimation HI method in 2015, and (h) winter wheat production estimation using HI method in 2016.

the future, the estimation of GPP will improve. Third, the uncertainties in VPM also come from the two down-regulation factors (W_{scalar} and T_{scalar}). Jin et al. (2013) performed a sensitivity analysis for VPM and found that it was important to account for both water and temperature to downscale the maximum LUE when estimating GPP in semi-arid climates, because the accuracy of both factors will directly affect the accuracy of GPP.

As for the WOFSOT model, only one set of crop parameters (calibrated by the wheat variety of Gallagher) was used in this study throughout the Oklahoma winter wheat area, which is difficult to represent the spatial variations. It was assumed that a single dominant cultivar was planted throughout the study area and that the crop characteristics and management measures did not vary spatially. However, more than 30 varieties of wheat are planted throughout Oklahoma and they differ in crop characteristics, and the most planted variety is Gallagher, which accounts for 14% of the total (ODAFF, 2017). To account for this variation, future research should regionalize the crop parameters. Secondly, the mismatch in spatial resolution among the dataset from the eddy covariance tower (at a point scale), the crop model (at a point scale) and the GPP_{VPM} pixels (500-m spatial resolution) will also affect the accuracy of the GPP_{EnKF} assimilation.

The data assimilation method used in this study was EnKF, while the standard EnKF method has its own assumptions and uncertainties. It tends to reject observations in the late period of CDMA, which is referred to "filter divergence" (Ines et al., 2013). To reduce the effect of the filter divergence, inflation factor can be conducted in future studies to enlarge the variance of the forecast ensemble (Huang et al., 2016; Lin et al., 2008). Besides, an important assumption of the EnKF is that both observation and model errors are Gaussian, while the reality is not the case. Particle Filter (PF) algorithm, which can assume that the error is non-Gaussian distribution, is a potential sequential assimilation method for CDMA (Moradkhani and Weihermüller, 2011; Nagarajan et al., 2011).

In the future study, GPP data at a finer spatial resolution should be applied to the GPP CDMA framework to reduce the spatial mismatch between crop models and GPP observations. Besides, given the current availability of Solar-induced chlorophyll fluorescence (SIF) observations, which can provide a direct link to the instantaneous photosynthetic activity (Guan et al., 2016; Guanter et al., 2014; Macbean et al., 2018), researchers begin to use SIF data for improving estimates of GPP (Bacour et al., 2019; Koffi et al., 2015; Macbean et al., 2018; Norton et al., 2018a, 2018b), which proved that the remote sensing derived SIF can largely reduce the uncertainty of GPP estimation (Norton et al., 2018b) and can further advance our understanding and help with a more accurate estimation of GPP.

5. Conclusions

In this study, the coupled VPM-WOFOST CDMA framework was used to estimate winter wheat yield in Oklahoma (USA) in 2015 and 2016 by assimilating 500-m-scale GPP_{VPM} product into the WOFOST model using EnKF algorithm. At the point scale, the WOFOST simulated GPP (GPP_{WOFOST}) had a good relationship with GPP_{EC} and can reflect winter wheat carbon sequestration well. Moreover, the VPM-WOFOST CDMA framework can improve winter wheat GPP estimation at the point scale. At the regional scale, assimilating GPP_{VPM} into the WOFOST model slightly improved the winter wheat yield estimation, this may be due to the scale mismatch and the uncertainty of remotely sensed GPP. However, this study is a successful first attempt that demonstrates the potential of GPP assimilation for regional-scale winter wheat yield estimation within a CDMA framework. In addition, it improves our understanding of the importance of carbon flux processes of crops in crop yield estimation.

CRediT authorship contribution statement

Wen Zhuo: Conceptualization, Methodology, Software, Writing – original draft. **Jianxi Huang**: Conceptualization, Supervision, Resources, Writing – review & editing. **Xiangming Xiao**: Conceptualization, Resources, Writing – review & editing. **Hai Huang**: Software, Writing – review & editing. **Rajen Bajgain**: Software, Formal analysis. **Xiaocui Wu**: Data curation, Visualization. **Xinran Gao**: Resources, Visualization. **Jie Wang**: Data curation, Visualization. **Xuecao Li**: Writing – review & editing. **Pradeep Wagle**: Data curation.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Project No. 41971383), Science and Technology Facilities Council of UK-Newton Agritech Programme (Project No. ST/N006798/1), the USDA National Institute of Food and Agriculture (NIFA) (Project No. 2013-69002 and 2016-68002-24967) and the China Scholarship Council (No. 201806350193). We thank professor Yao Zhang at Peking University for improving the manuscript. We are grateful to the Earth Observation and Modeling Facility (EOMF) members for their help and advice on this research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.eja.2022.126556.

References

Attia, A., Rajan, N., Xue, Q., Nair, S., Ibrahim, A., Hays, D., 2016. Application of DSSATCERES-Wheat model to simulate winter wheat response to irrigation management in the Texas High Plains. Agric. Water Manag 165, 50–60.

Bacour, C., Maignan, F., Macbean, N., Porcar-Castell, A., Flexas, J., Frankenberg, C., Peylin, P., Chevallier, F., Vuichard, N., Bastrikov, V., 2019. Improving estimates of Gross Primary Productivity by assimilating solar-induced fluorescence satellite retrievals in a terrestrial biosphere model using a process-based SIF model. J. Geophys. Res. Biogeosci. 124 (11), 3281–3306.

Bajgain, R., Xiao, X., Basara, J., Wagle, P., Zhou, Y., Mahan, H., Gowda, P., Mccarthy, H. R., Northup, B., Neel, J., Steiner, J., 2018. Carbon dioxide and water vapor fluxes in winter wheat and tallgrass prairie in central Oklahoma. Sci. Total Environ. 644, 1511–1524.

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens. Environ. 114 (6), 1312–1323.

Boogaard, H., Wolf, J., Supit, I., Niemeyer, S., Van Ittersum, M., 2013. A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. Field Crop Res. 143 (2013), 130–142.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int. 26, 341–358.

Brock, F.V., Crawford, K.C., Elliott, R.L., Cuperus, G.W., Stadler, S.J., Johnson, H.L., Eilts, M.D., 1995. The Oklahoma Mesonet: a technical overview. J. Atmos. Ocean. Technol. 12 (1), 5–19.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7 (4), 434–455.

Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126, 1719–1724.

Chen, Y., Zhang, Z., Tao, F., 2018. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. Eur. J. Agron. 101, 163–173.

Combe, M., De Wit, A.J., Vila`-Guerau de Arellano, J., van der Molen, M.K., Magliulo, V., Peters, W., 2017. Grain yield observations constrain cropland CO_2 fluxes over Europe. J. Geophys. Res. Biogeosci. 122 (12), 3238–3259.

De Wit, A., Duveiller, G., Defourny, P., 2012. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. Agric. For. Meteorol. 164, 39–52.

De Wit, A.D., Van Diepen, C., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Agric. For. Meteorol. 146, 38–56.

De Wit, C.T., 1965. Photosynthesis of Leaf Canopies. Pudoc, Wageningen.

Dente, L., Satalino, G., Mattia, F., Rinaldi, M., 2008. Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. Remote Sens. Environ. 112, 1395–1407.
Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., Shang, J., 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. Int. J. Appl. Earth Obs. Geoinf. 49, 63–74.
Doughty, R., Xiao, X., Wu, X., Zhang, Y., Bajgain, R., Zhou, Y., Qin, Y., Zou, Z., Mccarthy, H., Friedman, J., Wagle, P., 2018. Responses of gross primary production of grasslands and croplands under drought, pluvial, and irrigation conditions during 2010–2016, Oklahoma, USA. Agric. Water Manag. 204, 47–59.
Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J.-P., Destain, M.-F., 2014. Parameter identification of the STICS crop model, using an accelerated formal MCMC approach. Environ. Model. Softw. 52, 121–135.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99, 10143–10162.
Fao, 2017. The future of food and agriculture – trends and challenges.
Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Glob. Chang. Biol. 22 (2), 716–726.
Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.-E., Moran, M.S., Ponce-Campos, G., Beer, C., Campsvalls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proc. Natl. Acad. Sci. USA 111, E1327–E1333.
Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129 (1), 123–137.
Huang, J., Go´mez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.H., Wu, Y., Zhao, F., Wang, J., Xie, X.H., 2019a. Assimilation of remote sensing into crop growth models: current status and perspectives. Agric. For. Meteorol. 276–277.
Huang, H., Huang, J., Li, X., Zhuo, W., Wu, Y., Niu, Q., Su, W., Yuan, W., 2022. A dataset of winter wheat aboveground biomass in China during 2007–2015 based on data assimilation. Sci. Data 9 (200), 1–12. https://doi.org/10.1038/s41597-022-01305-6.
Huang, J., Ma, H., Sedano, F., Lewis, P., Liang, S., Wu, Q., Su, W., Zhang, X., Zhu, D., 2019b. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-prosail model. Eur. J. Agron. 102, 1–13.
Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., Wu, W., 2015a. Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 8 (8), 4060–4071.
Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., Wu, W., 2016. Assimilating a synthetic Kalman flter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. Agric. For. Meteorol. 216, 188–202.
Huang, J., Tian, L., Liang, S., Becker-Reshef, I., Su, W., Zhang, X., Zhu, D., Wu, W., 2015b. 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. 204, 106–121.
Huang, J., Zhuo, W., Li, Y., Huang, R., Sedano, F., Su, W., Dong, J., Tian, L., Huang, Y., Zhu, D., Zhang, X., 2018. Comparison of three remotely sensed drought indices for assessing the impact of drought on winter wheat yield. Int. J. Digit. Earth 11 (13), 1–23.
Illston, B.G., Basara, J.B., Crawford, K.C., 2004. Seasonal to interannual variations of soil moisture measured in Oklahoma. Int. J. Climatol. A J. R. Meteorol. Soc. 24 (15), 1883–1896.
Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149–164.
Ines, A.V., Honda, K., Gupta, A.D., Droogers, P., Clemente, R.S., 2006. Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. Agric. Water Manag. 83 (3), 221–232.
Iqbal, M.A., Shen, Y., Stricevic, R., Pei, H., Sun, H., Amiri, E., Penas, A., Del Rio, S., 2014. Evaluation of the FAO AquaCrop model for winter wheat on the North China Plain under deficit irrigation from field experiment to regional yield simulation. Agric. Water Manag. 135, 61–72.
Jiang, C., Ryu, Y., 2016. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). Remote Sens. Environ. 186, 528–547.
Jin, C., Xiao, X.M., Merbold, L., Arneth, A., Veenendaal, E., Kutsch, W.L., 2013. Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. Remote Sens. Environ. 135, 189–201.
Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, X., Wang, J., Li, X., Luo, J., 2017. 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. Photogramm. Remote Sens. 126, 24–37.
Jung, M., Reichstein, M., Schwalm, C.R., Huntingford, C., Sitch, S., Ahlstro¨m, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A.K., Kato, E., Papale, D., Poulter, B., Raduly, B., Ro¨denbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., Zeng, N., 2017. Compensatory water effects link yearly global land CO_2 sink changes to temperature. Nature 541, 516–520.
Kalfas, J.L., Xiao, X.M., Vanegas, D.X., Verma, S.B., Suyker, A.E., 2011. Modeling gross primary production of irrigated and rain-fed maize using MODIS imagery and CO_2 flux tower data. Agric. For. Meteorol. 151, 1514–1528.
Kang, X., Wang, Y., Chen, H., Tian, J., Cui, X., Rui, Y., Zhong, L., Kardol, P., Hao, Y., Xiao, X., 2014. Modeling carbon fluxes using multi-temporal MODIS imagery and CO_2 eddy flux tower data in Zoige Alpine Wetland, South-West China. Wetlands 34, 603–618.
Keenan, T.F., Prentice, I.C., Canadell, J.G., Williams, C.A., Wang, H., Raupach, M., Collatz, G.J., 2016. Recent pause in the growth rate of atmospheric CO_2 due to enhanced terrestrial carbon uptake. Nat. Commun. 7, 13428.
Koffi, E.N., Rayner, P.J., Norton, A.J., Frankenberg, C., Scholze, M., 2015. Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. Biogeosciences 12 (13), 4067–4084.
Kondo, M., Ichii, K., Takagi, H., Sasakawa, M., 2015. Comparison of the data-driven topdown and bottom-up global terrestrial CO_2 exchanges: GOSAT CO_2 inversion and empirical eddy flux upscaling. J. Geophys. Res. Biogeosci. 120, 1226–1245.
Li, Z.Q., Yu, G.R., Xiao, X.M., Li, Y.N., Zhao, X.Q., Ren, C.Y., Zhang, L.M., Fu, Y.L., 2007. Modeling gross primary production of alpine ecosystems in the Tibetan Plateau using MODIS images and climate data. Remote Sens. Environ. 107, 510–519.
Lin, C., Wang, Z., Zhu, J., 2008. An ensemble Kalman filter for severe dust storm data assimilation over China. Atmos. Chem. Phys. 8 (11), 2975–2983.
Lloyd, J., Taylor, J., 1994. On the temperature dependence of soil respiration. Funct. Ecol. 8 (3), 315–323.
Lobell, D.B., Hicke, J.A., Asner, G.P., Field, C., Tucker, C., Los, S., 2002. Satellite estimates of productivity and light use effciency in United States agriculture, 1982-98. Glob. Chang. Biol. 8, 722–735.
Lu, C., Fan, L., 2013. Winter wheat yield potentials and yield gaps in the North China Plain. Field Crop Res. 143, 98–105.
Ma, H.Y., Huang, J.X., Zhu, D.H., Liu, J.M., Zhang, C., Su, W., Fan, J.L., 2013. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD into WOFOST–ACRM model. Math. Comput. Model. Dyn. Syst. 58 (3–4), 753–764.
Macbean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., Ko¨hler, P., Go´mez-Dans, J., Disney, M., 2018. Author Correction: Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. Sci. Rep. 8, 10420.
Migliavacca, M., Meroni, M., Busetto, L., Colombo, R., Zenone, T., Matteucci, G., Manca, G., Seufert, G., 2009. Modeling gross primary production of agro-forestry ecosystems by assimilation of satellite-derived information in a process-based model. Sensors 9 (2), 922–942.
Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Beckstein, C., Braswell, B.H., Churkina, G., Desai, A.R., Falge, E., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric. For. Meteorol. 147 (3–4), 209–232.
Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob. Biogeochem. Cycles 22 (1).
Moradkhani, H., Weihermüller, L., 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399 (3), 410–421.
Nagarajan, K., Judge, J., Graham, W.D., Monsivais-Huertero, A., 2011. Particle filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions. Adv. Water Resour. 34 (4), 433–447.
Nearing, G.S., Crow, W.T., Thorp, K.R., Moran, M.S., Reichle, R.H., Gupta, H.V., 2012. Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: an observing system simulation experiment. Water Resour. Res. 48, W05525.
Norton, A.J., Rayner, P.J., Koffi, E.N., Scholze, M., Silver, J., Wang, Y., 2018a. Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. Biogeosci. Discuss. 2018, 1–40.
Norton, A.J., Rayner, P.J., Koffi, E.N., Scholze, M., 2018b. Assimilating solar-induced chlorophyll fluorescence into the terrestrial biosphere model BETHY-SCOPE v1.0: model description and information content. Geosci. Model Dev. 11 (4), 1517–1536.
ODAFF (Oklahoma Department of Agriculture, Food, and Forestry), 2017. Oklahoma Agricultural Statistics 2017. Oklahoma City, OK.
Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Taka´´c, J., Trnka, M., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35 (3), 103–114.
Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Chang. Biol. 11, 1424–1439.
Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M.S., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience 54, 547–560.
Ryu, Y., Berry, J.A., Baldocchi, D.D., 2019. What is global photosynthesis? History, uncertainties and opportunities. Remote Sens. Environ. 223, 95–114.
Silvestro, P.C., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. Remote Sens. 9 (5), 509.
Usda-Nass, 2014. Crop progress and condition. USDA's National Agricultural Statistics Service Oklahoma Field Office. ⟨https://www.nass.usda.gov/Statistics_by_State/Oklahoma/Publications/Crop_Progress_&_Condition/2014/index.php⟩ .
Vazifedoust, M., Dam, Van, Bastiaanssen, J., Feddes, R, W., 2009. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. Int. J. Remote Sens. 30 (10), 2523–2545.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlinear Sci. Numer. Simul. 10 (3), 273–290.

Wagle, P., Xiao, X., Torn, M.S., Cook, D.R., Matamala, R., Fischer, M.L., Jin, C., Dong, J., Biradar, C., 2014. Sensitivity of vegetation indices and gross primary production of tallgrass prairie to severe drought. Remote Sens. Environ. 152, 1–14.

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

Wang, Z., Xiao, X., Yan, X., 2010. Modeling gross primary production of maize cropland and degraded grassland in northeastern China. Agric. For. Meteorol. 150, 1160–1167.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402.

Wu, C., Munger, J.W., Niu, Z., Kuang, D., 2010. Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. Remote Sens. Environ. 114 (12), 2925–2939.

Wu, W.X., Wang, S.Q., Xiao, X.M., Yu, G.R., Fu, Y.L., Hao, Y.B., 2008. Modeling gross primary production of a temperate grassland ecosystem in Inner Mongolia, China, using MODIS imagery and climate data. Sci. China Ser. D Earth Sci. 51, 1501–1512.

Wu, X.C., Xiao, X., Yang, Z., Wang, J., Steiner, J., Bajgain, R., 2021a. Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008–2018. Agric. For. Meteorol. 297, 108240.

Wu, X.C., Xiao, X., Steiner, J., Yang, Z., Qin, Y., Wang, J., 2021b. Spatiotemporal changes of winter wheat planted and harvested areas, photosynthesis and grain production in the contiguous United States from 2008–2018. Remote Sens. 13, 1735.

Xiao, D., Tao, F., 2014. Contributions of cultivars, management and climate change to winter wheat yield in the North China Plain in the past three decades. Eur. J. Agron. 52, 112–122.

Xiao, J.F., Davis, K.J., Urban, N.M., Keller, K., Saliendra, N.Z., 2011. Upscaling carbon fluxes from towers to the regional scale: influence of parameter variability and land cover representation on regional flux estimates. J. Geophys. Res. Biogeosci. 116, 15.

Xiao, X., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q., Moore III, B., 2004a. Satellite-based modeling of gross primary production in an evergreen needleleaf forest. Remote Sens. Environ. 89 (4), 519–534.

Xiao, X.M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Berrien, M., Ojima, D., 2004b. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. Remote Sens. Environ. 91, 256–270.

Xiao, X.M., Zhang, Q.Y., Saleska, S., Hutyra, L., De Camargo, P., Wofsy, S., Frolking, S., Boles, S., Keller, M., Moore, B., 2005. Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. Remote Sens. Environ. 94, 105–122.

Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., Wang, L., 2017. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using landsat imagery and the CERES-wheat model. Agric. For. Meteorol. 246, 194–206.

Xin, F., Xiao, X., Zhao, B., Miyata, A., Baldocchi, D., Knox, S., Kang, M., Shim, K.M., Min, S., Chen, B., Li, X., 2017. Modeling gross primary production of paddy rice cropland through analyses of data from CO_2 eddy flux tower sites and MODIS images. Remote Sens. Environ. 190, 42–55.

Yuan, W.P., Zheng, Y., Piao, S.L., Ciais, P., Lombardozzi, D., Wang, Y.P., Ryu, Y., Chen, G.X., Dong, W.J., Hu, Z.M., Jain, A.K., Jiang, C.Y., Kato, E., Li, S.H., Lienert, S., Liu, S.G., Nabel, J., Qin, Z.C., Quine, T., Sitch, S., Smith, W.K., Wang, F., Wu, C.Y., Xiao, Z.Q., Yang, S., 2019. Increased atmospheric vapor pressure deficit reduces global vegetation growth. Sci. Adv. 5 (8), eaax1396. https://doi.org/10.1126/sciadv.aax1396.

Zhang, N., Zhao, C., Quiring, S.M., Li, J., 2017a. Winter wheat yield prediction using normalized difference vegetative index and agro-climatic parameters in Oklahoma. Agron. J. 109 (6), 2700–2713.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., Dong, J., 2017b. A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. Sci. Data 4, 170165.

Zhao, M., Running, S.W., 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. Science 329 (5994), 940–943.

Zhu, P., Kim, T., Jin, Z., Lin, C., Wang, X., Ciais, P., Mueller, N.D., Aghakouchak, A., Huang, J., Mulla, D., Makowski, D., 2022. The critical benefits of snowpack insulation and snowmelt for winter wheat productivity. Nat. Clim. Change 12, 485–490.

Zhuo, W., Fang, S., Gao, X., Wang, L., Wu, D., Fu, S., Wu, Q., Huang, J., 2022. Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: a case study for winter wheat in Hebei, China during 2009-2013. Int. J. Appl. Earth Obs. Geoinf. 106, 102668.

Zhuo, W., Fang, S., Wu, D., Wang, L., Li, M., Zhang, J., Gao, X., 2022. Integrating remotely sensed water stress factor with a crop growth model for winter wheat yield estimation in the North China Plain during 2008–2018. Crop J. https://doi.org/10.1016/j.cj.2022.04.004.

Zhuo, W., Huang, J., Gao, X., Ma, H., Huang, H., Su, W., Meng, J., Li, Y., Chen, H., Yin, D., 2020. Prediction of winter wheat maturity dates through assimilating remotely sensed leaf area index into crop growth model. Remote Sens. 12 (18), 2896.

Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Huang, H., Xu, B., Xiao, X., 2019. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. Remote Sens. 11 (13), 1618.