Agricultural and Forest Meteorology 216 (2016) 188–202

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation

Jianxi Huang[a,b,*], Fernando Sedano[c], Yanbo Huang[d], Hongyuan Ma[a], Xinlu Li[a], Shunlin Liang[e,c], Liyan Tian[f], Xiaodong Zhang[a], Jinlong Fan[g], Wenbin Wu[h]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, China
[c] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA
[d] United States Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, Stoneville, MS 38776, USA
[e] State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China
[f] Department of Geography, College of Geosciences, Texas A&M University, College Station, TX 77843, USA
[g] National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China
[h] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form 19 October 2015
Accepted 22 October 2015
Available online 14 November 2015

Keywords:
WOFOST
Leaf area index
Kalman filter
Ensemble Kalman filter
Winter wheat
Yield estimation

ABSTRACT

The scale mismatch between remote sensing observations and state variables simulated by crop growth models decreases the reliability of crop yield estimates. To overcome this problem, we implemented a two-step data-assimilation approach: first, we generated a time series of 30-m-resolution leaf area index (LAI) by combining Moderate Resolution Imaging Spectroradiometer (MODIS) data and three Landsat TM images with a Kalman filter algorithm (the synthetic KF LAI series); second, the time series were assimilated into the WOFOST crop growth model to generate an ensemble Kalman filter LAI time series (the EnKF-assimilated LAI series). The synthetic EnKF LAI series then drove the WOFOST model to simulate winter wheat yields at 1-km resolution for pixels with wheat fractions of at least 50%. The county-level aggregated yield estimates were compared with official statistical yields. The synthetic KF LAI time series produced a more realistic characterization of LAI phenological dynamics. Assimilation of the synthetic KF LAI series produced more accurate estimates of regional winter wheat yield (R^2 = 0.43; root-mean-square error (RMSE) = 439 kg ha^{-1}) than three other approaches: WOFOST without assimilation (determination coefficient R^2 = 0.14; RMSE = 647 kg ha^{-1}), assimilation of Landsat TM LAI (R^2 = 0.37; RMSE = 472 kg ha^{-1}), and assimilation of S-G filtered MODIS LAI (R^2 = 0.49; RMSE = 1355 kg ha^{-1}). Thus, assimilating the synthetic KF LAI series into the WOFOST model with the EnKF strategy provides a reliable and promising method for improving regional estimates of winter wheat yield.

1. Introduction

Crop growth and yield monitoring are essential to inform and develop national food policies and food security strategies. Several important factors, including population growth, cropland reductions, shortages of water, and extreme weather, can significantly influence agricultural production and threaten food security. Therefore, accurate monitoring of regional crop growth and estimation of grain yield are crucial for agricultural management and development of production-management strategies (Becker-Reshef et al., 2010; Franch et al., 2015).

Leaf area index (LAI) is an important vegetation biophysical variable that has been widely used for monitoring crop growth and estimating yields. LAI represents the ability of the crop to intercept solar radiation, which drives CO_2 assimilation and dry matter accumulation, and is therefore a key indicator for potential grain yield. There are three main methods to obtain crop LAI estimates. The first method is to directly measure LAI in the field. However, due to the large spatial heterogeneity of LAI, this approach requires intensive sampling over large areas, which is prohibitively expensive and time-consuming. A second alternative is to estimate LAI via crop growth models that use agricultural, meteorological and soil data together with crop parameters as inputs. These process-oriented crop growth models can potentially simulate temporal

* Corresponding author at: College of Information and Electrical Engineering, China Agricultural University, No. 17 Qinghua East Road, Haidian District, Beijing 100083, China.
E-mail address: jxhuang@cau.edu.cn (J. Huang).

http://dx.doi.org/10.1016/j.agrformet.2015.10.013
0168-1923/

patterns of crop LAI accurately for a given field or small area provided that the model input parameters and initial state variables have the required degree of precision and accuracy. However, the need to define the model's input parameters and initial conditions over large geographical regions restricts the application of crop growth models at a regional scale; the same cost and time problems then arise as in the case of field measurements. The third method is to retrieve crop LAI from remotely sensed reflectance values through inversion of radiative transfer models at large spatial scales, such as the MODIS LAI products (Myneni et al., 2002; Meroni et al., 2004). Unfortunately, because of mixed image pixels associated with heterogeneous land cover in most agricultural areas in northern China, the coarse 1-km resolution of MODIS LAI products leads to potentially large underestimations when compared to ground-based observations.

One solution to improve the methods above is to use data assimilation, a technique that incorporates field observations into dynamic mechanistic models, to produce more accurate estimates of the model's state variables and improve the model's outputs. Data assimilation has been used for LAI estimates and crop yield predictions with considerable success (Curnel et al., 2011; Xu et al., 2011; de Wit et al., 2012; Ma et al., 2013a,b; Wang et al., 2013; Li et al., 2014; Huang et al., 2015a, 2015b). Sequential assimilation is an important form of data assimilation that accounts for changes in image and crop characteristics over time (Liang and Qin, 2008). Sequential assimilation includes methods such as the ensemble Kalman filter (EnKF) that sequentially account for the uncertainties that arise during crop growth in both remotely sensed observations and crop simulation models (Dorigo et al., 2007; Quaife et al., 2008). The EnKF approach can be used to improve crop model performance without altering the model's structure by periodically updating state variables (e.g., LAI and soil moisture) within the growing season based on remote sensing observations (de Wit and van Diepen, 2007; Ines et al., 2013; Ma et al., 2013b; Li et al., 2014).

Several EnKF assimilation schemes with different degrees of complexity and integration have been developed and evaluated during the last decade. de Wit and van Diepen (2007) observed that EnKF-based assimilation of coarse-resolution satellite soil moisture estimates corrected errors in the WOrld FOod STudies (WOFOST) model's water balance and improved the relationship between modeled yields and official yield statistics for winter wheat. Curnel et al. (2011) found poor estimates of LAI when the differences in phenological development between assimilated and modeled LAI values were not corrected using an EnKF-based assimilation strategy. Vazifedoust et al. (2011) investigated the potential of assimilating LAI and relative evapotranspiration (as an indicator of agricultural drought) to improve estimates of total wheat production. They found that 1-month predictions using assimilated variables at a regional scale were more reliable than estimates based only on statistical yield data. Zhao et al. (2013) assimilated MODIS LAI time series data into the Python-WOFOST model using an EnKF algorithm to estimate maize yield, and achieved improved LAI monitoring and yield estimation in years with normal climatic conditions. Ines et al. (2013) assimilated remotely sensed soil moisture and LAI data into the Decision Support System for Agrotechnology Transfer-Cropping System Model (DSSAT-CSM)-Maize using an EnKF algorithm. They found that jointly assimilating soil moisture and LAI achieved better results than using only one of the two variables; furthermore, they demonstrated that the availability of downscaled remotely sensed soil moisture and LAI data made crop modeling considerably more accurate.

Crop growth models typically simulate a single relatively homogeneous site. However, the pixels of remote-sensing images in most landscapes represent the comprehensive effects of the reflectance signals from all land cover types within the pixel (Zhao et al., 2015). One of the most challenging issues in using remote sensing data-assimilation procedures has proven to be the heterogeneity of the land uses in each pixel. This is most apparent in coarse-resolution imagery (e.g., the 250-m to 1-km pixel size of the MODIS LAI). Several previous studies indicated that, because of such coarse spatial resolution, the MODIS LAI products that are commonly used for data assimilation cannot achieve satisfactory assimilation results for regions where wheat is cultivated (Duveiller et al., 2011; Xu et al., 2011; Ma et al., 2013a; Huang et al., 2015b). This is because the coarse-resolution pixels tend to result in larger scaling effects than would occur with higher-resolution sensors such as those on the Landsat TM, ASTER, and RapidEye satellites (Garrigues et al., 2006).

Remote-sensing data with medium or high spatial resolution provide an important data source for a successful crop model data-assimilation framework because such data reduce the scale mismatch effect between the observations and the model simulations. A number of previous studies demonstrated the potential of applying medium-resolution remote sensing data in agricultural data-assimilation procedures. Ma et al. (2013b) assimilated a normalized-difference vegetation index (NDVI) time series from the Chinese HJ-1 satellite into the coupled WOFOST–A two-layer Canopy Reflectance Model (WOFOST-ACRM) model using an EnKF-based assimilation strategy, and developed an improved relationship between the assimilated wheat yields and official statistical yield. Li et al. (2014) obtained improved maize yield estimates by assimilating multi-temporal LAI images from available 30-m Landsat ETM+ data into the WOFOST model using the EnKF algorithm. The availability of remote sensing data with high temporal resolution would potentially offer advantages in monitoring crop phenological development and variability during different developmental stages. However, satellite data with high temporal resolution generally tend to have a coarse spatial resolution and consistent time series of cloud-free images at medium spatial resolution are seldom available due to the frequent cloud cover during the crop growing season. One promising generalized solution to overcome the scale mismatch would be to combine accurate LAI values derived from medium-resolution images (e.g., Landsat TM) with phenological characteristics derived from sensors with a lower spatial resolution but a higher revisit frequency (e.g., MODIS), thereby generating an LAI trajectory with higher spatial and temporal resolution throughout the crop growing season.

Several methods have been proposed to integrate information from sensors with different spatial and temporal resolutions to simulate medium-resolution images for dates on which the images are not available (e.g., due to cloud cover). These methods rely on models with various degrees of complexity that are based on the empirical relationships between simultaneous images with different resolutions. Gao et al. (2006) developed an empirical fusion technique that blended 500-m MODIS and Landsat TM surface reflectance values using an initial Landsat image and a pair of MODIS images. Roy et al. (2008) developed a semi-empirical fusion approach using MODIS/Bidirectional Reflectance Distribution Function (BRDF) albedo data and Landsat TM surface reflectance data to generate synthetic Landsat images that accounted for the directional dependence of surface reflectance. Zurita-Milla et al. (2009) and Amorós-López et al. (2013) applied spectral unmixing to generate a time series of fine-resolution vegetation indices by combining MERIS and Landsat TM imagery with ancillary land-use information. Zhu et al. (2010) modified Gao's approach by implementing a spectral unmixing approach to improve the retrieval of data from pixels with heterogeneous cover types.

In general, these fusion methods successfully retrieved single synthetic fine-resolution scenes, provided that a pair of moderate- and fine-resolution images could be obtained that were not too far apart in time. The quality of the simulation is more likely to degrade when the simulated scene is further in time from the

reference pair, since the model's initial assumptions become weaker. Data-assimilation techniques optimally combine information from observation systems and models and their respective uncertainties to minimize the residual errors (Mathieu and O'Neill, 2008), and are an attractive alternative to produce synthetic medium-resolution images from existing medium- and moderate-resolution imagery. A clear advantage of explicitly accounting for these uncertainties in the retrieval of the simulated medium-resolution image is that this allows operational implementations in which a full time series of high-resolution synthetic images can be produced, rather than single synthetic images. Sedano et al. (2014) implemented a recursive Kalman filter algorithm (KF) to produce a time series of Landsat TM NDVI values at 16-day intervals using the available Landsat images and a time series of 250-m MODIS NDVI images.

In the present study, we developed a crop modeling framework that incorporates data assimilation at two stages: first, we used a KF algorithm to integrate MODIS and Landsat TM data and generate a complete time series of synthetic LAI values at a medium spatial resolution (30 m) at 4-day intervals. Second, we compared assimilation of the synthetic KF LAI time series with assimilation of the MODIS and Landsat products used separately. The main objective of this study was to evaluate whether the synthetic KF LAI time series could improve the accuracy of winter wheat yield estimates at a regional scale.

2. Study area

Our study area comprised 53 counties mainly in four prefecture-level cities (Baoding, Cangzhou, Hengshui, and Xingtai) in the central part of China's Hebei Province (Fig. 1). This area extends from 36°44′ N to 39°38′ N and from 115°28′ E to 117°16′ E, and it is dominated by winter wheat cultivation. The region is dominated by alluvial plains and has a typical temperate monsoon climate, with an average annual rainfall of 550 mm and an average temperature of 12 °C. Monthly mean temperatures range from a minimum of −3 °C in January to a maximum of 27 °C in July. A test site that provides complete experimental observation conditions for this region was established at the Ecological-Meteorological Experimental Station of the China Meteorological Administration in Gucheng (115.733° E, 39.148° N), in the northern part of Dingxing County.

The prevailing cropping system is a winter wheat–summer maize rotation, which represents the traditional planting pattern in the North China Plain and accounts for about 90% of the cereal cultivation area. Winter wheat is generally planted at the beginning of October and harvested in June of the next year. The important phenological stages for winter wheat include the green-up stage (early March), jointing stage (late March), elongation stage (early April), booting stage (mid-April), heading stage (late April to early May), anthesis stage (mid-May), and maturity (early June). The areas with high wheat planting densities are located in the western and northern parts of the study area, and the southern counties have lower planting densities.

3. Model and data

3.1. WOFOST model

WOFOST is a mechanistic model that simulates plant growth and physiological development processes at a daily time step for most crops. The model has been parameterized and calibrated for winter wheat in our study area. Ma et al. (2013b) and Huang et al. (2015b) provide details of the model and of its parameterization and calibration. LAI is an important output variable of the WOFOST model because it represents the crop's ability to capture light and assimilate carbon, which are crucial indicators of the potential grain yield. Therefore, in this study, LAI was adopted as the state variable in the data-assimilation procedure. The WOFOST model provides estimates of biomass and grain yields at a daily time step for a particular crop type. The model's outputs are directly usable for crop-specific yield estimation. The WOFOST model can be run in potential mode, with no limitations caused by water stress and other factors, and we chose potential mode in the present study because meteorological conditions during our study period suggested no unusual stress on the winter wheat, and there were no reports of significant outbreaks of insects or diseases.

3.2. Field data

We selected 53 sample plots representing different winter wheat growing conditions throughout the study area and monitored them from March to June 2009. Plots were relatively homogeneous and were established to match the corresponding Landsat TM 30-m-resolution pixels. Winter wheat LAI was measured using the LAI-2000 Plant Canopy Analyzer (LI-COR, Lincoln, NE, USA) during the seven key phenological stages: green-up, jointing, elongation, booting, heading, anthesis, and maturity. We also manually measured winter wheat yields in the plots after harvesting in mid-June. Official government statistics on winter wheat yields were obtained at a county level from the 2009 Hebei statistical yearbook.

3.3. Remote-sensing data

We acquired six cloud-free Landsat TM5 images of the study area on 14 March, 17 May, and 2 June 2009, close to the field-measured dates of green-up (5 March), anthesis (14 May), and maturity (10 June) from the EarthExplorer site (http://earthexplorer.usgs.gov/). The TM images were georeferenced to the Albers conical equal-area map projection using 45 field-measured ground control points. After geometric correction, the root-mean-square error (RMSE) of the calculated and measured locations was less than one pixel (30 m) for each TM image. An atmospheric correction was applied using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) model in the ENVI software (version 5.0) to obtain the reflectance in each band (RSI, 2001). The soil-adjusted vegetation index (SAVI) and NDVI were used to estimate LAI on each date. We also used the 4-day composite MODIS LAI product (MCD15A3), with 1-km spatial resolution, for a total of 45 dates from January to June 2009 (http://reverb.echo.nasa.gov/reverb). We applied an iterative Savitzky–Golay (S-G) filtering algorithm to the MODIS time series to obtain a smoothed LAI profile. In addition, we used a crop type map and a map of winter wheat pixel purity (the % of each pixel covered by winter wheat) to exclude pixels with less than 50% pixel purity from our analysis (Huang et al., 2015b).

4. Data assimilation

Our data assimilation process involved two main phases: first, we used a recursive Kalman filter algorithm to generate a complete LAI time series at a 30-m spatial resolution and a 4-day time interval from the time series of MODIS data and the three Landsat TM images; hereafter, the "synthetic KF LAI time series". Second, we assimilated the synthetic KF LAI series into the WOFOST model to create an ensemble Kalman Filter (EnKF) time series of daily values; hereafter, the "EnKF-assimilated LAI time series". Using the EnKF-assimilated LAI time series as input, the WOFOST model calculated the assimilated yields.

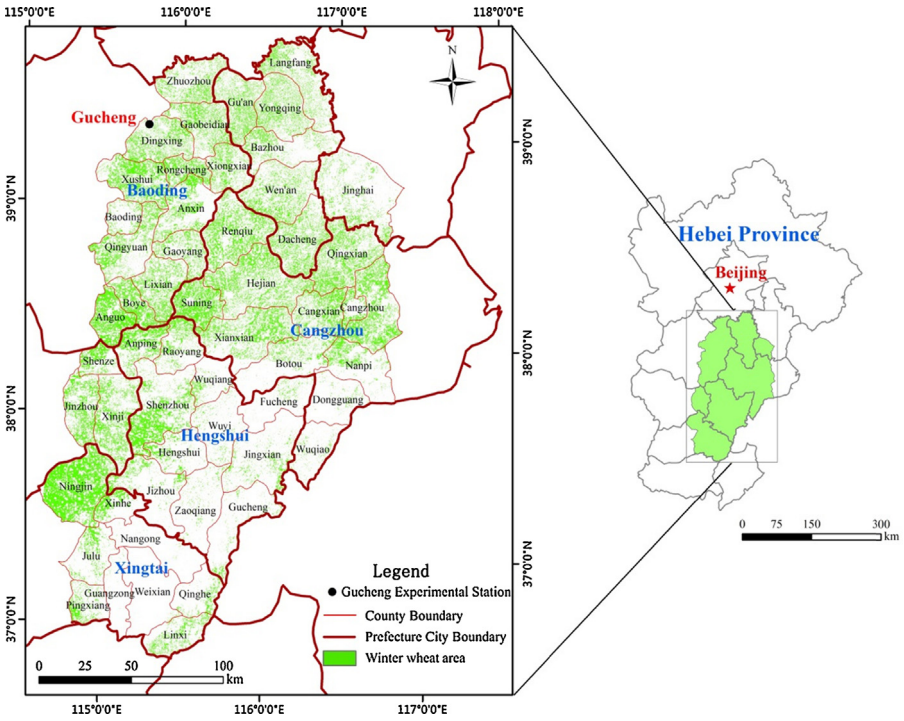

Fig. 1. The study area, and locations of the 53 counties in central Hebei Province.

4.1. Continuous time series of medium-resolution synthetic LAI images with a Kalman filter

We used a Kalman filter recursive algorithm (Kalman, 1960) to produce a 30-m-resolution LAI time series at a 4-day time step using the Landsat TM LAI data and the S-G filtered MODIS LAI time series described in the previous section as inputs. The KF algorithm integrates observations, models, and their respective uncertainties to estimate the state of a process with minimization of the RMSE (Maybeck, 1979; Welch and Bishop, 1995). KF algorithms have been successfully applied in soil hydrology (Huang et al., 2008), ecosystem modeling (Quaife et al., 2008; Samain et al., 2008), and crop phenology estimation (Vicente-Guijalba et al., 2014).

The KF algorithm retrieves the states of a process based on a combination of present measurements, a linear state-transition model, and the respective uncertainties of these elements:

$$x_k = Ax_{k-1} + w_{k-1} + w_{k-1} \tag{1}$$

$$z_k = Hx_k + v_k \tag{2}$$

where x_{k-1} and x_k are the model estimates in the previous $(k-1)$ and present (k) states, respectively; A represents a linear state-transition model that links x_k and x_{k-1}; z_k is the observation value in state k; and w and v are Gaussian random variables that represent white noise in the process being modeled, $N(0, Q)$, and white noise in the measurements, $N(0, R)$, respectively, with a mean of zero and with Q and R as the standard deviations; and H is an observation operator that relates the state to observation z_k. The KF algorithm is implemented in two steps. In the first step (the time update), the linear state-transition model propagates the estimate from the previous state $(k-1)$ and its uncertainty to provide prior estimates of the present state (k) of the model (Eq. (3)) and its uncertainty (Eq. (4)):

$$\hat{x}_k^- = A\hat{x}_{k-1} \tag{3}$$

$$P_k^- = AP_{k-1} + w_{k-1} \tag{4}$$

where $\hat{x}_k^-$ is the prior estimate of the present state (k); $\hat{x}_{k-1}$ is the posterior estimate of the variable in the previous state $(k-1)$; P_K^- is the prior uncertainty at the present state; and P_{k-1} is the posterior uncertainty at the previous state.

In the second step (the measurement update), the prior estimates (Eq. (5)) and their uncertainties (Eq. (6)) are updated with new observations via a linear combination of the prior model's estimate and the weighted difference between observation and prior estimate of model's state. The weights are defined by the "Kalman gain" at state k (K_k) based on the uncertainties of previous state and present observation (Eq. (7)):

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{5}$$

$$P_k = (1 - K_kH)P_k^- \tag{6}$$

$$K_k = P_k^-H^{\mathrm{T}}(HP_k^-H^{\mathrm{T}} + R)^{-1} \tag{7}$$

where T represents the matrix transpose operation, and $\hat{x}_k$ is the posterior estimate of the model's state. Large measurement noise (R) results in a low Kalman gain, which gives more weight to the model process. Conversely, a large process covariance (P) results in a high Kalman gain and more weight for the new measurements.

The implementation of the KF algorithm requires the definition of a set of observations and an underlying linear transition model. In this study, the available Landsat TM LAI images were used in the study period as the observations. The transition model was defined as the ensemble of two submodels, as was proposed by Sedano et al. (2014). The first submodel captures the phenological LAI trajectories over time, and is based on a linear regression between pixel values from successive MODIS LAI 4-day composites. This regression was calculated only for pixels with at least 50% pixel purity for winter wheat. The second submodel captures the relationship between MODIS and Landsat LAI pixel values and provides an additional constraint to the seasonal trajectories defined by the first submodel. This submodel implements a linear regression between the concurrent MODIS and Landsat TM LAI images to implicitly account for potential image noise and non-linear relationship between MODIS and Landsat LAI pixel values that could

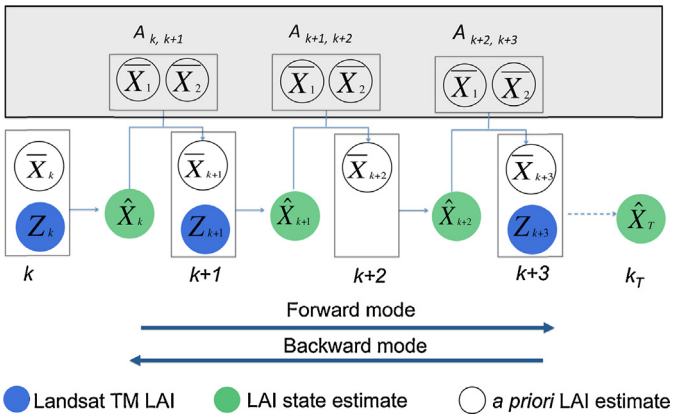

Fig. 2. Flow chart of Kalman filter recursive approach used to produce a continuous time series of synthetic LAI images at a 30-m spatial resolution. At each time-step, the transition model, *A*, combines estimates from submodels 1 and 2 ($\bar{x}_1$ and $\bar{x}_2$) to produce a single *a priori* estimate, X_{k-} from previous state estimate (time update). The final estimate of the state, X_k, is the weighted average of the time update and the measurement update. If a new Landsat observation is not available, the estimate of the state at that time-step is the result of the measurement update from the previous time step. The model is alternatively run in forward and backward modes and their corresponding estimates were subsequently combined (smoothing mode).

result in weaker relationships between the pairs of images and increase uncertainties in the LAI estimate. The uncertainties in each submodel were calculated as the standard errors of the linear regressions.

The initial state of the model was defined by using the first MODIS LAI 4-day composite. This approach followed the work of Sedano et al. (2014) by assigning the standard deviation of the values in the first image as the initial uncertainty. Then, the KF algorithm was applied in smoothing mode, in which forward and backward recursions are combined to estimate the LAI state values (Rauch, 1963). The smoothing mode results in lower residuals and uncertainties than forward or backward filtering modes.

Fig. 2 illustrates the KF approach used to produce a continuous time series of LAI images at a 30-m spatial resolution. At each time-step, the transition model (*A*) projects the estimate from the previous state (the time update step) by combining the estimates of submodels 1 and 2 ($\bar{X}_1$ and $\bar{X}_2$ respectively) to produce a single *a priori* estimate, $\bar{X}_k$. If available, a Landsat LAI observation, Z_k, provides a new estimate for the state (the measurement update step). The final estimate of the state, X_k, is the weighted average of the time update (via the transition model) and the measurement update (based on the Landsat TM LAI observation), with weights for the two updates inversely proportional to their respective uncertainties. If a new Landsat TM LAI observation is not available, the estimate of the state at that time-step is the result of just the time update, and is calculated using the result of the measurement update from the previous time step and the transition model. The model is alternately run in forward and backward modes (the filtering process) and their corresponding estimates are subsequently combined (the smoothing process). The forward and backward models were combined under the assumption that their estimates were both independent and Gaussian by applying the following equations:

$$\hat{x}_{FBk} = \hat{x}_{Fk}[P_{Fk}/(P_{Fk} + P_{Bk})] + \hat{x}_{Bk}[P_{Bk}/(P_{Fk} + P_{Bk})] \quad (8)$$

$$\frac{1}{P_{FBk}} = \left(\frac{1}{P_{Fk}}\right) + \left(\frac{1}{P_{Bk}}\right) - (1/R_k) \quad (9)$$

where $\hat{x}_{Fk}$, $\hat{x}_{Bk}$ and $\hat{x}_{FBk}$ denote the posterior estimates at state *k* in the forward, backward and combined mode, respectively; P_{Fk}, P_{Bk} and P_{FBk}, are the uncertainties at state *k* in forward, backward and combined mode, respectively; R_k is the measurement uncertainty at state *k*.

4.2. Assimilation of remotely sensed LAI datasets into the WOFOST model using the EnKF algorithm

The EnKF algorithm proposed by Evensen (2003) is based on Monte Carlo ensemble generation, in which the approximation of an *a priori* state error covariance matrix is forecasted by propagating an ensemble of model states using updated states (ensemble members) from the previous time step. As in the case of KF, the EnKF-based assimilation also requires the definition of a transition model and a set of observations. In this study, the WOFOST model was used as a nonlinear dynamic transition model and the satellite remotely sensed LAI (i.e., the synthetic KF LAI series, S-G filtered MODIS LAI, and three TM LAI) were used as the observations. The WOFOST model can be used to simulate growth variations over time based on the model's input parameters and initial conditions during various phenological stages. In contrast, the remotely sensed LAI values represent the actual crop growth status over a large spatial scale. We used the EnKF algorithm to assimilate the remotely sensed LAI datasets (i.e., the synthetic KF LAI series, S-G filtered MODIS LAI or three TM LAI) into the WOFOST model in an effort to optimize the estimates of wheat LAI by reducing the uncertainties that exist in both the remotely sensed LAI and the WOFOST model. The ultimate goal was to assess the potential of regional estimation of winter wheat yield using the synthetic KF LAI data.

4.2.1. Uncertainties in the crop model parameters and the remotely sensed observations

Determining and accounting for the uncertainties in the crop model parameters and the remotely sensed observations is an essential step to improve the performance of the EnKF-based assimilation system. Crop growth development phases and grain yield are bounded by the crop model's input parameters and initial conditions. The initial crop total dry weight (the TDWI parameter) influences the initial growth rate and affects the maximum LAI that can be reached during the growing season. The lifespan of leaves growing at 35 °C (the SPAN parameter) determines the rate and the time when LAI begins to decrease (de Wit et al., 2012; Huang et al., 2015b). There are two common methods to generate the ensemble members of modeled LAI (i.e., the forecast ensemble members). The first method directly adds a Gaussian perturbation to the WOFOST-simulated LAI (Ma et al., 2013b). The second method adds a Gaussian perturbation to WOFOST's uncertain input parameters, and then uses these disturbed values as inputs for WOFOST to simulate the LAI ensemble members. This second method assumes that the WOFOST model's uncertainty comes from errors in the input parameters (Curnel et al., 2011; de Wit and van Diepen, 2007). The method based on perturbation of input parameters was used in the present study to generate the forecast ensemble members.

Based on a sensitivity analysis and successful use of TDWI and SPAN in previous data-assimilation practices for estimating regional winter wheat yields (de Wit et al., 2012; Ma et al., 2013a; Huang et al., 2015b), these two parameters were chosen to account for the uncertainty of the WOFOST model in this study. The WOFOST simulation started at the true emergence date, which was set to 18 October 2008. The two model parameters (TDWI and SPAN) were perturbed once by adding white noise with a zero mean and an appropriate standard deviation, and were used in the WOFOST model to generate the time series of simulated LAI with 100 forecast ensemble members. Values of the standard deviations of TWDI and SPAN were set to 7.8 kg ha^{-1} and 0.7 days respectively, according to the results of our previous study (Huang et al., 2015b).

The field-measured LAI data were upscaled by averaging the LAI values for all 90 m × 90 m subplots that fell within a 1-km grid cell. Then, linear regression equations between the resulting upscaled 1-km LAI and the corresponding S-G filtered MODIS LAI at the seven phenological stages were built. RMSE values of these regression

equations were used as the observational errors for the 1-km S-G filtered MODIS LAI in the EnKF algorithm. Linear regression equations between LAI values from the field-measured subplots and the corresponding Landsat TM LAI values during the three phenological stages for which Landsat data were available (green-up, anthesis, and maturity) were also built. Details of these regression models can be found in Huang et al. (2015b). The RMSE values were also used for these regression equations as the observational errors for TM LAI in the EnKF algorithm. For the synthetic KF LAI series, we believe that the distribution of all 30-m synthetic KF LAI values within a 1-km grid cell represents the observational errors well for each 1-km grid cell. Thus, random sampling of 100 synthetic KF LAI values within a 1-km grid cell was used as the observational ensemble members in the EnKF algorithm (Evensen, 2003).

4.2.2. The ensemble Kalman filter assimilation procedure

In the ensemble Kalman filter system, we assumed that the observations can be linked to the state variable (e.g., LAI at time k) by the following equation:

$$B = HS + v \tag{10}$$

where B is the observation vector, S is the state variable vector, and v is a Gaussian-distributed random error vector with a zero mean and certain observation error covariance, and H is the operator that maps the model state variable to the observation space. Based on this assumption, the relationship between the estimated state and error covariance can be described in the following equations:

$$S_k^a = S_k^f + K(B_k - HS_k^f) \tag{11}$$

$$K = P_k^f H^{\mathrm{T}} (HP_k^f H^{\mathrm{T}} + R_k)^{-1} \tag{12}$$

$$P_k^a = (I - KH)P_k^f \tag{13}$$

where S_k^a and S_k^f are the analysis and forecast vectors, respectively; H^{T} is the transpose matrix of H; P_k^a and P_k^f denote the error covariance of the analysis and forecast ensembles, respectively; R_k is the error covariance of the observation ensemble, I is the identity matrix, and K is the Kalman gain.

In the present study, the simulated value of LAI was the only state variable that was directly assimilated from external observations (i.e., the S-G filtered MODIS LAI, Landsat TM LAI from three dates, and the synthetic KF LAI). Therefore, H can be taken as an identity matrix, and Eqs. (11) and (12) can be rewritten as Eqs. (14) and (15), respectively:

$$S_k^a = S_k^f + K(B_k - S_k^f) \tag{14}$$

$$K = P_k^f (P_k^f + R_k)^{-1} \tag{15}$$

The EnKF forecast and analysis error covariances were calculated directly from an ensemble of model simulations:

$$P_k^f = \frac{1}{N-1} \sum_{n=1}^{N} (S_n^f - \overline{S}^f)(S_n^f - \overline{S}^f)^T \tag{16}$$

where N is the number of ensemble members, n represents a running index for ensemble member, and $\overline{S}^f$ represents the mean of the n ensemble members.

The standard EnKF method tends to reject observations in favor of the ensemble forecast in the late period of data assimilation, which could lead the analysis to deviate incrementally from the reality, which is referred to as "filter divergence" (Burgers et al., 1998; Ines et al., 2013). To reduce the effect of filter divergence, we adopted the similar idea of an inflation factor to enlarge K, as described by Lin et al. (2008). We designed an inflation factor $E \geq 1$ to change the (P_k^f) in the second half of the EnKF assimilation procedure, and assumed that E changes in response to changes in the values of P_k^f and R_k:

$$E = r\left(\frac{k}{150}\right)\left(\frac{R_k}{P_k^f}\right) \tag{17}$$

where r $(0 < r < 1)$ is a random value, 150 represents the total number of days in the data assimilation cycle from emergence to maturity, and k represents the number of days (from 1 to 150). The term $k/150$ is meant to gradually increase the inflation factor, so that after 150 days of running time, the inflation factor reaches its maximum value. In our assimilation procedure, we needed to judge whether $(R_k)/(P_k^f) > 4$ and $E \geq 1$, and if so, we calculated the inflation factor and used it to enlarge K.

The basic filtering steps applied in the EnKF-based assimilation of the remotely sensed LAI into the WOFOST model can be described as follows. The WOFOST simulation starts with the true emergence date, which was set to 18 October based on observations at agricultural meteorological stations throughout the study area. The two model parameters (TDWI and SPAN) were perturbed once by adding white noise with a zero mean and the previously determined standard deviations (see Section 4.2.1), and were then used in the WOFOST model to simulate the time series of LAI with 100 forecast ensemble members $\{\mathrm{fLAI}^1, \mathrm{fLAI}^2, \ldots \mathrm{fLAI}^{100}\}$. If an observed LAI was available at time k, we conducted random sampling from all 30-m synthetic KF LAI values within the 1-km grid cell to generate a 100-member observation ensemble $\{m\mathrm{LAI}^1, m\mathrm{LAI}^2, \ldots m\mathrm{LAI}^{100}\}$, then the forecast ensemble and the observation ensemble using the EnKF to obtain the optimal estimated ensemble $\{\mathrm{LAI}^1, \mathrm{LAI}^2, \ldots \mathrm{LAI}^{100}\}$ at time k. If no LAI observation was available at time k, then the forward simulation was conducted using the WOFOST model. This process is repeated until the crop simulated by WOFOST reaches the maturity stage. The mean of the optimal estimated ensemble provides the best estimate of LAI during the assimilation process, which is finally input into the WOFOST model to estimate the winter wheat yield for each 1-km cell in the grid. Fig. 3 provides a flowchart for the estimation of regional winter wheat yield using the EnKF-based assimilation of remotely sensed LAI into the WOFOST model. The EnKF-based assimilation algorithm was coded and coupled with the WOFOST model using the FORTRAN computer programming language.

5. Results

5.1. Synthetic KF LAI

The Kalman filter algorithm was used to generate a time series of 30-m-resolution LAI images at a 4-day time step. To account for differences in winter wheat coverage, we stratified all pixels into five classes based on the wheat vegetation cover: 0–20%, 21–40%, 41–60%, 61–80%, and >80%. Then, the Kalman filter was applied separately to each winter wheat cover class.

Fig. 4 shows the synthetic KF LAI time series for four different wheat cover classes. The growing season pattern was similar for all cover classes. LAI remained < 0.5 for the first 70 days of the year, followed by a rapid rise to reach the maximum LAI after DOY 120, followed by a rapid post-harvest decline. For pixels with winter wheat cover between 20% and 40% (Fig. 4a), the influence of other vegetation with different phenological cycles resulted in sustained high LAI values over a longer period (by 20 days). The peak LAI values within the growing season were higher for pixels with a higher winter wheat cover. Peak seasonal LAI values reached 4 in pixels with winter wheat cover between 20% and 60% (Fig. 4a and b), 6 when winter wheat cover was between 60% and 80% (Fig. 4c), and

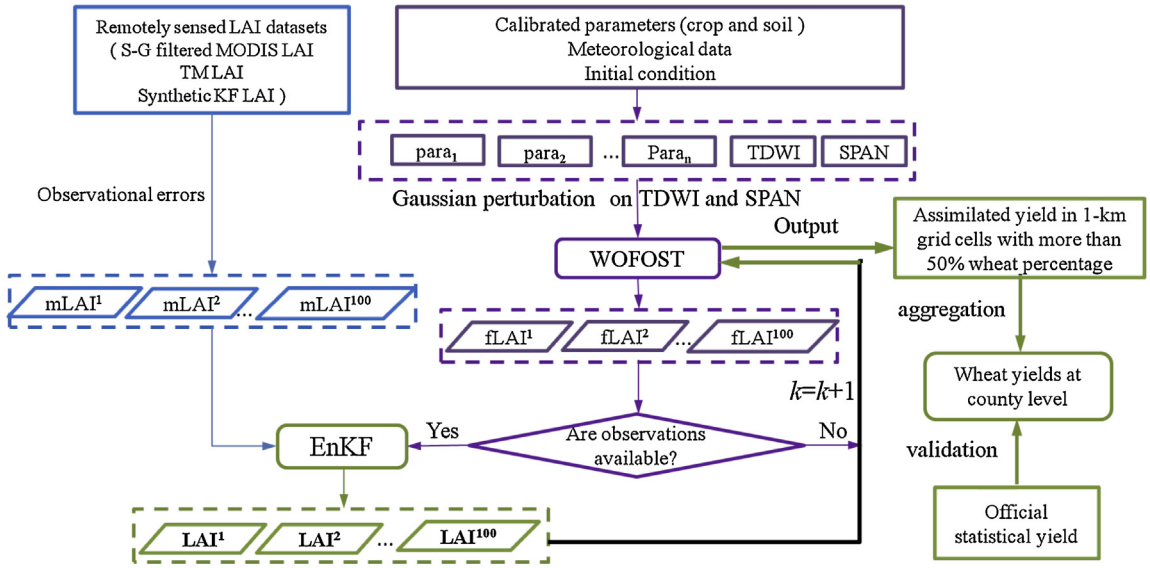

Fig. 3. Flowchart for the winter wheat yield estimation using the EnKF-based assimilation algorithm.

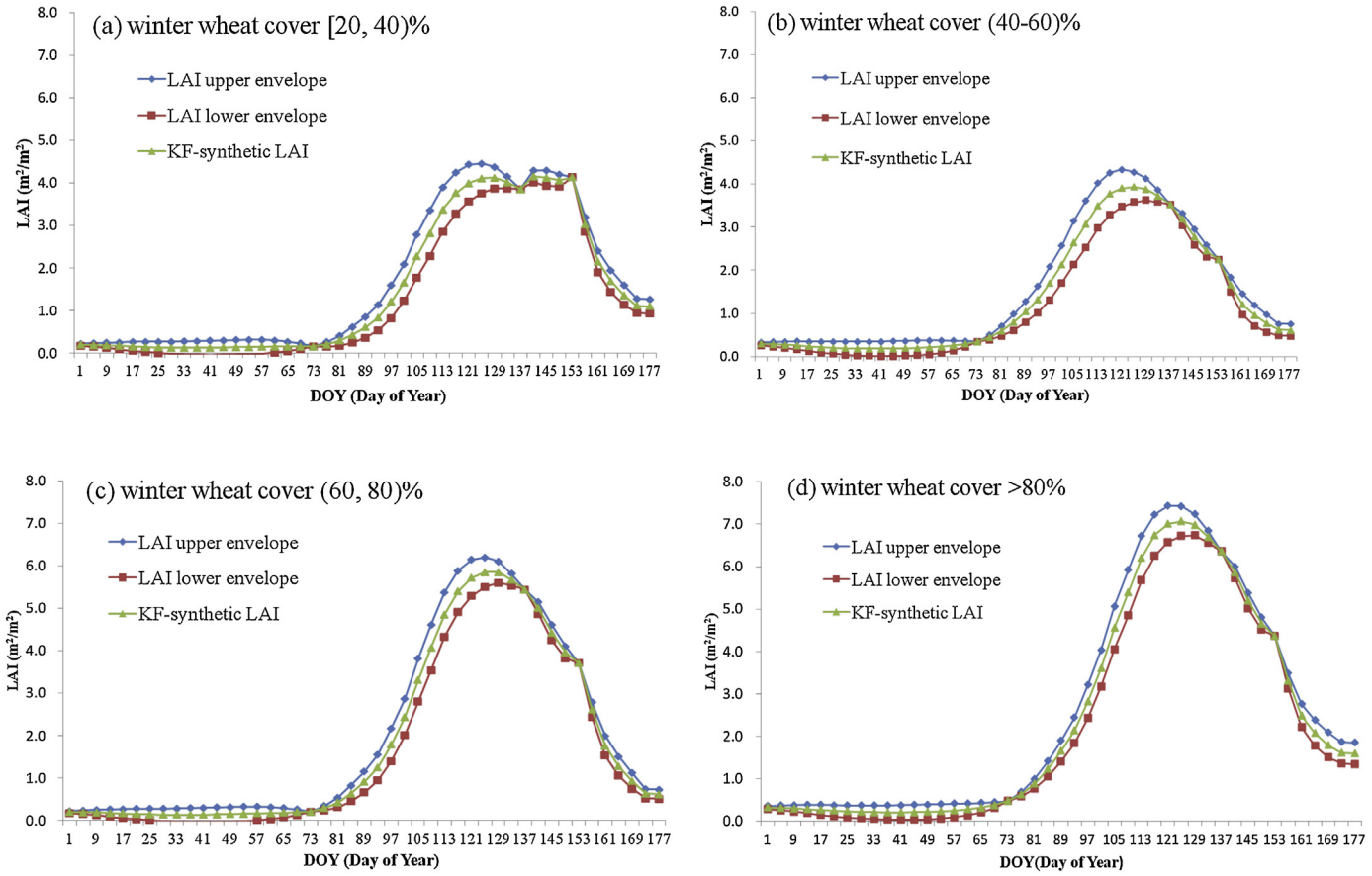

Fig. 4. Synthetic KF LAI and uncertainties in the different winter wheat cover classes.

7 when cover was higher than 80% (Fig. 4d). The uncertainties of the LAI estimates (the range between the upper and lower envelopes) varied throughout the growing season. The uncertainties of the synthetic KF LAI time series are the result of combining, through the Kalman gain (Eq. (7)), uncertainty in the previous state and observation uncertainties (Eqs. (4) and (6)). The presence of TM LAI observations at a given state implies that more precise information is available, and this reduces the uncertainty of that state. Hence, the uncertainties of the LAI estimates (the range between the upper and lower envelopes) varied throughout the growing season, depending on the distribution of the TM LAI observations. Uncertainties were lowest when TM LAI observations were available (14 March: DOY 73, 17 May: DOY 137, and 2 June: DOY 153) and were higher for time steps further from these LAI observations.

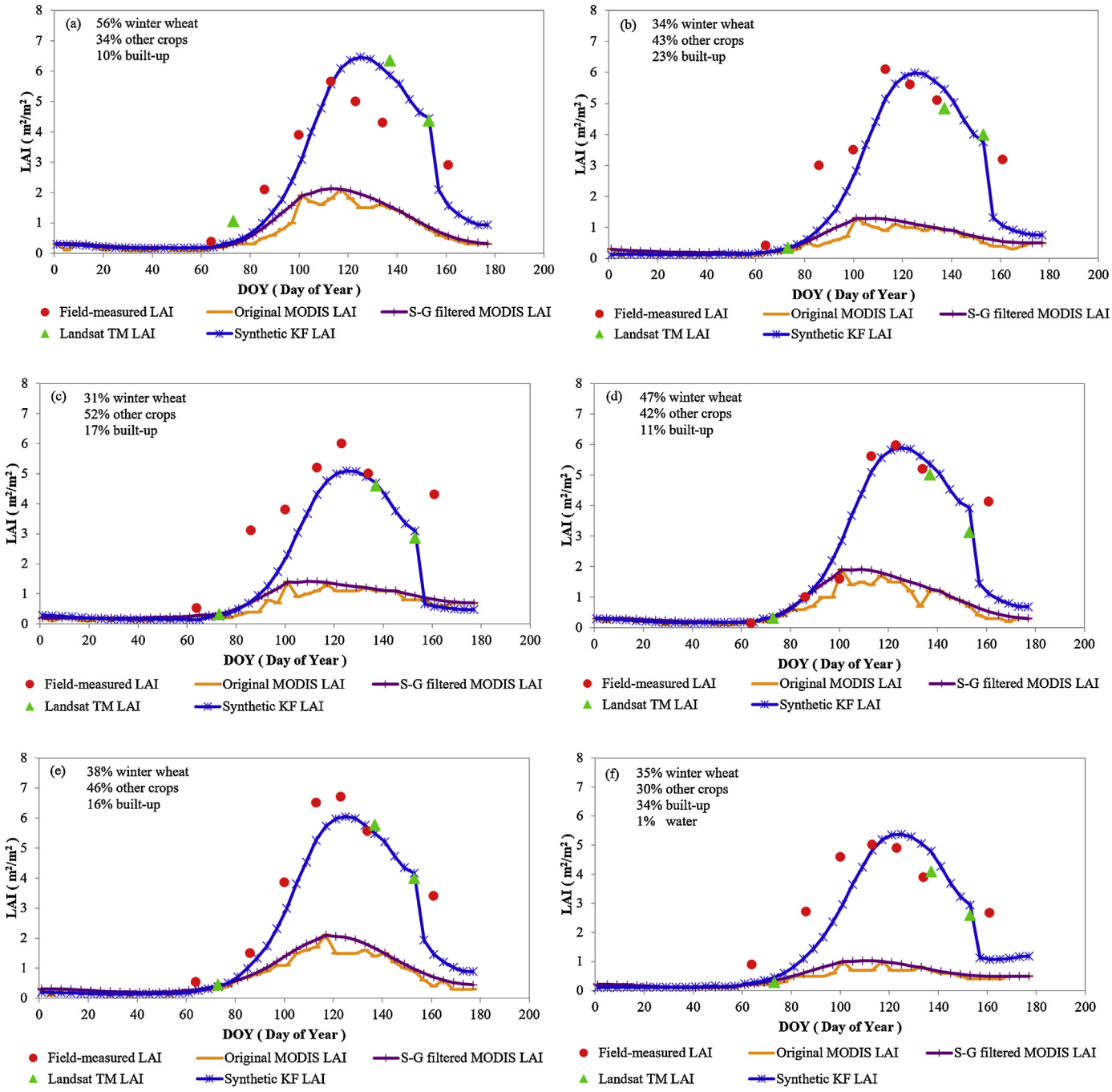

Fig. 5. Phenological changes in the LAI of winter wheat for the field-measured LAI, synthetic KF LAI, TM LAI, original MODIS LAI, and S-G filtered MODIS LAI profiles. Values represent the means for field sample plots with the indicated cover of winter wheat and other crops.

5.2. Assessment of the accuracy of the synthetic KF LAI using field-measured LAI

A direct comparison of field-measured LAI and remote-sensing data is not reasonable because of the scale mismatch between ground-level *in situ* measurements and the coarse-resolution measurements provided by wide-angle remote sensors (Liang et al., 2002; Hufkens et al., 2008). Thus, to provide a more reasonable comparison, the field measurements were upscaled to the 1-km spatial resolution based on the average of the subplot LAI values within a 1-km cell, and the synthetic KF LAI temporal profiles were resampled for 1-km pixels using the average of the original 30-m Landsat TM pixels.

One of the important enhancements provided by the KF algorithm is the improvement in the spatial and temporal characterization of the phenological dynamics for winter wheat compared with the S-G filtered MODIS LAI. The synthetic KF LAI maintained the general shape of the field-measured data (Fig. 5), with a peak around DOY 120 to 130, whereas the MODIS LAI time series (both the raw data and the S-G filtered data) produced seasonal peaks with an LAI value lower than the field-measured LAI.

The synthetic KF LAI estimates followed the field-measured data more closely than the other remote sensing-based estimates. The seasonal peaks retrieved by the KF algorithm reached values similar to those measured in the field during the booting stage (DOY 110) and the heading stage (DOY 125). However, there were still

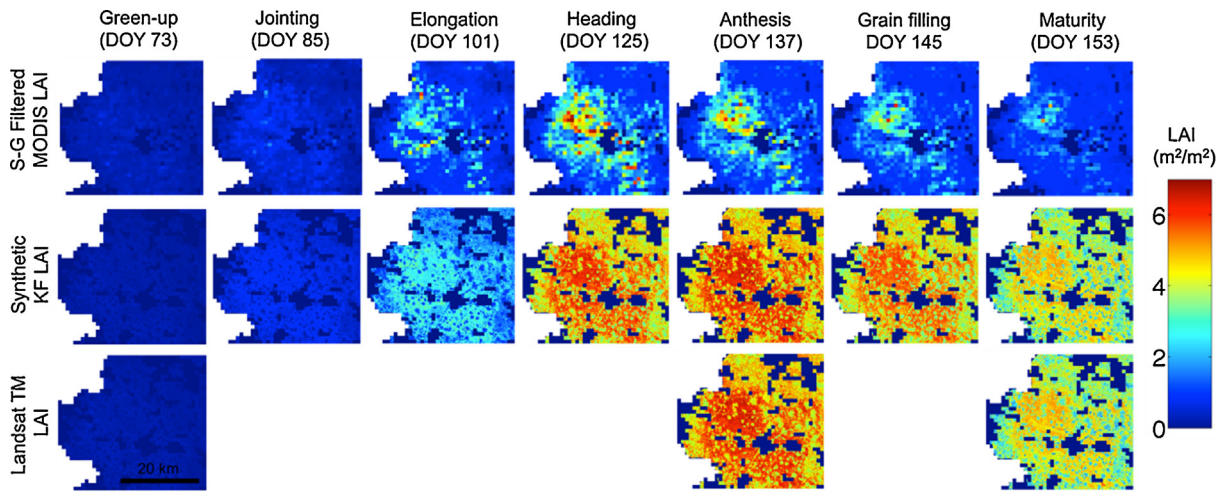

Fig. 6. Temporal sequences of the spatial variation in the S-G filtered MODIS LAI 4-day composite LAI values, the synthetic KF LAI, and LAI values from the existing Landsat TM LAI images during the main winter wheat development stages.

noticeable differences between the field-measured LAI and the synthetic KF LAI estimates at some sites. For instance, field LAI measurements at some sample sites were higher than the synthetic KF LAI estimates at the booting and heading stages (Fig. 5c and e). Also, at some locations, field measurements showed an earlier start to the growing season than in the TM LAI and synthetic KF LAI estimates (Fig. 5a, b and f). These differences can be explained by the characteristics of the input data used to generate the synthetic KF LAI time series. These time series strongly depended on the accuracy of the input TM LAI and on the phenological evolution over time derived from the S-G filtered MODIS LAI data. Our results showed that the synthetic KF LAI tended to have larger deviations from field values when the Landsat TM LAI values were not accurate (Fig. 5a and c).

Furthermore, the differences between field LAI measurements and the synthetic KF LAI estimates show that the approach is sensitive to the number of TM images and their acquisition dates. Rapid changes in LAI detected in the field around the booting and heading stages cannot be captured by the KF synthetic time series, since there were no TM LAI images close to the moment when these rapid changes occurred. The presence of TM LAI close in time to the periods when such key fluctuations in LAI values occur would enhance the predictive power of the synthetic time series. The prediction power of the KF LAI time series is also constrained by the 4-day temporal frequency of the input MODIS LAI, which limits the ability to capture LAI fluctuations that occur within shorter time periods.

An example of this is found in the fact that most of the synthetic KF LAIs suggested later heading (DOY 125) than field measurements (DOY 121). The nearest TM LAI observations, at the green-up stage (DOY 73) and the anthesis stage (DOY 137), cannot provide precise information about the non-linear changes in LAI that occur around DOY 121. The existing TM LAI images force the synthetic KF LAI series to delay the heading of the winter wheat by 4 days. Since this delay is within the temporal resolution of the 4-day composites, the transition model based on the S-G filtered MODIS LAI cannot detect such rapid changes.

5.3. Spatial distribution of the synthetic KF LAI

Because most wheat fields in the study area are relatively fragmented and interspersed with other crops and other land uses, the 1-km spatial resolution of the 4-day MODIS LAI composites generally cannot capture the spatial distribution of the winter wheat fields (Fig. 6). In particular, parts of the study area with non-vegetated cover types (e.g., roads, buildings) had lower LAI values than fully vegetated agricultural surfaces.

The 30-m spatial resolution of the TM LAI images allows an improved representation of the spatial structure of the crop fields and other land uses. However, these images were available at only three stages (green-up, anthesis, and maturity), and therefore cannot accurately track the spatial and temporal variations of winter wheat LAI throughout the growing season.

Our results nonetheless show that the synthetic KF LAI showed distinct LAI trends in both cultivated and non-cultivated land, retained the spatial structure of the cultivated fields during the growing season, and corrected the underestimation in the S-G filtered MODIS LAI (Fig. 6). Although the general LAI trends in both the S-G filtered MODIS LAI and the synthetic KF-LAI had the same shape, there were key differences in the magnitude of the LAI values, with the main differences occurring at the heading stage. The S-G filtered MODIS LAI time series values at the green-up stage averaged 0.25 m^2/m^2 (RMSE = 0.54), and increased to 1.22 m^2/m^2 (RMSE = 3.24) at the elongation stage. At the heading stage, LAI averaged 1.27 (RMSE = 4.51), and the sites dominated by wheat had LAI values of around 3 m^2/m^2. The synthetic KF LAI increased rapidly from green-up (an average of 0.22 m^2/m^2 and RMSE = 0.062) to the elongation stage (an average of 2.79 m^2/m^2 and RMSE = 1.62), reached its maximum value at the heading stage (an average of 5.52 m^2/m^2 and RMSE = 3.27), and then gradually decreased at the maturity stage (an average of 1.09 m^2/m^2 and RMSE = 0.61). This agrees with what is expected based on the LAI pattern for this crop.

The RMSE for KF-LAI at DOY 113, 121, and 133 was fairly high considering that the field LAI measurements and the synthetic KF LAI estimates were not so different. This could be related to the procedure used to up-scale the data to the 1-km grid resolution because of the border effect that would cause such errors. Another possible reason could be the LAI in these stages varied more widely than it did during other growing stages (Fig. 7).

5.4. Observational ensemble members of the synthetic KF LAI in the EnKF algorithm

Accurately determining the observational ensemble members is essential for the success of a data-assimilation scheme based on a crop model. In fact, the standard deviation of the Gaussian white noise error needs to be a realistic value for it to represent the uncertainty of the remotely sensed LAI. In most cases, it is difficult to accurately determine this standard deviation. In this study, we represented the uncertainty of the 1-km observational

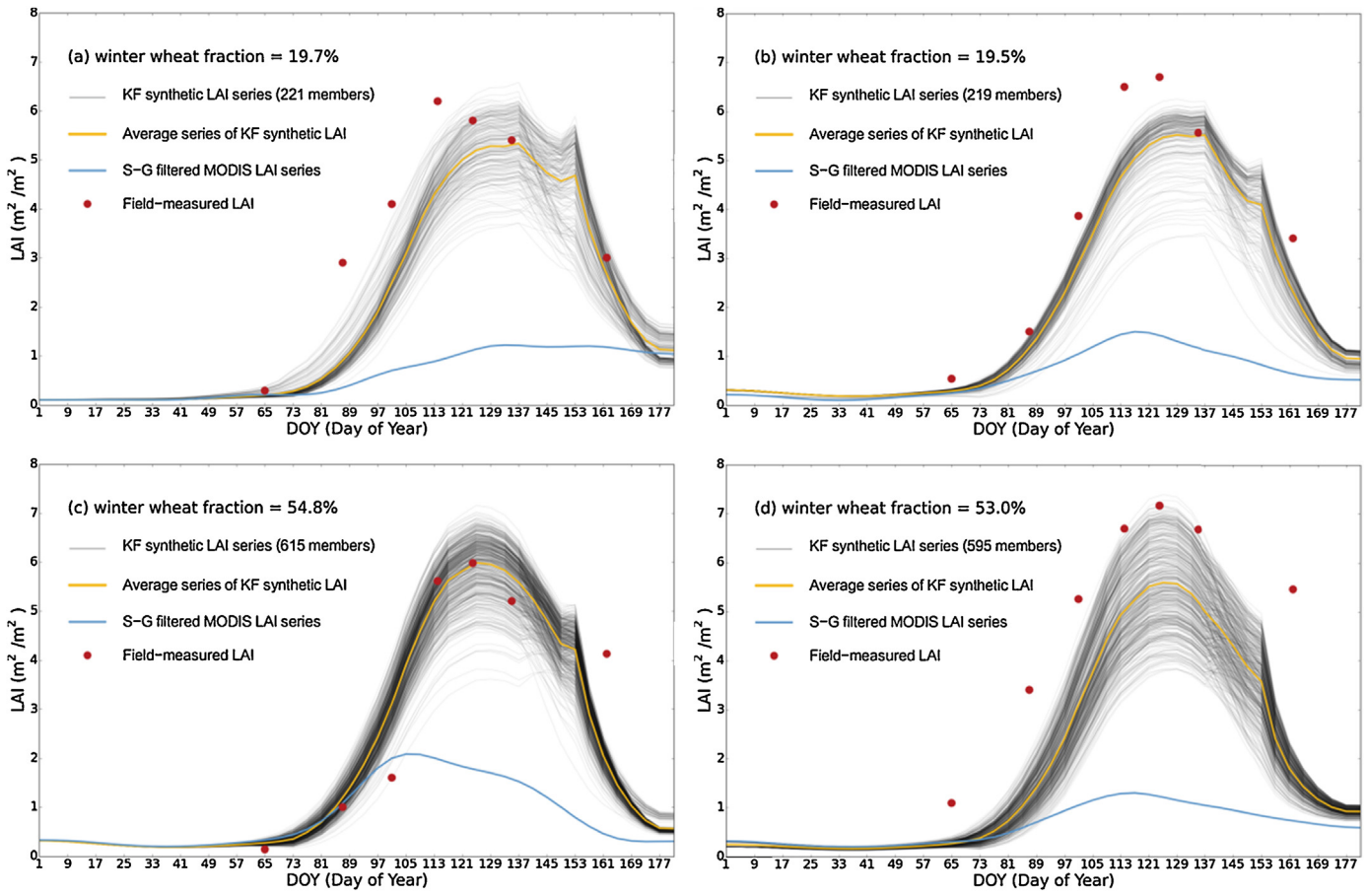

Fig. 7. Temporal variations in the 30-m synthetic KF LAI series within a 1-km grid cell.

LAI data using an ensemble of 30-m synthetic KF LAI values within each 1-km grid cell, which allows the propagation of non-Gaussian distributions through nonlinear models. Fig. 7 shows the temporal variations of the 30-m synthetic KF LAI series within a 1-km grid cell. The phenological trends were similar for the synthetic LAI series, with differences only in the magnitude of the LAI values, possibly caused by different growing conditions for winter wheat. Our results showed that the 30-m synthetic KF LAI usually contained large errors when a pixel contained a large proportion of built-up land and other crops because the S-G filtered MODIS LAI profile cannot adequately represent the temporal variation in the phenological characteristics of winter wheat under these conditions (Fig. 7a and b). The average 30-m synthetic KF LAI was close to the field-measured LAI when the fraction of winter wheat was more than 50% and the built-up fraction was less than 25% (Fig. 7c and d). Based on this analysis, higher pixel purity improves the accuracy of the synthetic KF LAI. Thus, a threshold for at least 50% pixel purity for winter wheat was chosen and another threshold for less than 25% built-up land in the rest of the analyses. The results indicated a significant improvement over the traditional method that used the Gaussian distribution to generate the observational ensemble members.

5.5. Comparison of the LAI trajectories after data assimilation at a field scale

We compared the three assimilated LAI trajectories with and without EnKF-based assimilation at a field scale (Fig. 8). The results show that the S-G filtered MODIS LAI values were lower than the other values throughout the growing season, thus the values in the assimilated LAI series remained low. This can be explained by the fact that the low LAI values in the S-G filtered series would force the WOFOST model to simulate unrealistically low assimilated LAI values.

The Landsat TM LAI dataset contains data from only three growth stages (green-up, anthesis, and maturity). After the Landsat TM LAI data from the green-up stage were assimilated into WOFOST, the LAI trajectory depended strongly on WOFOST for forward simulation. When the Landsat TM LAI data at the anthesis and maturity stages were assimilated into WOFOST, drastic fluctuations appeared in the assimilated LAI trajectories due to the limited number of LAI observations used in the data-assimilation procedure. From the green-up to the anthesis stage, the assimilated LAI curve followed the trend of the WOFOST-simulated LAI curve without assimilation.

The time series data for the synthetic KF LAI values with a 4-day time step were also assimilated into the WOFOST model. Both the MODIS LAI time series and the relatively accurate Landsat TM LAI values are inputs for the synthetic KF LAI. Despite having lower assimilated LAI values than both the WOFOST-simulated LAI and the synthetic KF LAI during the main growing season, the assimilated LAI series improved the representation of the heading stage. In addition, the assimilated LAI gradually diverged from the synthetic KF LAI and became closer to the WOFOST-simulated LAI (in the case of slight filter divergence) in the post-anthesis period despite the use of an inflation factor.

We also compared the importance of the pre-heading and post-heading stages in the assimilation scheme. The results showed that assimilating LAI during the pre-heading stage ($R^2 = 0.51$ and RMSE = 580 kg ha^{-1}) achieved better accuracy than assimilating LAI during the post-heading stage ($R^2 = 0.42$ and RMSE = 630 kg ha^{-1}). There are three possible reasons for this. The first is that our

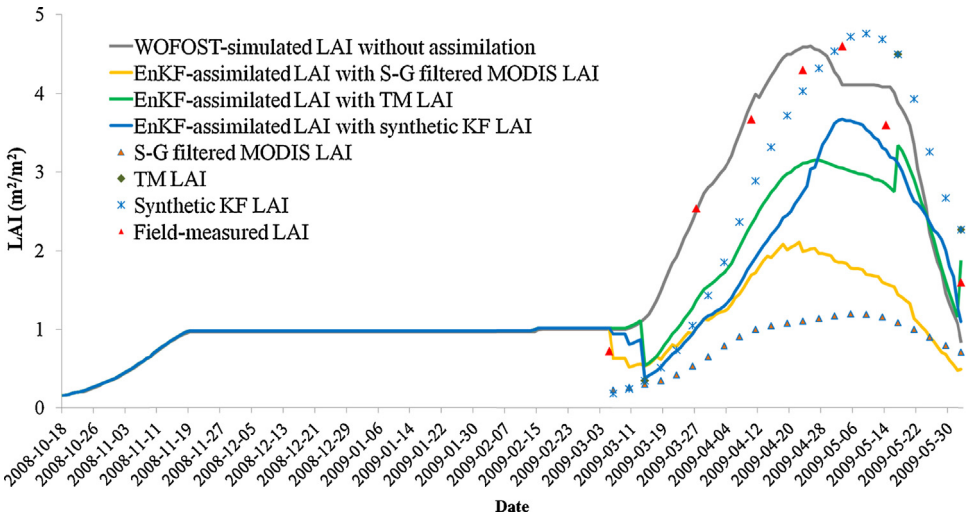

Fig. 8. Comparison of the LAI curves with and without EnKF-based assimilation at a field scale.

selection of pixels with at least 50% wheat means that during the pre-heading period, these pixels were generally not affected by the presence of other crops (i.e., because winter wheat is the only crop before the heading stage in the study area). In contrast, the LAI trajectory during the post-heading period tends to be influenced by other simultaneously emerging summer crops (e.g., cotton, soybean). The second reason is that the crop leaves start turning yellow and drying during the post-heading stage, and LAI decreases, which results in a weak correlation between the LAI at this stage and the final yield. The third possible reason is that a shift in the timing of the heading stage in the EnKF-assimilated LAI during the later stages of assimilation is caused by filter divergence, which occurred despite our use of an inflation factor to correct for this problem. Our results agree with the results from several previous studies, which suggested that late-season data are not suitable for data assimilation (Launay and Guerif, 2005; Dente et al., 2008; Machwitz et al., 2014; Huang et al., 2015b). However, our results disagree with those of Kouadio et al. (2012). The reason for this discrepancy may be the different crop phenological and planting patterns of the different agricultural areas in the two studies and the different data-assimilation strategies that were applied.

5.6. EnKF assimilation of the three remotely sensed LAI datasets into WOFOST at a regional scale

The EnKF assimilation was implemented for the 1-km MODIS grid cells with at least 50% winter wheat pixel purity. Then, these simulated wheat yields were aggregated to obtain county-level yield estimates, and the estimated yields based on the three remotely sensed LAI datasets were compared with government statistics for each region. The wheat yield throughout the study region averaged 4482 kg ha^{-1}, and most of the yield estimates were within the range from 2000 to 5000 kg ha^{-1}.

The WOFOST simulation without LAI assimilation could not estimate the winter wheat yield well. The aggregated winter wheat yields without LAI assimilation had a low coefficient of determination despite the small error (R^2 = 0.12, RMSE = 647 kg ha^{-1}; Fig. 9a). Although the WOFOST simulation captured some of the spatial variability of wheat yield (Fig. 10a), it generally overestimated the wheat yields, with an average value of 6104 kg ha^{-1} compared to the mean official government statistical value of 5711 kg ha^{-1}.

The 1-km S-G filtered MODIS LAI time series from emergence to maturity was directly assimilated into the WOFOST model using the EnKF-based assimilation strategy. The results indicated a large estimation error but a moderate coefficient of determination (R^2 = 0.49, RMSE = 1355 kg ha^{-1}; Fig. 9b). This can be explained by the low LAI values from the S-G filtered MODIS LAI in wheat fields, which would force the WOFOST model to produce unrealistically low wheat yield values (Fig. 10b). Thus, the MODIS LAI products for winter wheat areas are not suitable for being directly used to estimate crop yield in the crop model-data assimilation framework.

The 30-m-resolution Landsat TM LAI values for three dates that corresponded to three stages of crop development (green-up, anthesis, and maturity) were upscaled to the 1-km resolution, and were also assimilated into the WOFOST model. The yield estimation accuracy improved (R^2 = 0.37, RMSE = 472 kg ha^{-1}) compared to the case without data assimilation (R^2 = 0.12, RMSE = 647 kg ha^{-1}). The winter wheat yield with data assimilation averaged 5539 kg ha^{-1}, versus 5711 kg ha^{-1} in the official statistics (Fig. 9c). Assimilation of the TM LAI data captured more information on the spatial variability of winter wheat yields throughout the study area because of the high spatial resolution of the TM LAI (Fig. 10c). This demonstrates the importance of accurate remotely sensed LAI values for improving the performance of the crop model data-assimilation model.

The synthetic KF LAI values from seedling emergence to maturity, at a 4-day time step, were also assimilated into the WOFOST model, and this noticeably improved the accuracy of the estimates, with a better coefficient of determination (R^2 = 0.43) and the smallest RMSE (439 kg ha^{-1}) (Fig. 9d). This can be explained by the higher temporal and spatial resolution of the synthetic KF LAI series, which resulted from integrating the S-G MODIS LAI time series with the more accurate Landsat TM LAI values. The average yield throughout the study region was 5603 kg ha^{-1}, which is very close to the average official statistical yield of 5711 kg ha^{-1}. The estimated yields were reasonable, and ranged from 4591 to 6483 kg ha^{-1}. Fig. 10d shows that the eastern counties (e.g., Wen'an, Jinghai, Dacheng, Qingxian, Cangxian, Cangzhou) had the lowest wheat yield (less than 5000 kg ha^{-1}) and that the western counties (e.g., Shenze, Jinzhou, Xinji) had the highest wheat yield (greater than 6500 kg ha^{-1}), whereas the central counties had intermediate wheat yields (from 5000 to 6500 kg ha^{-1}). The estimated yield with assimilation of the synthetic KF LAI agreed well with the spatial pattern of official statistical yield at the county level (Fig. 10d).

The relative error (RE) between the simulated and statistical yields was calculated as follows:

$$\mathrm{RE} = \left[\frac{(\mathrm{Yield}_{\mathrm{simulated}} - \mathrm{Yield}_{\mathrm{statistics}})}{\mathrm{Yield}_{\mathrm{statistics}}}\right] \times 100\%$$

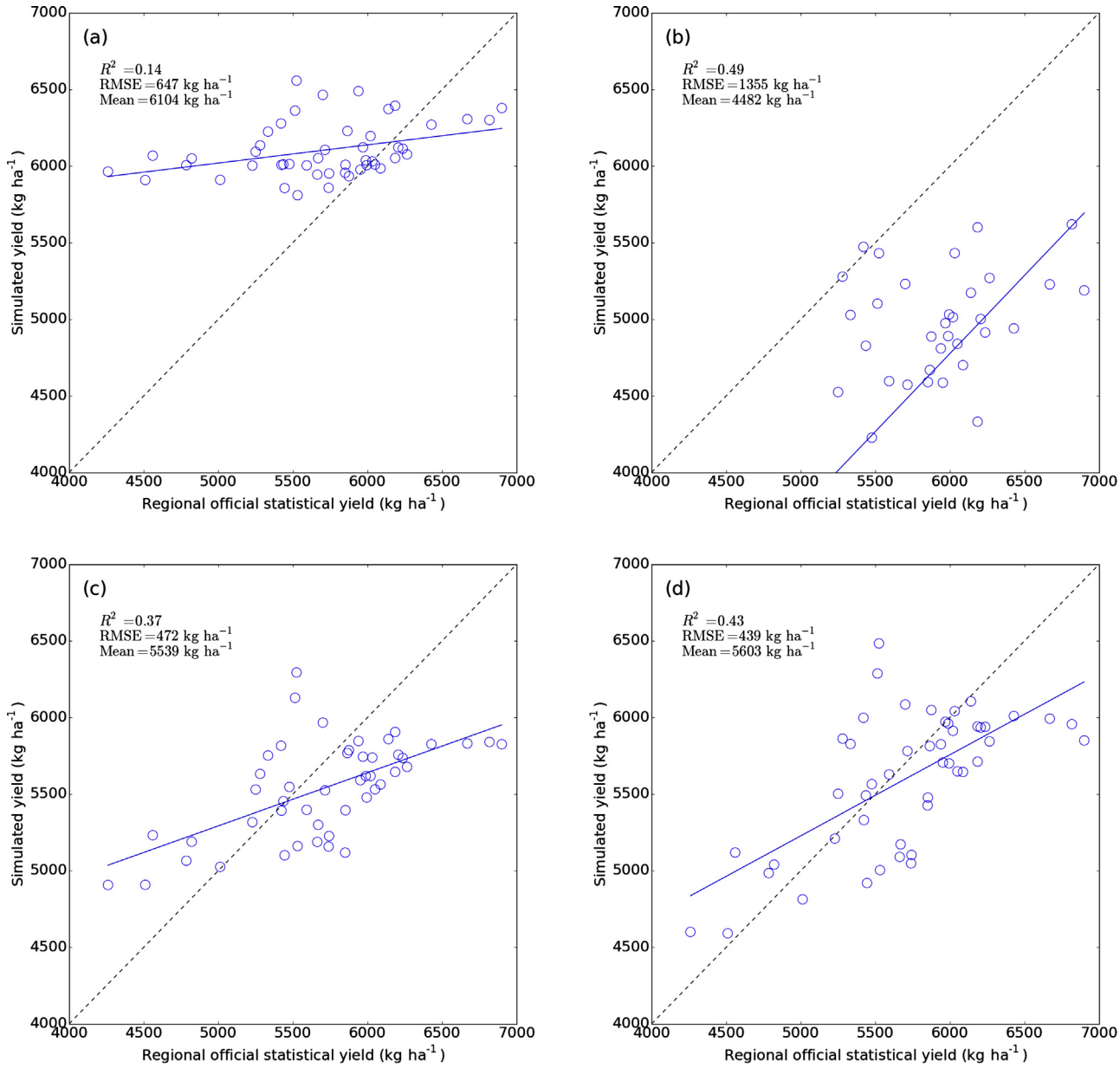

Fig. 9. Estimated winter wheat yield at a regional scale (a) without assimilation and with assimilation using (b) the S-G filtered MODIS LAI, (c) the TM LAI, and (d) the synthetic KF LAI.

This error was used to evaluate the accuracy of the yield estimates using WOFOST without and with data assimilation based on the three remote-sensing LAI datasets for the 53 counties in the study area. The results show that the calibrated WOFOST model without assimilation had a relatively high accuracy, with an average RE = 9.6%. Assimilating the S-G filtered MODIS LAI produced a moderately strongly biased estimate (average RE = 21.8%). The largest RE was observed in Jinghai County (RE = −44.0%), followed by Yongqing County (RE = −40.7%). Assimilating the Landsat TM LAI data from three dates significantly decreased the error (average RE = 7.1%). The largest RE was observed in Jizhou County (RE = −15.5%, followed by Dacheng County (RE = 15.2%). Assimilation of the synthetic KF LAI time series produced the highest accuracy and the smallest RE (RE = 6.1%). For most of the counties, assimilation of the synthetic KF LAI produced the lowest RE of the four strategies. Especially for Wuqiang and Anping counties, the RE with assimilation of the synthetic KF LAI produced much lower values (−0.5 and 0.04%, respectively) than assimilation with the Landsat TM LAI values (−6.2% and −3.8%, respectively). However, RE with assimilation of the synthetic KF LAI was higher than with assimilation of the Landsat TM LAI values in several counties; for example, RE was 11.0 and 6.9% for Fucheng and Hengshui County respectively. Several counties (e.g., Cangzhou, Nangong, Guangzong, Weixian, Dongguang, Wuqiao) were not included in the assimilation analysis because pixels with more than 50% winter wheat were unavailable within these counties.

Using an EnKF-based assimilation process, Curnel et al. (2011) showed that the existence of a phenological shift induced by the joint uncertainties on TSUM1 and crop emergence date biased the yield estimation accuracy. In our study, we had a good estimate of the true sowing date and emergence date throughout the study area. Our results indicated that the EnKF-based assimilation based on two parameters (TDWI and SPAN) that are less sensitive to such phenological shifts significantly improved the relationship between the assimilated yields and the official yield statistics for winter wheat. However, when remotely sensed LAI are assimilated,

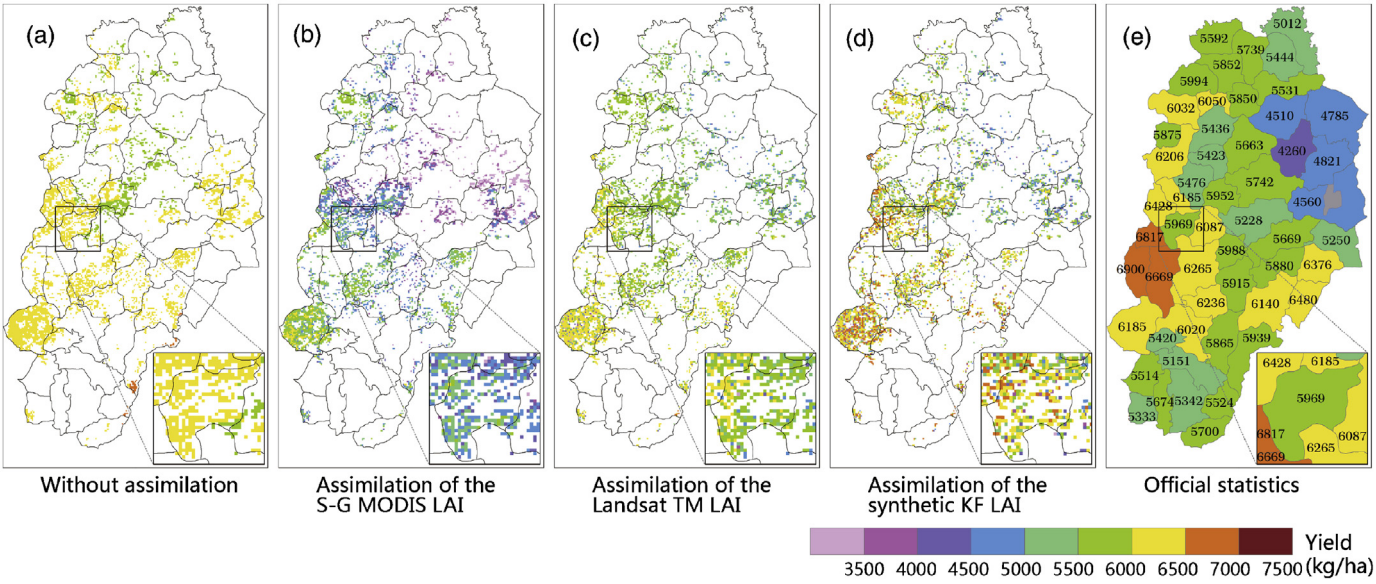

Fig. 10. Spatial distribution of winter wheat yield in the WOFOST model (a) without data assimilation and with assimilation based on (b–d) three different remotely sensed LAI datasets. (e) Official government statistics.

a phenological change (especially at the heading stage) would occur in the assimilated LAI due to the filtering divergence that occurs during the late growing period despite the use of an inflation factor to correct for this problem. Some cells in the grid of EnKF-assimilated LAI values contained large phenological biases when assimilating the synthetic KF LAI series. This might be the main reason why several counties (e.g., Jinzhou, Gu'an, Hejian) had more biased wheat yields when assimilating the synthetic KF LAI series than when assimilating the TM LAI and in the situation without assimilation.

6. Discussion

In the study region, the high heterogeneity of the cultivated land presents significant challenges for the assimilation of remote sensing data to improve estimates of crop yield. This is exacerbated by the fact that field size within these fragmented landscapes is often smaller than the pixel size of the coarse-resolution remote sensing data. This increases pixel heterogeneity and complicates the data-assimilation analysis, particularly when there is a scale mismatch between the remotely sensed image pixels and the model's field simulation. Coarser pixels such as those in the MODIS reflectance data at a scale of 250 m or 1 km usually result in LAI values obtained from highly heterogeneous surfaces, leading to larger scaling errors than would occur with higher-resolution data from satellites such as Landsat TM, ASTER, and RapidEye. Therefore, effective techniques for downscaling 1-km MODIS LAI data to medium-resolution 30-m values will be very important for improving the performance of data assimilation. In this study, we generated the synthetic KF LAI time series based on three Landsat TM image scenes and MODIS LAI time series, and used this synthetic LAI dataset to address the scale mismatch between the satellite observations and the crop model.

The synthetic KF LAI estimates improved the characterization of winter wheat's phenological cycle and the subsequent yield estimates obtained from other remote-sensing sources. Still, the precision and accuracy of these estimates depends on the number and position of the Landsat TM observations. In this study, we acquired cloud-free Landsat TM images during the growing season close to the winter wheat growth stages of green-up, anthesis, and maturity stages. The accuracy of the estimates would benefit from additional Landsat TM observations at key moments in the phenological development of winter wheat. This would result in a more precise representation of crop phenology, its unique features, and any potential anomalies, and would result in lower uncertainties during time steps for which actual LAI images were unavailable and synthetic LAI images were produced. As Earth-observation sensors with better resolution become available (e.g., the Disaster Monitoring Constellation, IRS-LISS III, SPOT, Sentinel 2, CBERS-3 and CBERS-4, Landsat 8 OLI, and China's GF1), the integration of improved images in the present method will enhance our ability to characterize LAI and other biophysical variables. However, as we noted earlier, phenological changes can occur rapidly at certain key times during the development of wheat plants (e.g., at the booting and heading stages), and it may not be possible to capture these changes until remote sensing data with higher temporal resolution and finer spatial resolution becomes available. It is also worth mentioning that the synthetic medium-resolution KF NDVI time series could be assimilated into a coupled crop growth and radiative transfer model to improve the estimation of crop yields over large spatial scales and avoid the problem of LAI retrieval.

In our analysis, we assumed that the dominant winter wheat cultivar for our study area was planted throughout the study area and that the crop characteristics and management measures did not vary spatially, which allowed us to assume that the calibrated model's parameters did not vary within this region. However, both the cultivar choice and the crop management conditions are likely to vary significantly, thereby affecting crop phenology and yield. In future research, it would be interesting to account for those variations in the WOFOST model. In addition, we used the potential mode of the WOFOST model, and therefore did not account for the possible effects of water stress and other yield-limiting factors (e.g., nutrients, pests, and weeds). In future research, we should examine the effects of accounting for these factors. Despite these problems, LAI remained a useful proxy for these and other factors because it comprehensively reflects the net effects of many factors on winter wheat growth in the study area.

In this study, we generated a continuous synthetic KF LAI time series with a 30-m spatial resolution and a 4-day time step. However, data assimilation was conducted at a 1-km scale rather than the 30-m scale that would theoretically have been possible if more Landsat TM data had been available. This also reflects limitations

imposed by the available computational efficiency. The preliminary estimates suggest that the study area includes more than 200 000 winter wheat grid cells at a 30-m resolution, so WOFOST simulations at this resolution would take hundreds of hours. We chose to use 1-km assimilation scale to conduct the data assimilation; even at this coarse scale, with 30-m synthetic KF LAI values within a 1-km grid cells that were used as the ensemble members, computations took more than 50 h with 5352 grid cells that had a winter wheat fraction of at least 50% using a single standard computing system. In future research, it will be helpful to include more detailed meteorological data and data on crop parameters in the simulations, but this will also require advances in computing technology, such as access to powerful multi-core, parallel-processing computers.

In this study, we compared two remotely sensed LAI datasets (e.g., S-G filtered MODIS LAI and TM LAI) based on methods developed during our previous study (Huang et al., 2015b) with a new synthetic LAI dataset using a KF algorithm to evaluate the impact of scale effects on the assimilation accuracy. The S-G filtered MODIS LAI time series included large errors due to the high pixel heterogeneity. The assimilation showed a good coefficient of determination, but a high RMSE compared with the estimates created without data assimilation. Yield estimates improved with the TM LAI images from three dates (i.e., a limited number of measurements but with higher spatial detail), and the yield estimates improved, with a slightly lower coefficient of determination but a significantly decreased RMSE. We obtained the best assimilation results (R^2 = 0.43 and the smallest RMSE, 439 kg ha^{-1}) by assimilating the synthetic KF LAI time series. These results confirmed that the data-assimilation performance of the WOFOST model depends heavily on the accuracy of LAI retrieval and on the spatial and temporal scales of the assimilation.

In our previous research (Huang et al., 2015b), we used a double-logistic regression function to generate a scale-adjusted LAI time series based on a ratio adjustment technique, and assimilated the resulting scale-adjusted LAI into the WOFOST model to significantly reduce RMSE for the estimated winter wheat yields using a recalibration strategy with the 4DVar variational assimilation algorithm. However, to use this method, we assumed that the pre-heading and post-heading stages had a constant ratio between the MODIS LAI and the corresponding Landsat LAI. In reality, this ratio is a phenology-dependent variable. In the present study, we observed overly high LAI values at the heading stage (e.g., greater than 8) when using the ratio adjustment method. However, more reasonable LAI values (i.e., ranging from 3 to 7) at the heading stage were obtained and the overly high LAI values were avoided when we used the KF algorithm in the present study. At a regional scale, both the 4DVar and the EnKF strategies significantly improved the estimation accuracy for winter wheat yield. We found that 4DVar (R^2 = 0.48; RMSE = 151.92 kg ha^{-1}) outperformed EnKF (R^2 = 0.43; RMSE = 439 kg ha^{-1}). This can be explained by the fact that EnKF-based assimilation techniques aim to sequentially correct the uncertainty observed in the modeled state variable without correcting for the causes of this uncertainty. However, variational assimilation attempts to optimize some of the uncertain model input parameters or their initial state using all of the available observations throughout the assimilation window (e.g., the whole growing season), and this demonstrated the advantage of using a larger dataset to improve the performance of the data assimilation. This conclusion is consistent with that of Curnel et al. (2011).

A number of crop model data-assimilation studies have been conducted to estimate regional wheat or maize yields in China (Ma et al., 2008, 2013a, 2013b; Xu et al., 2011; Tian et al., 2013; Wang et al., 2013; Li et al., 2014; Huang et al., 2015a,b). However, it remains unclear which combination of satellite remote sensing data and crop modeling (e.g., input data; calibration; data assimilation scheme) will be most effective in China. The WOFOST model is suitable for data assimilation to support estimation of crop yields by integrating Landsat TM and MODIS data using an EnKF algorithm in the present study or the 4DVar algorithm in our previous study (Huang et al., 2015b). As remote sensing data with a spatial resolution of 10–50 m becomes available, this will improve the effectiveness of data-assimilation schemes to support agricultural production management in China, even in the case of areas with small and heterogeneous fields.

7. Conclusions

In this study, we used the WOFOST process-based crop growth model to estimate winter wheat yield at a regional scale, and compared the yield estimates produced using three LAI datasets with different temporal and spatial resolutions. We found that the 1-km MODIS LAI products could not be directly used to estimate crop yield in the EnKF data-assimilation procedure because of their coarse spatial resolution; despite the high coefficient of determination, the RMSE was unacceptably large. The results based on assimilating Landsat TM data from three dates indicated that accurate remotely sensed LAI values are essential for improving the performance of EnKF-based data assimilation. The integration of LAI information obtained with better resolution sensors (Landsat TM) and lower-resolution sensors (MODIS) within the KF algorithm produced a continuous time series of LAI values with high temporal and spatial resolution. The synthetic KF LAI estimates produced a more realistic characterization of the crop's phenological dynamics. When the 30-m synthetic KF LAI values within a 1-km grid cell were used as the observational ensemble members in the EnKF algorithm, this approach improved the estimates of winter wheat yields at a regional scale. In addition, using pre-heading LAI was more effective for improving the EnKF-based data assimilation's performance than using post-heading LAI. These results confirm the importance of LAI retrieval accuracy and of scaling adjustments between the pixel scale of the remotely sensed observations and the single-point simulation scale of the crop model in the data-assimilation scheme. Furthermore, these results indicate that assimilating the synthetic KF LAI into the WOFOST model with the EnKF strategy is a promising approach to improving crop yield estimation at a large spatial scale in agricultural production operations.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Project No. 41371326, 41471342), National High Technology Research and Development 863 Program in China (2013AA10230103), and the fundamental research funds for the Chinese central universities (No. 2015XD004), and National Aeronautics and Space Administration (NASA) of United States (Grant No. NNX13AJ26G). We thank the journal's editors and reviewers for their efforts to improve the quality of our paper.

References

Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., Camps-Valls, G., 2013. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. Int. J. Appl. Earth Obs. Geoinf. 23, 132–141.

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.

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

Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS experiment. Agric. For. Meteorol. 151 (12), 1843–1855.

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 (4), 1395–1407.

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.J.W., van Diepen, C.A., 2007. Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts. Agric. For. Meteorol. 146 (1–2), 38–56.
Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinf. 9 (2), 165–193.
Duveiller, G., Baret, F., Defourny, P., 2011. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. Remote Sens. Environ. 115 (10), 2686–2701.
Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53 (4), 343–367.
Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., Sobrino, J.A., 2015. Improving timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day. Remote Sens. Environ. 161, 131–148.
Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44 (8), 2207–2218.
Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. Remote Sens. Environ. 105 (4), 286–298.
Huang, C., Li, X., Lu, L., Gu, J., 2008. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens. Environ. 112 (3), 888–900.
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. 8 (8), 4060–4071.
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.
Hufkens, K., Bogaert, J., Dong, Q.H., Lu, L., Huang, C.L., Ma, M.G., Che, T., Li, X., Veroustraete, F., Ceulemans, R., 2008. Impacts and uncertainties of upscaling of remote-sensing data validation for a semi-arid woodland. J. Arid Environ. 72 (8), 1490–1505.
Ines, A., Das, N., Hansen, J., Njoku, E., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149–164.
Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agric. Ecosyst. Environ. 111 (1–4), 321–339.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. ASME J. Basic Eng. 82 (Series D), 35–45.
Kouadio, L., Duveiller, G., Djaby, B., Jarroudi, E., Defourny, P., Tychon, B., 2012. Estimating regional wheat yield from the shape of decreasing curves of green area index temporal profiles retrieved from MODIS data. Int. J. Appl. Earth Obs. Geoinf. 18, 111–118.
Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., Wang, J., 2014. Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. Ecol. Model. 291, 15–27.
Liang, S., Fang, H., Chen, M., Shuey, C.J., Walthall, C., Daughtry, C., Morisette, J., Schaaf, C., Strahler, A., 2002. Validating MODIS land surface reflectance and albedo products: methods and preliminary results. Remote Sens. Environ. 83, 149–162.
Liang, S., Qin, J., 2008. Data assimilation methods for land surface variable estimation. In: Liang, S. (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application. Springer, New York, pp. 319–339.
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.
Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., Wu, S., 2013a. Assimilation of MODIS-LAI into WOFOST model for forecasting regional winter wheat yield. Math. Comput. Modell. 58 (3–4), 634–643.
Ma, H., Huang, J., Zhu, D., Liu, J., Zhang, C., Su, W., Fan, J., 2013b. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD into WOFOST-ACRM model. Math. Comput. Modell. 58 (3–4), 753–764.
Ma, Y., Wang, S., Zhang, L., How, Y., Zhang, L., He, Y., Wang, F., 2008. Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 10 (4), 426–437.
Machwitz, M., Guistarini, L., Bossung, C., et al., 2014. Enhanced biomass prediction by assimilating satellite data into a crop growth model. Environ. Modell. Softw. 62, 437–453.
Mathieu, P., O'Neill, A., 2008. Data assimilation: from photon counts to Earth System forecasts. Remote Sens. Environ. 112 (4), 1258–1267.
Maybeck, P., 1979. Stochastic Models, Estimation, and Control, Vol. 1. Academic Press, New York, pp. 1–16.
Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. Remote Sens. Environ. 92 (2), 195–206.
Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G., Lotsch, A., Friedl, M., Morisette, J., Votava, P., Nemani, R., Running, S., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83 (1–2), 214–231.
Quaife, T., Lewis, P., Dekauwe, M., Williams, M., Law, B., Disney, M., Bowyer, P., 2008. Assimilating canopy reflectance data into an ecosystem model with an ensemble Kalman filter. Remote Sens. Environ. 112 (4), 1347–1364.
Rauch, H.E., 1963. Solutions to the linear smoothing problem. IEEE Trans. Autom. Control 8 (4), 371–372.
Roy, D., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112 (6), 3112–3130.
RSI, 2001. ENVI User's Guide, September 2001 edition. Research Systems.
Samain, O., Roujean, J., Geiger, B., 2008. Use of a Kalman filter for the retrieval of surface BRDF coefficients with a time-evolving model based on the ECOCLIMAP land cover classification. Remote Sens. Environ. 112 (4), 1337–1346.
Sedano, F., Kempeneers, P., Hurtt, G., 2014. A Kalman filter-based method to generate continuous time series of medium-resolution NDVI images. Remote Sens. 6 (12), 12381–12408.
Tian, L., Li, Z., Huang, J., Wang, L., Su, W., Zhang, C., Liu, J., 2013. Comparison of two optimization algorithms for estimating regional winter wheat yield by integrating MODIS leaf area index and world food studies model. Sensor Lett. 11 (6–7), 1261–1268.
Vazifedoust, M., van Dam, J.C., Bastiaanssen, W.G.M., Feddes, R.A., 2011. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. Int. J. Remote Sens. 30 (10), 2523–2545.
Vicente-Guijalba, F., Martinez-Marin, T., Lopez-Sanchez, J.M., 2014. Crop phenology estimation using a multitemporal model and a Kalman filtering strategy. IEEE Trans. Geosci. Remote Sens. 11 (6), 1081–1085.
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.
Welch, G., Bishop, G., 1995. An introduction to the Kalman filter. Technical Report 95-041. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, pp. 1–16.
Xu, W., Jiang, H., Huang, J., 2011. Regional crop yield assessment by combination of a crop growth model and phenology information derived from MODIS. Sensor Lett. 9 (3SI), 981–989.
Zhao, F., Li, Y., Dai, X., Verhoef, W., Guo, Y., Shang, H., Gu, X., Huang, Y., Yu, T., Huang, J., 2015. Simulated impact of sensor field of view and distance on field measurements of bidirectional reflectance factors for row crops. Remote Sens. Environ. 156, 129–142.
Zhao, Y., Chen, S., Sheng, S., 2013. Assimilating remote sensing information with crop model using ensemble Kalman filter for improving LAI monitoring and yield estimation. Ecol. Model. 270, 30–42.
Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114 (11), 2610–2623.
Zurita-Milla, R., Kaiser, G., Clevers, J.G.P.W., Schneider, W., Schaepman, M.E., 2009. Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. Remote Sens. Environ. 113 (9), 1874–1885.