2015

Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model

Jianxi Huang
China Agricultural University, jxhuang@cau.edu.cn

Liyan Tian
China Agricultural University

Shunlin Liang
Beijing Normal University

Hongyuan Ma
China Agricultural University

Inbal Becker-Reshef
University of Maryland

See next page for additional authors

Follow this and additional works at: https://digitalcommons.unl.edu/usdaarsfacpub

Huang, Jianxi; Tian, Liyan; Liang, Shunlin; Ma, Hongyuan; Becker-Reshef, Inbal; Huang, Yanbo; Su, Wei; Zhang, Xiaodong; Zhu, Dehai; and Wu, Wenbin, "Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model" (2015). *Publications from USDA-ARS / UNL Faculty*. 2088.
https://digitalcommons.unl.edu/usdaarsfacpub/2088

Authors

Jianxi Huang, Liyan Tian, Shunlin Liang, Hongyuan Ma, Inbal Becker-Reshef, Yanbo Huang, Wei Su, Xiaodong Zhang, Dehai Zhu, and Wenbin Wu
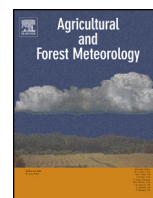
Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model

Jianxi Huang [a,*], Liyan Tian [a], Shunlin Liang [b,c], Hongyuan Ma [a], Inbal Becker-Reshef [c], Yanbo Huang [d], Wei Su [a], Xiaodong Zhang [a], Dehai Zhu [a], Wenbin Wu [e]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China
[c] Department of Geographical Sciences, University of Maryland, College Park, MD 20740, USA
[d] United States Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, Stoneville, MS, USA
[e] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

A B S T R A C T

To predict regional-scale winter wheat yield, we developed a crop model and data assimilation framework that assimilated leaf area index (LAI) derived from Landsat TM and MODIS data into the WOFOST crop growth model. We measured LAI during seven phenological phases in two agricultural cities in China's Hebei Province. To reduce cloud contamination, we applied Savitzky–Golay (S–G) filtering to the MODIS LAI products to obtain a filtered LAI. We then regressed field-measured LAI on Landsat TM vegetation indices to derive multi-temporal TM LAIs. We developed a nonlinear method to adjust LAI by accounting for the scale mismatch between the remotely sensed data and the model's state variables. The TM LAI and scale-adjusted LAI datasets were assimilated into the WOFOST model to allow evaluation of the yield estimation accuracy. We constructed a four-dimensional variational data assimilation (4DVar) cost function to account for the observations and model errors during key phenological stages. We used the shuffled complex evolution–University of Arizona algorithm to minimize the 4DVar cost function between the remotely sensed and modeled LAI and to optimize two important WOFOST parameters. Finally, we simulated winter wheat yield in a 1-km grid for cells with at least 50% of their area occupied by winter wheat using the optimized WOFOST, and aggregated the results at a regional scale. The scale adjustment substantially improved the accuracy of regional wheat yield predictions ($R^2 = 0.48$; RMSE = 151.92 kg ha^{-1}) compared with the unassimilated results ($R^2 = 0.23$; RMSE = 373.6 kg ha^{-1}) and the TM LAI results ($R^2 = 0.27$; RMSE = 191.6 kg ha^{-1}). Thus, the assimilation performance depends strongly on the LAI retrieval accuracy and the scaling correction. Our research provides a scheme to employ remotely sensed data, ground-measured data, and a crop growth model to improve regional crop yield estimates.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Climate fluctuations and reductions in the area of cultivated land have increasingly threatened the wheat crop of China, the world's second-largest wheat producer (FAO, 2012), creating a major national concern over food security. Winter wheat comprises about 85% of China's total summer grain production, Therefore, accurate regional monitoring of wheat growth and yield prediction have become crucial for national food security and sustainable agricultural development in China. However, most yield-prediction methods still depend on conventional techniques, including predictions from agro-meteorological models and empirical statistical regression models between spectral vegetation indices and field-measured yields. One of the main drawbacks of such empirical regression models for estimating crop yields is that the models are only applicable for specific crop cultivars, crop growth stages, or certain geographical regions (Doraiswamy et al., 2003; Fang et al., 2011).

In contrast, process-oriented crop simulation models based on mathematical descriptions of key physical and physiological processes offer powerful tools to simulate the physiological development, growth, and yield of a given crop based on the interactions among environmental characteristics such as the climate, crop

management, soil conditions, and plant physiological processes such as photosynthesis and respiration. Several previous studies have confirmed that such crop growth models can be successfully applied to crop yield prediction at a field scale (Jégo et al., 2012; Moulin et al., 1998). However, their practical application at a regional scale is restricted by uncertainties in the model's structure and processes, and especially uncertainties in the input parameters and initial conditions of the model. Therefore, there is increasing interest in providing better estimates of model state variables and input parameters so as to improve the model's ability to simulate crop growth (Dorigo et al., 2007).

Remotely sensed data offers strong advantages over other monitoring techniques by providing a timely, synoptic, and up-to-date overview of actual crop growing conditions over large areas at multiple stages during the growing season, and the data can be utilized in conjunction with crop models to improve prediction of crop yields at a range of spatial scales (Liang and Qin 2008). Furthermore, remotely sensed data can be used to complement crop model simulation results under situations that are not accounted for by the model (de Wit et al., 2012). Thus, data assimilation, an approach that incorporates field or other observations into dynamic mechanistic models, can produce more accurate estimates of model input parameters and state variables, and this approach has increasingly been used for crop growth monitoring and yield prediction, with considerable success (Curnel et al., 2011; Dente et al., 2008; de Wit and van Diepen, 2007; de Wit et al., 2012; Fang et al., 2008, 2011; Ma et al., 2013a; Tian et al., 2013; Wang et al., 2013).

It is widely acknowledged that regional crop yield estimates using crop models can be improved by assimilating the values of biophysical variables derived from remotely sensed data obtained during the growing season. There are two overall groups of strategies for data assimilation: variational assimilation algorithms and sequential algorithms (Liang and Qin 2008). The main difference between the two groups is that each subsequent observation for sequential assimilation will influence the nature of the change from the current state of the model; in contrast, variational assimilation adjusts the estimation using all of the available observations throughout the assimilation window. Variational assimilation offers the advantage of using a larger dataset to improve the precision of each estimation (Curnel et al., 2011; Liang and Qin 2008). The variational methods start by constructing a cost function with respect to the control variables, which comprise state variables and model parameters that must be estimated for the system simulation.

Several variational assimilation schemes with different degrees of complexity and model integration have been developed and evaluated during the last decade, and the results suggested that they have tremendous potential for predicting regional crop yield. Curnel et al. (2011) compared a variational algorithm with a sequential algorithm (the ensemble Kalman filter: EnKF) to estimate wheat yield, and found that the variational algorithm achieved better accuracy. Fang et al. (2011) integrated the CERES-Maize model with the MODIS LAI products using a simplified variational method based on the Powell optimization algorithm to predict corn yield in Indiana, United States. They found that the predicted corn yield agreed well with the USDA statistical data for most of the study area. Dente et al. (2008) assimilated the LAI values derived from ENVISAT ASAR and MERIS data into the CERES-Wheat model using a variational algorithm to improve prediction of the regional wheat yield. This process reinitialized the model by optimizing the input parameters, which required good temporal agreement between the LAI values simulated by the crop model and estimates derived from remote-sensing data. Xu et al. (2011) used the shuffled complex evolution–University of Arizona (SCE–UA) algorithm to assimilate the phenological information derived from the MODIS LAI trajectory into the WOFOST model after optimizing

the emergence date and minimum temperature for growth, and improved the prediction of regional winter wheat yield.

Due to the variability of land cover and the complexity of the crop planting pattern in agricultural landscapes, the scale mismatch between the remotely sensed observations and the state variables of crop growth models remains a difficult challenge. In most of the reported approaches for agricultural data assimilation frameworks, the scale mismatch between pixel-scale remotely sensed observational data and the single-point scale of the crop models has not been fully taken into account, and this can greatly decrease the performance of the data assimilation. To support an agricultural data assimilation system, remote sensing must combine short revisit intervals with large geographical coverage. Most widely used satellite sensors provide low spatial resolution (e.g., the AVHRR, MODIS, MERIS, and SPOT Vegetation instruments). Although these sensors have the advantage of capturing crop phenological development and variability for pixels that contain a high proportion of a single crop due to their high temporal resolution, their coarse spatial resolution increases the intra-pixel heterogeneity. Thus, most researchers have only investigated data assimilation practices in relatively homogeneous agricultural areas to reduce these errors (Bastiaanssen, 2003; Fang et al., 2008; Ma et al., 2008; Mo et al., 2005; Xu et al., 2011). Furthermore, the retrieval algorithm for MODIS LAI products was designed for global-scale applications with all vegetation types, not to account for specific agricultural crops, and generally tends to underestimate crop LAI (Duveiller et al., 2013; Fang et al., 2012). Finally, there is a mismatch between the nature of the remotely sensed LAI values and the LAI simulated by crop models. For example, the LAI used by the WOFOST model is actually a "green area index" (GAI), since it includes the contributions of stems and storage organs (Duveiller et al., 2011b; de Wit et al., 2012). When LAI is required for specific crop monitoring and applications, several studies have improved retrieval performance by using a filtering procedure (e.g., a canopy structural dynamics model) to generate a time series for crop LAI based on the 250-m-scale daily reflectance data and thermal data during several phenological stages (Duveiller et al., 2012, 2013).

The scale mismatch between remotely sensed observations and a crop model's state variables can be largely overcome by using instruments with high spatial resolution and wide swath coverage, such as the Disaster Monitoring Constellation and the forthcoming Sentinel 2. Unfortunately, a series of cloud-free images with fine spatial resolution can seldom be acquired, because the time when the crop canopy is growing most actively coincides with the cloudy and rainy season in many parts of the world. Furthermore, modeling the spatial heterogeneity with two widely used methods (correcting the scaling bias and downscaling) can be a complex issue, requiring rigorous approximations and a priori knowledge that might not be readily available for operational applications (Duveiller et al., 2011a). One potential solution for the scale mismatch is to combine the phenological information from sensors with low spatial resolution but high revisit frequency (e.g., MODIS) and relatively accurate LAI values derived from medium-resolution images (e.g., Landsat TM) to produce a scale-adjusted LAI trajectory during the crop growing season.

Intra-pixel heterogeneity is also a challenging issue when conducting data assimilation using remote-sensing data with coarse spatial resolution, particularly over complex agricultural landscapes. The analysis can be focused on a subset of the pixels that contain a high fraction of a single crop instead of using all of the pixels (de Wit et al., 2012). Becker-Reshef et al. (2010) used a mask based on the percentage of a pixel covered by the target crop as a filter to identify the purest winter wheat pixels at a county level, and used the mask to obtain high-accuracy predictions of regional wheat yields. A related problem in an agricultural data assimilation framework is that crop growth models are often specific to a given

crop (or crop variety). In order to couple them with remotely sensed data, it is preferable to use crop-specific masking (Duveiller et al., 2011a), thereby reducing the noise generated by other land cover or crop types. However, there has been little research to determine the effects of the spatial and temporal scales of remotely sensed LAI datasets combined with a crop-specific mask on the accuracy of data assimilation.

The overall goal of the present study was to improve the accuracy of estimation of winter wheat yield based on variational assimilation using LAI datasets derived from Landsat TM and MODIS data. To accomplish this goal, we defined the following specific objectives:

1. To determine whether integration of multi-temporal Landsat TM data and MODIS LAI time series can improve on the crop LAI time series provided by a single sensor.
2. To explore the relative importance of LAI at different phenological stages for estimating wheat yields at the field scale.
3. To validate the assimilation accuracy for LAI datasets derived from Landsat TM and MODIS data by comparing the assimilated LAI and wheat yield with field-measured data at the field scale, and further compare the assimilated wheat yields with official statistics at a regional scale.

2. Data and methods

2.1. Study area

The study was conducted in a planting area dominated by winter wheat in two prefecture-scale agricultural cities, Baoding and Hengshui (115°10′E to 116°34′E and 37°03′N to 39°36′N), in the southern part of China's Hebei Province (Fig. 1). The region covers approximately 16,335 km², and consists of 24 counties. The prevailing planting pattern is dominated by an intensive dual-cropping system based on winter wheat and summer crops, including maize, soybean, and cotton. The region is characterized by alluvial plains. The climate is a continental monsoon climate with average annual rainfall ranging from 400 mm to 800 mm, and an average annual temperature ranging from 9 °C to 15 °C. The soil texture is primarily a loam, with abundant organic matter. High winter wheat yields are traditionally reported from this region, where the soil and climate conditions and adequate irrigation from groundwater make the region suitable for winter wheat growth. Generally, winter wheat is planted at the beginning of October and harvested in early or mid-June in the next year.

2.2. Field data

We chose a mechanistic model to simulate crop growth in this study (see Section 2.3 for details). To calibrate and validate the model for use in our study region, field experiments were performed at the Gucheng Ecological–Meteorological Integrated Observation Experiment Station, which is located east of Gucheng Town, in Dingxing County of Hebei Province (39°08′N, 115°40′E, elevation of 15.2 m above sea level). An automated weather station was installed for long-term observation at this station to measure daily maximum and minimum temperatures, solar radiation, wind speed, actual vapor pressure, and precipitation, which were used to drive the crop growth model at the field scale. Winter wheat (the 'Hengguan 35' cultivar) was planted at the station. The phenological stages and genetic-specific parameters required by the WOFOST model during the growing season and the soil physical properties were measured at the station. Two important crop-specific
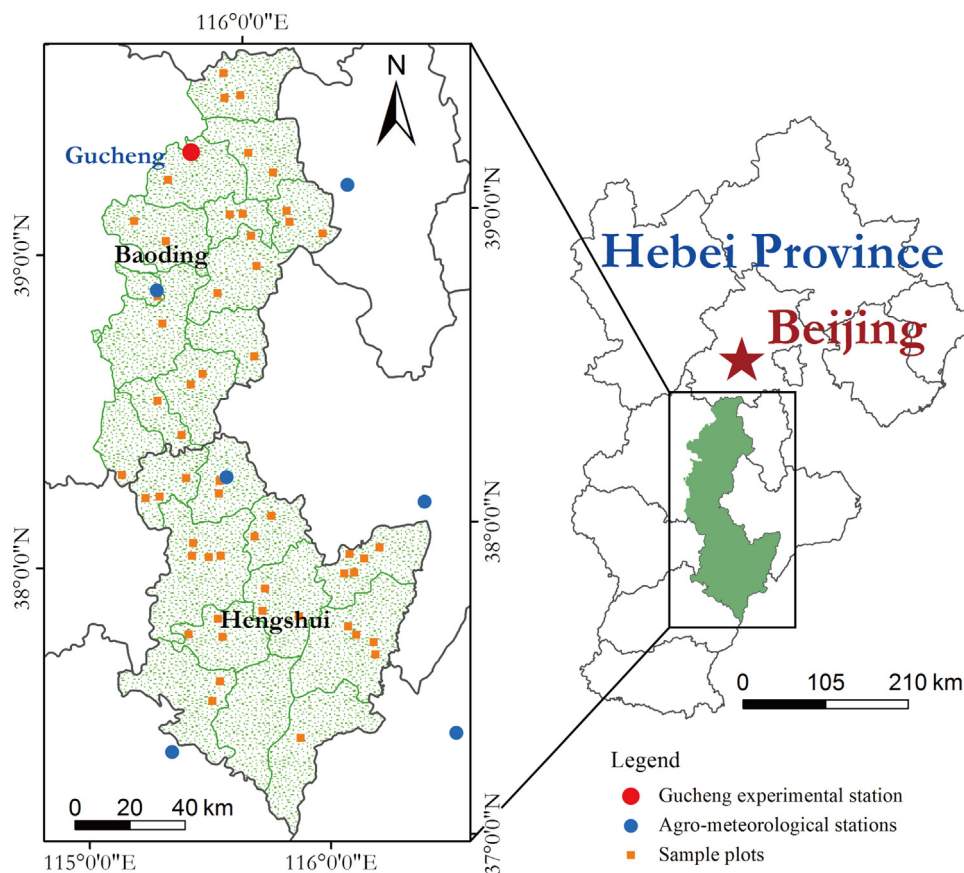

Fig. 1. Location of the study area and of the sample plots in the 24 counties of southern Hebei Province, China.

Table 1
Parameter values used for calibration of the WOFOST model.

Parameter	Description	Units	Value	Source
Crop initial condition parameters				
TDWI[a]	Initial total crop dry weight	kg ha^{-1}	210	Field measurements
LAIEM	Leaf area index at emergence	ha ha^{-1}	0.13	Default value for wheat in WOFOST
RGRLAI	Maximum relative rate of increase in LAI	ha ha^{-1} d^{-1}	0.00817	Default value for wheat in WOFOST
Green area parameters				
TBASE	Lower threshold temperature for aging of leaves	°C	0	Ma et al. (2013a)
SPAN[a]	Life span of leaves growing at 35 °C	days	27	Field measurements
SLATB00	Specific leaf area at DVS = 0	ha kg^{-1}	0.00224	Field measurements
SLATB050	Specific leaf area at DVS = 0.5	ha kg^{-1}	0.00210	Field measurements
SLATB200	Specific leaf area at maturity	ha kg^{-1}	0.00195	Field measurements
Phenological parameters				
DLO	Optimum day length for development	h	14	Ma et al. (2013a)
DLC	Critical day length	h	8	Ma et al. (2013a)
TSUM1	Cumulative temperature from emergence to anthesis	°C	891.0	Field measurements
TSUM2	Cumulative temperature from anthesis to maturity	°C	672.0	Field measurements
Assimilation parameters				
AMAXTB00	Maximum leaf CO_2 assimilation rate (DVS = 0)	kg ha^{-1} h^{-1}	45	Field measurements
AMAXTB100	Maximum leaf CO_2 assimilation rate (DVS = 1)	kg ha^{-1} h^{-1}	45	Field measurements
AMAXTB130	Maximum leaf CO_2 assimilation rate (DVS = 1.3)	kg ha^{-1} h^{-1}	45	Field measurements
AMAXTB200	Maximum leaf CO_2 assimilation rate (DVS = 2)	kg ha^{-1} h^{-1}	4.8	Field measurements
KDIFFTB00	Extinction coefficient for diffuse visible light (DVS = 0)	–	0.61	Default value for wheat in WOFOST
KDIFFTB200	Extinction coefficient for diffuse visible light (DVS = 2)	–	0.61	Default value in WOFOST
EFFTB0	Light-use efficiency of a single leaf (T = 0 °C)	kg ha^{-1} h^{-1} J^{-1} m^2 s^{-1}	0.47	Field calibration
EFFTB40	Light-use efficiency of a single leaf (T = 40 °C)	kg ha^{-1} h^{-1} J^{-1} m^2 s^{-1}	0.35	Field calibration
TMPFTB10	Reduction factor of AMAX (T = 10 °C)	–	0.60	Field calibration
TMPFTB15	Reduction factor of AMAX (T = 15 °C)	–	0.70	Field calibration
TMPFTB25	Reduction factor of AMAX (T = 25 °C)	–	0.85	Field calibration
TMPFTB35	Reduction factor of AMAX (T = 35 °C)	–	0.97	Field calibration
Conversion of assimilation into biomass parameters				
CVL	Conversion efficiency of assimilates into leaf tissue	kg kg^{-1}	0.740	Field measurements
CVO	Conversion efficiency of assimilates into storage organs	kg kg^{-1}	0.791	Field measurements
CVR	Conversion efficiency of assimilates into root tissue	kg kg^{-1}	0.694	Field measurements
CVS	Conversion efficiency of assimilates into stem tissue	kg kg^{-1}	0.740	Field measurements
Maintenance respiration parameters				
Q10	Relative change in respiration rate per 10 °C temperature increase	–	2	Default value for wheat in WOFOST
RML	Relative maintenance respiration rate of leaves	kg CH_2O kg^{-1} d^{-1}	0.03	Default value for wheat in WOFOST
RMO	Relative maintenance respiration rate of storage organs	kg CH_2O kg^{-1} d^{-1}	0.01	Default value for wheat in WOFOST
RMR	Relative maintenance respiration rate of roots	kg CH_2O kg^{-1} d^{-1}	0.015	Default value for wheat in WOFOST
RMS	Relative maintenance respiration rate of stems	kg CH_2O kg^{-1} d^{-1}	0.015	Default value for wheat in WOFOST
Partitioning parameters				
FRTB00	Fraction of total dry matter to roots at DVS = 0	kg kg^{-1}	0.50	Field calibration
FRTB040	Fraction of total dry matter to roots at DVS = 0.4	kg kg^{-1}	0.17	Field calibration
FRTB070	Fraction of total dry matter to roots at DVS = 0.7	kg kg^{-1}	0.07	Field calibration
FRTB090	Fraction of total dry matter to roots at DVS = 0.9	kg kg^{-1}	0.03	Field calibration
FLTB00	Fraction of total dry matter to leaves at DVS = 0	kg kg^{-1}	0.682	Field calibration
FLTB015	Fraction of total dry matter to leaves at DVS = 0.15	kg kg^{-1}	0.560	Field calibration
FLTB025	Fraction of total dry matter to leaves at DVS = 0.25	kg kg^{-1}	0.623	Field calibration
FLTB050	Fraction of total dry matter to leaves at DVS = 0.5	kg kg^{-1}	0.310	Field calibration
FLTB065	Fraction of total dry matter to leaves at DVS = 0.65	kg kg^{-1}	0.220	Field calibration
FLTB095	Fraction of total dry matter to leaves at DVS = 0.95	kg kg^{-1}	0	Field calibration
FLTB200	Fraction of total dry matter to leaves at DVS = 2.0	kg kg^{-1}	0	Field calibration
FOTB095	Fraction of total dry matter to storage organs at DVS = 0.95	kg kg^{-1}	0.73	Field calibration
FOTB200	Fraction of total dry matter to storage organs at DVS = 2.0	kg kg^{-1}	1	Field calibration
FSTB00	Fraction of total dry matter to stems at DVS = 0	kg kg^{-1}	0.318	Field calibration
FSTB015	Fraction of total dry matter to stems at DVS = 0.15	kg kg^{-1}	0.450	Field calibration
FSTB025	Fraction of total dry matter to stems at DVS = 0.25	kg kg^{-1}	0.377	Field calibration
FSTB050	Fraction of total dry matter to stems at DVS = 0.5	kg kg^{-1}	0.690	Field calibration
FSTB065	Fraction of total dry matter to stems at DVS = 0.65	kg kg^{-1}	0.780	Field calibration
FSTB095	Fraction of total dry matter to stems at DVS = 0.95	kg kg^{-1}	0.270	Field calibration
Death rate parameters				
PERDL	Maximum relative death rate of leaves due to water stress	kg kg^{-1} d^{-1}	0.03	Default value for wheat in WOFOST
RDRSTB15001	Relative death rate of stems at DVS = 1.50	kg kg^{-1} d^{-1}	0.02	Default value for wheat in WOFOST
RDRDTB200	Relative death rate of stems at DVS = 2.0	kg kg^{-1} d^{-1}	0.02	Default value for wheat in WOFOST
Rooting parameters				
RDI	Initial rooting depth	cm	10	Default value for wheat in WOFOST
RRI	Maximum daily increase in rooting depth	cm d^{-1}	1.2	Default value for wheat in WOFOST
RDCMR	Maximum rooting depth	cm	125	Default value for wheat in WOFOST
Soil and management parameters				
CRAIRC	Critical soil air content for aeration	cm^3 cm^{-3}	0.06	Default value in WOFOST
DD	Depth of drainage	cm	20	Default value in WOFOST

Table 1 (*Continued*)

Parameter	Description	Units	Value	Source
IDEM	Day of emergence	DOY (day of year)	290	Field measurements
K0	Hydraulic conductivity of saturated soil	cm d^{-1}	10	Default value in WOFOST
KSUB	Maximum percolation rate of water into subsoil	cm d^{-1}	10	Default value in WOFOST
SM0	Soil moisture content of saturated soil	cm^3 cm^{-3}	0.506	Field measurements
SMFCF	Soil moisture content at field capacity	cm^3 cm^{-3}	0.325	Field measurements
SMLIM	Maximum moisture content in topsoil	cm	0.065	Default value in WOFOST
SMW	Soil moisture content at wilting point	cm^3 cm^{-3}	0.072	Field measurements
SOPE	Maximum percolation rate in the root zone	cm d^{-1}	10	Default value in WOFOST
WAV	Initially available water in total root-exploitable soil	cm	20	Default value in WOFOST

[a] Reinitialized parameters that must be optimized using the 4DVar assimilation procedure.

parameters, including the initial crop total dry weight (TDWI) at the true emergence and the life span of leaves growing at 35 °C (SPAN), were also measured at the station.

To acquire the growth conditions for winter wheat at a regional scale during different growth stages, we selected 53 sample plots representing different winter wheat growing conditions throughout the study area (Fig. 1) and monitored the plots from March to June 2009. Each sample plot covered an area of 90 m × 90 m, with four subplots (30 m × 30 m) enclosed in each plot. The plots were relatively homogeneous. LAI was measured in 2009 for six 5 × 5 m areas uniformly distributed within each subplot using the LAI-2000 LAI meter (LI-COR, Lincoln, NE, USA) during the seven key phenological stages for winter wheat: the green-up stage (early March), jointing stage (late March), elongation stage (early April), booting stage (late April), heading stage (early May), anthesis stage (mid-May), and maturity (mid-June). LAI values from the four subplots were averaged to represent the unique LAI value at each phenological stage in each sample plot. The position at the center of each sample site was recorded with a global positioning system receiver with an accuracy of about ±10 m so that it could be georeferenced to the remote-sensing images. In addition, the photosynthetic parameters of two plants in each subplot were measured using the LI-6400 portable photosynthesis system (LI-COR, Lincoln, NE, USA). Finally, winter wheat yields in the 53 sample plots were manually measured after harvesting in mid-June of 2009.

2.3. WOFOST crop model

2.3.1. Description of WOFOST

Several crop growth models are available, including EPIC (Williams and Singh, 1995), WOFOST (Boogaard et al., 1998; Van Diepen et al., 1989), DSSAT (Jones et al., 2003), and AquaCrop (Steduto et al., 2009). We chose the WOFOST model for this study because the model uses a generic process description that is suitable for large-scale and regional simulations. We used version 7.1.4 of the WOFOST model (Van Diepen et al., 1989) to simulate winter wheat growth. WOFOST is a mechanistic process-based model that describes plant growth based on light interception and CO_2 assimilation as the growth-driving processes and that uses crop phenological development as the growth-controlling process. The three crop development stages (DVS) are expressed using dimensionless variables, with zero representing emergence, one representing anthesis, and two representing maturity. The model can be applied in two different modes: the potential mode, in which crop growth is purely driven by temperature and solar radiation and no growth-limiting factors are considered, and the water-limited mode, in which crop growth is limited by the availability of water, root characteristics, soil physical characteristics, rainfall, and evapotranspiration (ET) during the growing season. The difference in yield between the potential and water-limited modes can be interpreted as the effect of soil moisture stress. We used the potential mode and water stress was not accounted for in this study, because the basic meteorological conditions in 2009, and especially the

temperature and rainfall, met the cumulative temperature and water consumption requirements for wheat throughout the winter wheat growing season. LAI is one of the most important state variables in the WOFOST model, as it represents the ability of the crop to intercept solar radiation, which drives CO_2 assimilation and is a crucial indicator for potential grain yield. Currently, WOFOST does not account for other yield-limiting factors, such as nutrients, pests, weeds, and farm management (Boogaard et al., 1998; de Wit et al., 2012).

2.3.2. Calibration of WOFOST

The WOFOST model requires data on a range of weather, soil, crop, and management parameters for each cell in the grid to simulate the spatial distribution of crop yield (Boogaard et al., 1998). Before a crop model can be used for a given agro-environmental region, it must be calibrated and the performance of the calibration must be evaluated to ensure that the model can accurately simulate the entire crop growth process by accounting for the variability of various local environmental parameters and the characteristics of the crop. We calibrated WOFOST for 'Hengguan 35', the dominant wheat cultivar that is grown in Baoding and Hengshui districts.

In this study, we calibrated the WOFOST input parameters using data collected in the fields at the Gucheng Experiment Station. The WOFOST parameters were determined from four sources: field measurements, field calibrations, published values, and the default values for wheat in WOFOST (Table 1). Details of calibration of WOFOST were reported previously (Ma et al., 2013b). Field validation showed that the errors of the WOFOST-simulated emergence, anthesis, maturity, and dry matter in storage organs were +2 days, −1 day, −2 days, and +63.5 kg ha^{-1}, respectively. This calibration indicated that simulation with WOFOST agreed well with the observed phenological and yield data at the field scale.

To regionalize the WOFOST model, meteorological inputs must be spatialized based on data from observational stations at the single point scale. In our study, we used a common Kriging interpolation routine to estimate the values of weather variables for each 10 km × 10 km grid size, including daily maximum and minimum temperature, solar radiation, wind speed, and actual vapor pressure. For precipitation, we used a daily regional precipitation dataset with a 25 km × 25 km grid size obtained from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/).

The soil data for each site were collected from the Gucheng station and from six National Agrometeorological Observation Stations of China in the study area (Fig. 1), which provided data on soil moisture, bulk density, water content at field capacity and at the wilting point, soil texture, and pH. The soil parameters used for the WOFOST modeling included the field capacity, wilting point, and initial available soil data. Other soil parameters for the study area were derived from the 1:1000000 Chinese soil database (http://www.soil.csdb.cn). We used inverse distance-weighted interpolation to generate the gridded values of the soil parameters at a 10 km × 10 km scale.
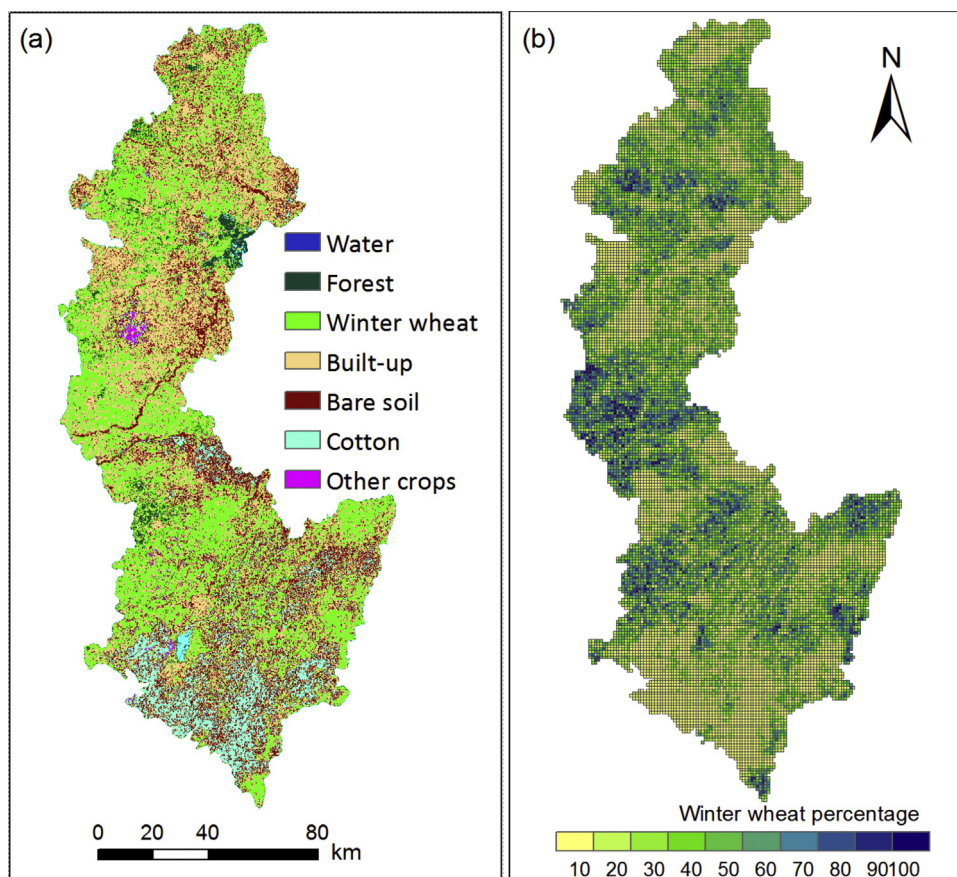
Fig. 2. (a) Map of the land-use and crop types in the study area. (b) Winter wheat pixel purity map (% of a pixel occupied by winter wheat) at a 1-km scale.

2.4. Crop type and pixel purity maps

We acquired six cloud-free Landsat TM scenes of the study area during the winter wheat growing season: on 14 March, 17 May, and 2 June 2009. These were close to the field-measurement dates during the green-up stage (5 March), anthesis stage (14 May), and maturity stage (10 June), respectively. In early March, winter wheat is growing fast during the green-up period, when other crops are not planted yet and natural shrub and forest vegetation has not yet begun to turn green. In early May winter wheat is in the middle to late part of its growing season and crop LAI reaches a maximum, as cotton and soybean are just beginning to turn green. In early June, winter wheat is becoming mature and LAI decreases to its minimum post-anthesis values, while the cotton and soybean canopies completely cover the surface of the fields.

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 version 5.0 of the ENVI software to obtain the reflectance in each band (RSI, 2001). The sample data from 152 ground parcels representing seven land-use or crop classes were collected during the field campaign in 2009; the classes were winter wheat, cotton, other crops, built-up, bare soil, forest, and water. A spatially distributed crop type map was obtained by means of supervised classification using the Mahalanobis distance algorithm in ENVI 5.0 based on the three TM

images (RSI, 2001); the overall accuracy of the classification was 90.3% and the kappa coefficient was 0.87. The crop type map was employed to mask all the pixels that were not classified as winter wheat fields (Fig. 2a).

A 1-km grid was overlaid on the 30-m land-use and crop type map to obtain a pixel purity map based on the percentage of winter wheat in each cell of the grid (Fig. 2b). There were a total of 16,335 pixels in the 1-km grid that covered the study area, and wheat occupied more than 50% of the cell in 2751 of these pixels (16.8% of the total).

2.5. Remotely sensed LAI datasets

2.5.1. TM LAI

Numerous studies have demonstrated that there is a strong link between spectral vegetation indices derived from remotely sensed observations and the field-measured LAI (Colombo et al., 2003; Peterson et al., 1987). Since spectral measurements of winter wheat are strongly influenced by the effects of the soil background in the reflectance signal during the green-up phenological phase, before the soil is covered by vegetation, a soil-adjusted vegetation index (SAVI) is most suitable for constructing the statistical relationship during this period (Huete, 1988). When the soil is fully covered by the winter wheat canopy in May and June, the normalized-difference vegetation index (NDVI) can be used to establish statistical regression relationships (Sellers, 1985). Two of the spectral bands from the TM images (band 3 [red] and band 4 [NIR]) within a 3×3 window size were used to establish the relationship between the field-measured LAI and SAVI in March when the canopy not yet closed or between LAI and NDVI in May and June
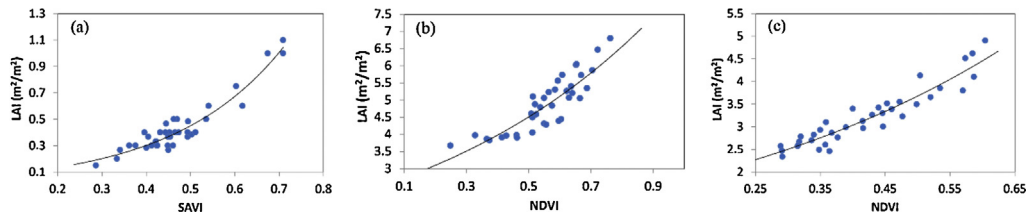
Fig. 3. Statistical regressions for the relationship between field-measured LAI and two TM vegetation indices (SAVI in March and NDVI in May and June) on (a) 14 March, (b) 17 May and (c) 2 June in 2009.

when the canopy was closed:

$$14 \text{March} 2009 : \text{LAI} = \frac{\ln[(1 - \text{SAVI}/1.2581)/0.9130]}{-0.8377},$$

$$R^2 = 0.849; p < 0.001 \tag{1}$$

$$17 \text{May} 2009 : \text{LAI} = \frac{\ln[(1 - \text{NDVI}/1.0866)/3.3790]}{-0.3994},$$

$$R^2 = 0.742; p < 0.001 \tag{2}$$

$$2 \text{June} 2009: \text{LAI} = \frac{\ln[(1 - \text{NDVI}/9.7639)/1.0081]}{-0.0155},$$

$$R^2 = 0.874; p < 0.001 \tag{3}$$

The parameter values in these equations were obtained from regressions of field-measured LAI (using 37 of the 53 plots) against SAVI and NDVI during the three time periods (Fig. 3).

The three statistical regression equations were applied to the TM data to obtain regional TM LAI maps. Then, these maps were

validated by comparing the LAI values measured in the remaining 16 sample plots with the corresponding values estimated from the LAI maps. The RMSE of the average LAI values was 0.52 m² m⁻² for the three dates. The average LAI values for the maps from 14 March, 17 May and 2 June were 0.35 m² m⁻², 5.1 m² m⁻², and 3.5 m² m⁻², respectively. The LAI maps (Fig. 4) show an LAI increase from mid-March to mid-May when heading occurs, and then a decrease to the beginning of June. The validation showed that the TM LAI maps at the three phenological stages accurately represented the LAI values during the given time periods and the temporal changes in LAI.

2.5.2. Savitzky–Golay (S–G) filtered MODIS LAI

MODIS offers the advantages of short revisit intervals and large geographical coverage, which facilitates efforts to capture crop growth signatures and their spatial and temporal variability. A series of MODIS land products have been provided free of charge and they are readily available from the Earth Observation System gateway (http://eospso.gsfc.nasa.gov/). We employed the 4-day MODIS LAI product (MCD15A3), with 1-km spatial resolution, to cover the majority of the growing season from January to June 2009, including 45 dates between day of year (DOY) 1 and DOY
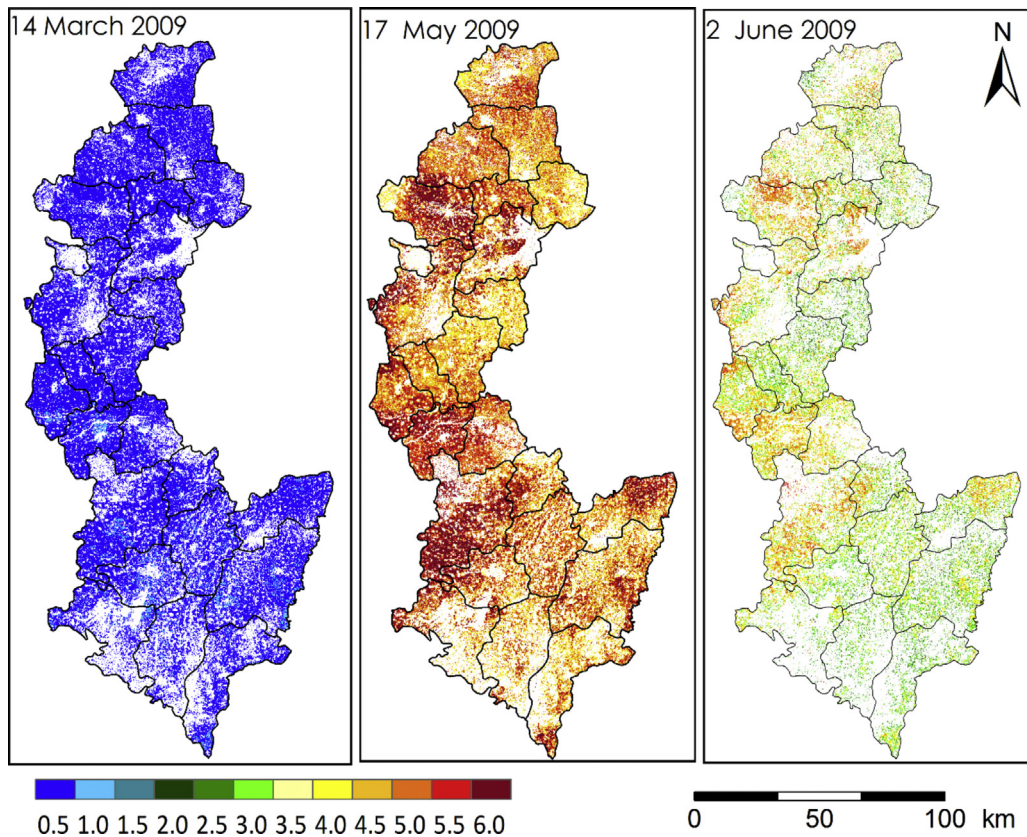

Fig. 4. Maps of winter wheat LAI in the study area retrieved from the Landsat TM data.
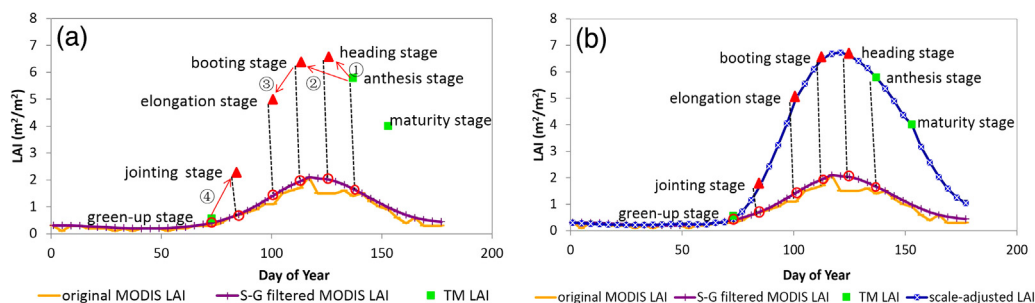
Fig. 5. (a) The procedures used to adjust the LAI values. Arrowheads represent the procedure to calculate adjusted LAI at different phenological stages using the ratio method; the circled numbers represent the sequence of four steps in the ratio calculation. The triangles represent the adjusted LAI calculated at different phenological stages. (b) The adjusted LAI profile simulated by the logistic function. Triangles and squares represent LAI values during the phenological stages used to fit the logistic function.

177. The h26v05 and h27v05 tiles, which cover the study area, were re-projected from the sinusoidal projection to the Albers conical equal-area projection using bilinear re-sampling in the MODIS reprojection tool.

Although the LAI dataset is a 4-day maximum-value-composited product, it still includes considerable noise caused by cloud contamination and atmospheric variability. Therefore, it needs to be processed for denoising before constructing realistic LAI time-series data to clearly reflect the growth status and temporal (phenological) variation of the crop. It was hypothesized that the local minimum of the LAI profile would represent noise due to cloud contamination, whereas the local maximum would represent the true value. On this basis, we applied an iterative S–G filtering algorithm to the MODIS LAI profile (Chen et al., 2004; Xu et al., 2011). The daily S–G filtered MODIS LAI profile captured the phenological characteristics of winter wheat reasonably well, with $R^2 = 0.62$, $p = 0.002$, and RMSE = 4 days, based on a comparison with phenological data measured during the seven phenological periods in the 53 sample plots. An exploratory analysis showed that most of the S–G filtered MODIS LAI values were too small ($<2.0\,\text{m}^2/\text{m}^2$) in the study area. A MODIS pixel with 1-km resolution in northern China is always a mixture of land-use and crop types rather than pure pixels that contain only winter wheat, which usually leads to low values in the MODIS LAI time series.

2.5.3. Scale-adjusted LAI

In our previous work (Ma et al., 2013a), we proposed a simple scaling method based on adjusting the S–G filtered MODIS LAI to the field-measured LAI on the basis of a logistic function (hereafter, the "old" algorithm). One drawback of the old algorithm relates to generating regional wheat LAI through spatial interpolation of ground-based LAI from sample plots. Since crop LAI does not meet the requirement of spatial continuity, this produces errors in the wheat LAI. In the present study, we improved the old algorithm by introducing medium-resolution data (i.e., TM data) from three phenological stages. The new scaling method is based on two assumptions: first, that the S–G filtered MODIS LAI correctly represents the crop phenological characteristics in pixels with high crop purity, and second, that the TM LAI is also relatively accurate. The new scaling method merges these two datasets to produce a scale-adjusted LAI.

The new scaling method consists of two steps. First, LAI maps with a 30-m spatial resolution are derived using the empirical relationship between the field-measured LAI and the TM-derived SAVI or NDVI; second, the TM LAI data from three dates were integrated with the S–G filtered MODIS LAI phenological information to generate a scale-adjusted LAI. We used a double-logistic regression function to simulate the adjusted LAI profile (Ma et al., 2013a; Xu et al., 2011). Logistic functions are monotonic, and therefore cannot simulate the LAI profile during all growth stages. Thus, different

logistic functions were used to simulate the pre-heading and post-heading LAI profiles, with the division chosen at the heading stage because this was the stage with maximum LAI. A logistic function requires at least four points to establish the equation. For the pre-heading LAI, the ratio (α) of the TM LAI ($LAI_{\text{TM-anthesis}}$) to the S–G MODIS LAI ($LAI_{\text{MODIS–SG-anthesis}}$) at the anthesis stage was multiplied by the S–G MODIS LAI ($LAI_{\text{MODIS–SG-heading}}$) at the heading stage to calculate the adjusted LAI at the heading stage, as shown in Eq. (4). α is a phenology-dependent variable. The same algorithm was applied to calculate the adjusted LAI at the booting stage (late April), elongation stage (mid-April), and jointing stage (late March). The four steps in the calculation are shown in Fig. 5a. The adjusted LAI values at these stages were used together with the TM LAI at the green-up stage to construct the logistic function. For the post-heading LAI, the adjusted LAI at the heading and booting stages was used together with TM LAI at the anthesis and maturity stages to establish the logistic function. Fig. 5a shows the calculation process.

$$\text{Adjusted LAI}_{\text{heading}} = \alpha \text{LAI}_{\text{MODIS–SG-heading}} \qquad (4)$$

Then, the logistic equation was fitted to the data:

$$y(t) = \left[\frac{c}{1 + e^{a+bt}}\right] + d \qquad (5)$$

where $y(t)$ is the adjusted LAI at time t (days), a and b are the fitting parameters, c is the maximum LAI, and d is the initial LAI. Fig. 5b shows the adjusted LAI profile simulated by the logistic function. It is important to note that scale adjustment was not conducted for the entire 1-km pixel, just for the area planted with winter wheat within the 1-km pixel. Variation in the fraction of winter wheat within a 1-km pixel greatly affects the characteristics of the MODIS LAI curve, and therefore influences the adjusted LAI curve. Thus, the fraction of wheat within a 1-km pixel is an important implicit parameter in the scaling model.

Comparison of satellite biophysical parameters with corresponding in situ measurements should be conducted with caution due to the scale mismatch between in situ measurements and the spatial resolution of the remote sensing data. In the study, all winter wheat TM LAI pixels that fell within the 1-km grid were averaged so that they could be scaled up to the 1-km scale. Furthermore, we investigated how the differences in the proportions of the different land cover and crop types in a 1-km pixel affected the scale-adjusted LAI. Fig. 6 compares the original MODIS LAI, the S–G filtered MODIS LAI, the TM LAI, and the scale-adjusted LAI with the field-measured LAI for pixels with different land use and crop type proportions. We found that the scale-adjusted LAI was systematically lower than the field-measured LAI when built-up land occupied 25% or more of the pixel (Fig. 6a and b). The LAI scale adjustment may not be successful when pixels contain a large proportion of built-up land and other crops because the original MODIS LAI profile cannot adequately represent the temporal variation in
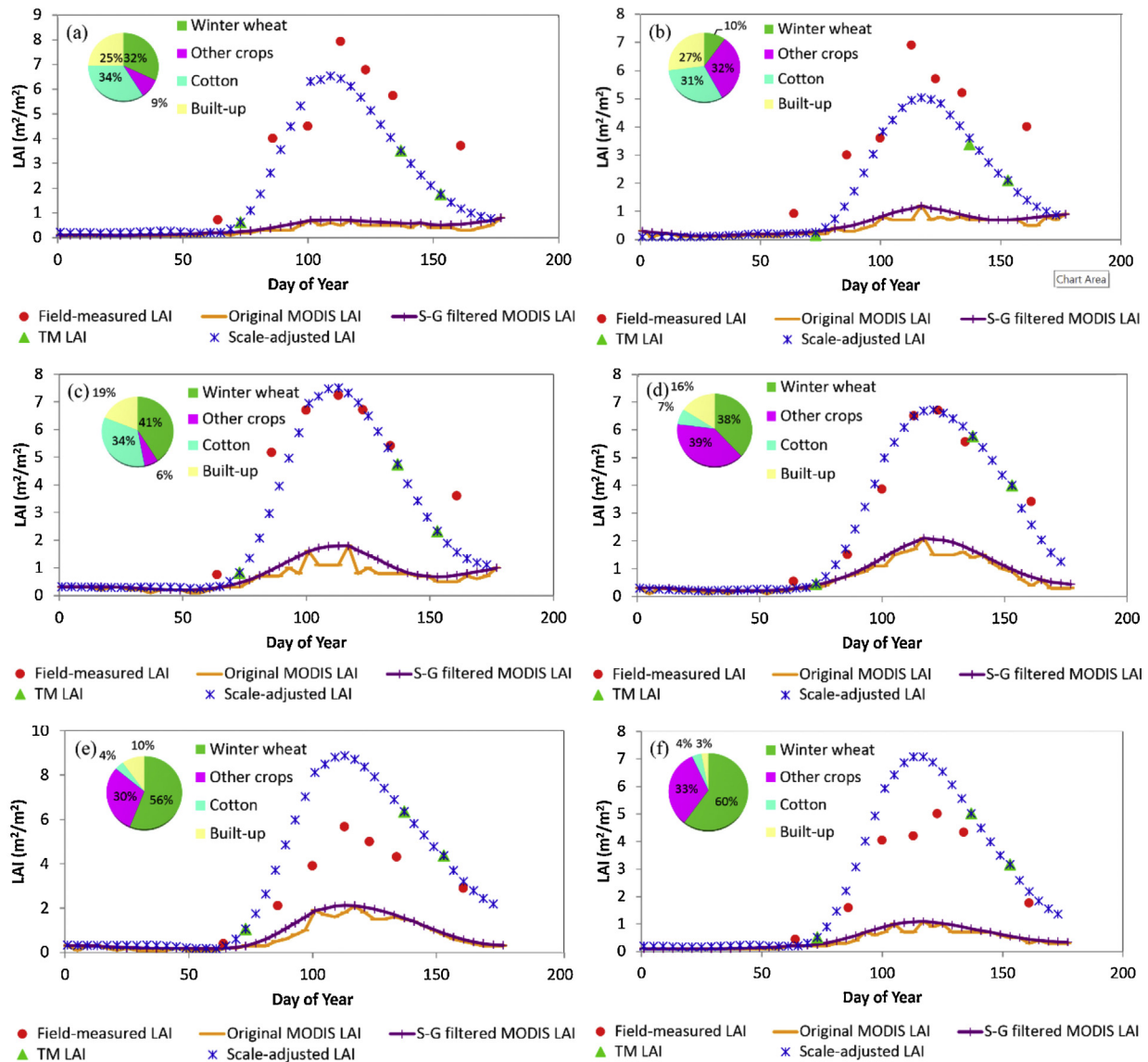
Fig. 6. Comparisons of the various LAI profiles for pixels with different land use and crop type proportions (defined in the pie charts) vs. DOY, the day of year.

the phenological characteristics of winter wheat under these conditions. A lower TM LAI can also lead to a lower scale-adjusted LAI, as shown in Fig. 6a and b. Conversely, a high TM LAI led to a high scale-adjusted LAI (Fig. 6e and f). The scale-adjusted LAI was close to the field-measured observations when the fraction of winter wheat was 38% or more and the built-up fraction was less than 20% (Fig. 6c and d). Based on this analysis, higher pixel purity improves the accuracy of the scale-adjusted LAI, and to ensure adequate purity, we chose a threshold of at least 50% pixel purity for winter wheat and a threshold of less than 25% built-up land in the rest of the analyses.

3. Assimilation method

3.1. Selection of reinitialized parameters for WOFOST

TDWI strongly influences the initial growth rate and represents an important uncertainty in the WOFOST model. Variability of TDWI greatly influences the rate of increase of the crop LAI and also affects the maximum LAI that can be reached during the growing season (de Wit et al., 2012). The initial day of emergence (IDEM) is also an important parameter that directly influences the biomass

of the storage organs, crop LAI, and phenology. IDEM and TDWI are strongly related because changes in one of the two factors can offset the effect of changes in the other parameter; for example, combining a later emergence date (IDEM) with a higher value for TDWI would lead to a similar grain yield. We focused on TDWI in this study, because TDWI reflects the actual biomass that generates subsequent growth.

The SPAN parameter represents the lifespan (in days) of leaves growing at 35 °C. Thus, SPAN determines the rate and timing of leaf senescence, and therefore, determines the time when LAI begins to decrease after heading. This parameter is also influenced by nutrients (e.g., a lack of nitrogen leads to early browning of leaves) as well as by pests and diseases, but these aspects in crop growth are not accounted for by WOFOST (Curnel et al., 2011). However, SPAN accounts to some extent for the effects of insects and diseases factors.

Considering the important roles of TDWI and SPAN, both parameters can alter both the trajectory of LAI during the growing season and the maximum LAI achieved by the crop. Thus, we simultaneously reinitialized TDWI and SPAN during the data assimilation. In the study, the default values for TDWI and SPAN were set as 210 kg ha^{-1} and 27 days, respectively. TDWI ranged between 50
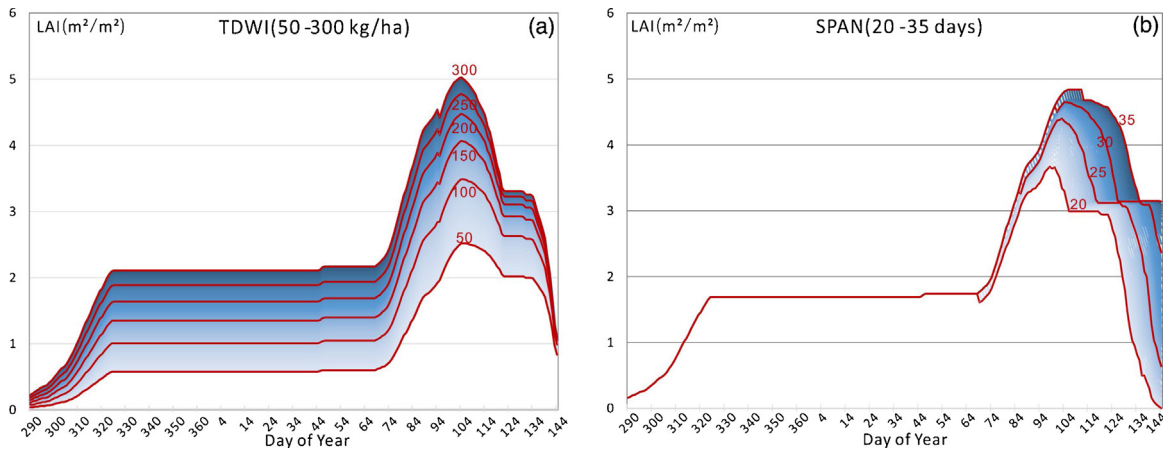
Fig. 7. Evolution of the simulated LAI profiles produced by the WOFOST model (a) with TDWI ranging between 50 and 300 kg ha^{-1} (251 iterations) and (b) with SPAN ranging from 20 to 35 days (160 iterations).

and 300 kg ha^{-1}, whereas SPAN ranged from 20 to 35 days. Fig. 7 shows the evolution of the simulated LAI profiles in response to changes in TDWI and SPAN within the defined ranges, respectively. TDWI mainly influenced the rate of increase of the crop LAI and the maximum LAI values, whereas SPAN only influenced the phase of decreasing LAI after heading.

3.2. 4DVar cost function method

The 4DVar method (Liang and Qin, 2008; Dente et al., 2008) was chosen to construct the cost function that we used to assimilate the remotely sensed LAI data into the WOFOST model for estimating the winter wheat yield at the field and regional scales. This cost function includes a term that measures the distance between the reinitialized parameter, x_k, and the background value, x_{k0}, at the beginning of the interval $(x_k - x_{k0})$, with a summation over time of the cost function at each observational increment for integration over the duration of the observations. The cost function in this

study, $J(x)$, was constructed as follows:

$$J(x) = \frac{1}{2}\sum_{k=1}^{2}(x_k - x_{k0})^T \mathbf{B}^{-1}(x_k - x_{k0})$$
$$+ \frac{1}{2}\sum_{i=1}^{N}(y_i - H_i(x))^T \mathbf{Q}_0^{-1}(y_i - H_i(x)) \qquad (6)$$

where k represents the number of reinitialized parameters; x_k represents the value of the WOFOST model input parameters (TDWI or SPAN); x_{k0} represents the prior information on these two parameters; $\mathbf{B}$ is the error covariance matrix for the two WOFOST model parameters; N represents the total number of LAI values derived from the remotely sensed data; y_i represents the LAI values derived from the remotely sensed data; $H_i(x)$ represents the LAI simulated by the WOFOST model; and $\mathbf{Q}_0$ represents the error covariance matrix for LAI derived from the remotely sensed data.

In the 4DVar cost function, the error covariance matrix for the WOFOST model ($\mathbf{B}$) and the observational error covariance matrix ($\mathbf{Q}_0$) are extremely important for determining the final assimilation accuracy. In this study, $\mathbf{B}$ mainly arises from uncertainties in the TDWI and SPAN values. $\mathbf{B}$ was developed using statistical
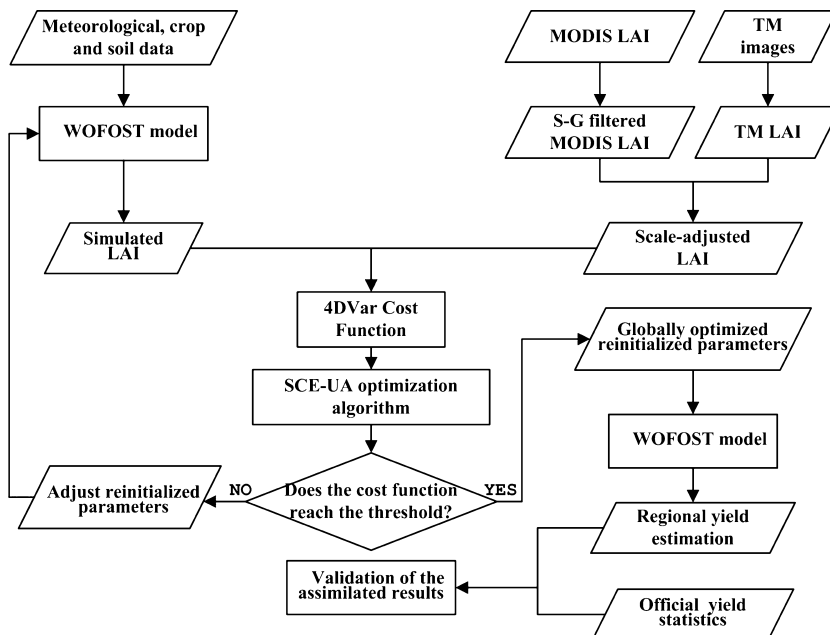

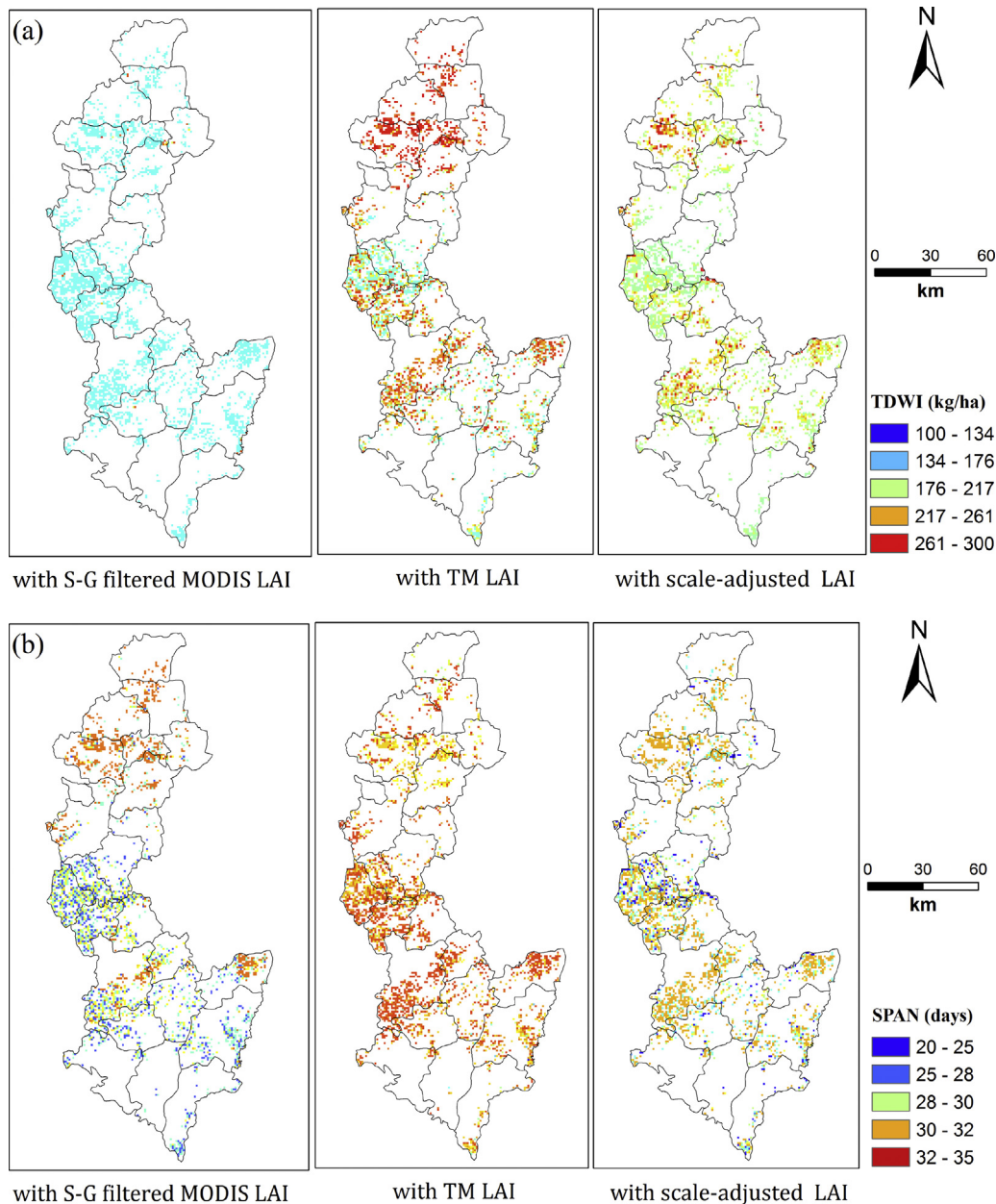**Fig. 8.** Flowchart of the 4DVar assimilation procedure.

Fig. 9. Spatial distributions of the (a) TDWI and (b) SPAN parameters after data assimilation using the 4Dvar cost function.

information from an empirical model and prior knowledge of the study area. The default values for TDWI and SPAN were set to $210 \, kg \, ha^{-1}$ and 27 days, respectively, as the starting point, with an error covariance of 7.8 and 0.7, respectively. $\mathbf{Q}_0$ was developed using the standard deviation of the field-measured LAI during the seven phenological stages from the 53 sample plots, which minimized the errors between the remotely sensed LAI and the WOFOST-modeled LAI values by providing less weight to LAI observations at different phenological stages when the uncertainty of these estimates was highest. The relative observational errors for the TM LAI during the three phenological stages equaled 0.42 at the green-up stage, 0.24 at the anthesis stage, and 0.33 at the maturity stage. The relative observational errors of the scale-adjusted LAI based on the 53 LAI sample plots were 0.06 at the green-up stage, 0.13 at the jointing stage, 0.26 at the elongation stage, 0.17 at the booting stage, 0.14 at the heading stage, 0.12 at the anthesis stage, and 0.12 at the maturity stage.

To find the optimal parameter sets, we used the SCE–UA optimization algorithm (Duan et al., 1994) to minimize the value of the cost function J by iterating the initial parameters of the WOFOST model. When the cost function reaches the given threshold value for its minimum value, the WOFOST-simulated yield using the optimized parameter set is the final assimilated yield. The SCE–UA algorithm repeats the comparison of two optimal parameter datasets until one of the following three predefined conditions is satisfied: (1) the cost function value does not improve by more than 0.0001% after five iterations; (2) the objective function has been calculated more than 10,000 times; or (3) the initial parameters of the model have converged to within a predetermined small range. We used two different remotely sensed LAI datasets at the regional scale: assimilation using the three dates for the TM LAI data, and using the time series for the scale-adjusted LAI. Fig. 8 shows the flowchart followed by the 4DVar assimilation procedure.
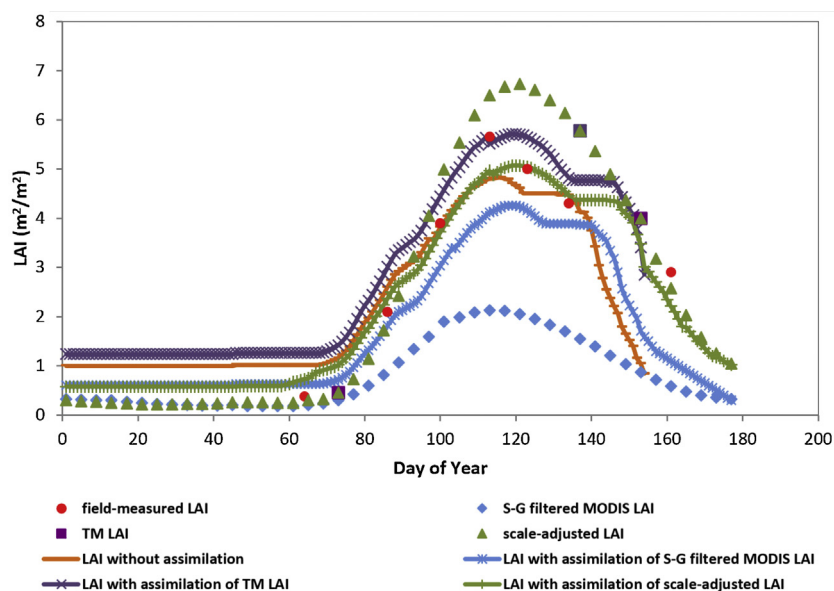
Fig. 10. Comparison of the WOFOST-simulated LAI with the field-measured LAI based on assimilation using different LAI datasets.

4. Results

4.1. Spatially distributed reinitialized parameters

The assimilation procedure described in Section 3 was applied to the 1-km grid cells that had at least 50% of the pixel occupied by winter wheat throughout the study area. For each cell, the WOFOST simulation starts at the true emergence date, which was set to 16 October 2008 based on the field observations. The WOFOST model was executed using the input parameters from the weather, soil, and crop management information corresponding to the cell of the grid for which the LAI observations were obtained. The optimization algorithm attempted to minimize the value of the cost function. The method generated a combination of TDWI and SPAN values for each cell in the grid. Fig. 9 shows the regional spatial distribution of the TDWI and SPAN values. Explaining and validating these spatial distributions are difficult due to a lack of corresponding regional-scale field-measured data. However, low TDWI values were obtained when assimilating the S–G filtered MODIS LAI, and this may have been caused by the low LAI values in the S–G filtered MODIS LAI during the green-up and heading stages. When assimilating the TM LAI and the scale-adjusted LAI, the reinitialized TDWI exhibited more realistic spatial variability (Fig. 9a). Assimilating TM LAI data from the three dates for which data was available achieved the best TDWI values, with an average value of 212.41 kg ha^{-1} and an error of only 1.2% compared with the field-measured mean value of 210 kg ha^{-1}. This can be explained by the accuracy of the TM LAI values during the green-up stage, which played the dominant role in retrieving accurate TDWI values. We obtained an average TDWI value of 196.25 kg ha^{-1}, with an error of 6.6%, after assimilating the scale-adjusted LAI.

For the SPAN parameter, assimilating the S–G filtered MODIS LAI values achieved the best accuracy, with an average value of 27.76 days and an error of 2.8% compared with the field-measured mean value of 27 days. It is possible that the temporal pattern of wheat phenological characteristics played a dominant role in optimizing the SPAN parameter. The spatial variability of SPAN is obvious after assimilating the S–G filtered MODIS LAI, TM LAI, and scale-adjusted LAI (Fig. 9b). However, the SPAN value of 32.43 days was higher when assimilating data from the three dates for the TM LAI due to the limited numbers of LAI values during the growing season. Relatively accurate SPAN values were obtained, with an average value of 29.59 days and an error of 9.6%, after assimilating the scale-adjusted LAI.

From the joint errors of TDWI and SPAN, assimilating the scale-adjusted LAI achieved better accuracy (a total error of 16.2%) than using either the S–G filtered MODIS LAI (52.8%) or the TM LAI (21.3%) alone. These results showed that assimilating the scale-adjusted LAI provided a more detailed image of the spatial distributions of TDWI and SPAN throughout the study area.

4.2. Comparison of the LAI trajectories after data assimilation

To validate the assimilation accuracy of LAI at the field scale, we compared the WOFOST-simulated LAI (the assimilated LAI) using the re-initialized TDWI and SPAN parameters with the field-measured LAI. Fig. 10 compares the WOFOST-simulated LAI with the field-measured LAI based on assimilation using the different LAI datasets. The assimilated LAI with the S–G filtered MODIS LAI was systematically below the field-measured LAI and the WOFOST-simulated LAI without assimilation as a result of the low values for the re-initialized TDWI. This was especially true during the

Table 2
Comparison of the estimated yield at the field scale using various LAI datasets with the field-measured yield.

	Mean (kg ha^{-1})	Maximum (kg ha^{-1})	Minimum (kg ha^{-1})	R^2	p	RMSE (kg ha^{-1})
Field-measured yield at the 53 sample plots	7326	8295	6435	–	–	–
Estimated yield without data assimilation	6386	6701	6253	0.69	0.012	1083
Estimated yield with 4DVar using the 53 LAI sample plots	6816	7485	6222	0.83	0.002	585
Estimated yield with 4DVar using the TM LAI from three dates	6722	7286	5732	0.70	0.004	692
Estimated yield with 4DVar using the scale-adjusted LAI	6815	7298	6223	0.77	0.003	654

Table 3
Accuracy of the estimated wheat yield using LAI from different phenological periods or combinations of these periods compared with the field-measured yield at the field scale.

|	Mean (kg ha^{-1})	Maximum (kg ha^{-1})	Minimum (kg ha^{-1})	R^2	p	RMSE (kg ha^{-1})
Field-measured yield at 53 sample plots	7326	8295	6435	–	–	–
Estimated yield with field-measured LAI at the 7 phenological stages	7531	8557	6944	0.77	0.004	373
Estimated yield with field-measured LAI at the green-up stage	7229	7731	6321	0.32	0.018	547
Estimated yield with field-measured LAI at the jointing stage	7227	8265	6655	0.36	0.019	533
Estimated yield with field-measured LAI at the elongation stage	7185	7949	6247	0.42	0.008	517
Estimated yield with field-measured LAI at the booting stage	7461	8071	6311	0.51	0.004	477
Estimated yield with field-measured LAI at the heading stage	7467	8021	6598	0.56	0.003	463
Estimated yield with field-measured LAI at the anthesis stage	7240	7842	6275	0.45	0.006	512
Estimated yield with field-measured LAI at the maturity stage	7145	7563	7038	0.28	0.017	607
Estimated yield with scale-adjusted LAI during the pre-heading stage (green-up, elongation, and heading stages)	7472	7980	7065	0.67	0.001	455
Estimated yield with scale-adjusted LAI during the post-heading stage (heading, anthesis, and maturity stages)	7368	7964	5950	0.47	0.005	510
Estimated yield with scale-adjusted LAI for all seven phenological stages	7503	8050	6863	0.72	0.003	410
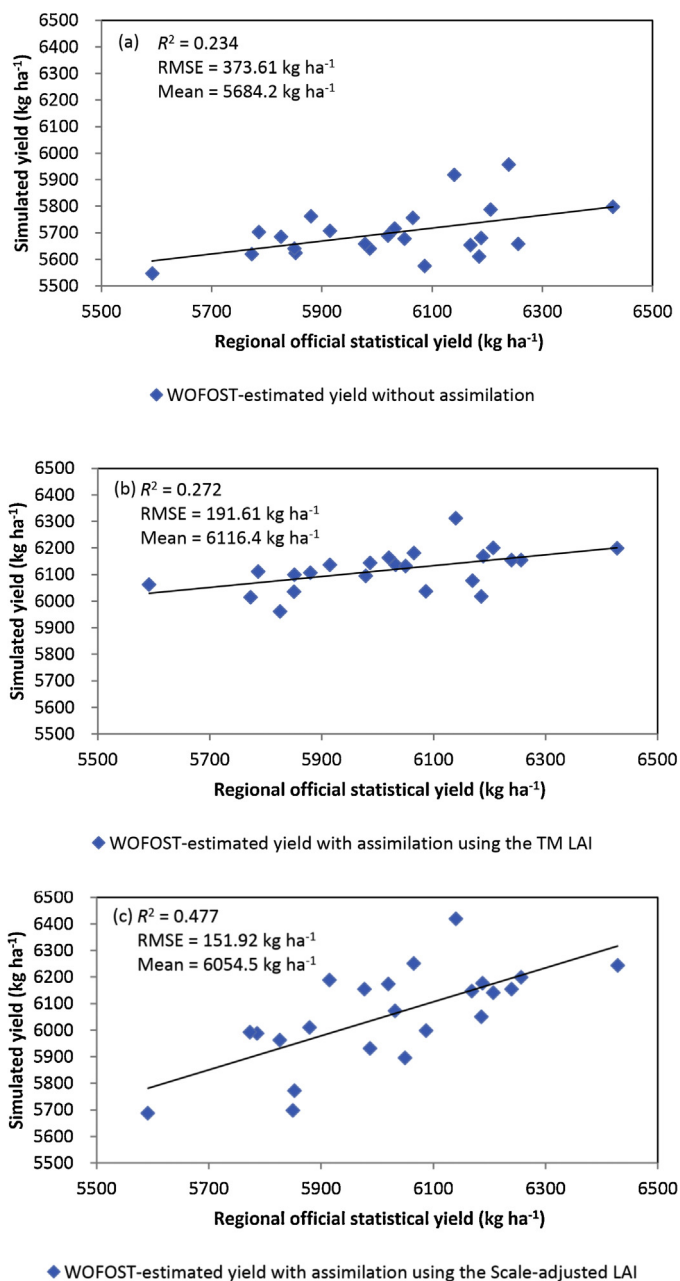
Fig. 11. Scatterplots of the simulated and official statistical yields.

post-heading stages, which were not consistent with the field-measured LAI, possibly due to the inclusion of green biomass from the summer crops (e.g., soybean and cotton) within the 1-km pixels rather than due to senescence of the winter wheat. Although the LAI assimilated with the TM LAI was relatively accurate (RMSE = 0.66 m^2 m^{-2}, R^2 = 0.97, p = 0.0043), the shape of the assimilated LAI curve was not consistent with the field-measured LAI due to the limited number of assimilation data. The validation results showed that the scale-adjusted LAI achieved the best accuracy compared with the field-measured LAI during the growing season, with RMSE = 0.34 m^2 m^{-2}, R^2 = 0.98, and p = 0.004.

4.3. Comparison of estimated wheat yield with the results based on data assimilation at the field scale

We assessed the accuracy of the estimated yield at the field scale based on validation using the field-measured yields from the 53 sample plots. Table 2 compares the estimated yield with the field-measured data for the three LAI datasets (i.e., the 53 LAI sample plots, TM LAI based on three dates, and the scale-adjusted LAI). Overall, the three LAI datasets underestimated the actual wheat yield. As expected, directly assimilating the field-measured LAI from the 53 sample plots achieved the best accuracy, with R^2 = 0.83 and RMSE = 585 kg ha^{-1}. The scale-adjusted LAI achieved the second-highest accuracy, with R^2 = 0.77 and RMSE = 654 kg ha^{-1}. Assimilating the TM LAI from the three dates improved the yield estimation compared with the WOFOST-simulated yield without data assimilation, with RMSE = 692 and 1083 kg ha^{-1}, respectively.

4.4. Importance of LAI in different phenological stages for wheat yield estimation at the field scale

To assess the impacts of LAI in different phenological stages on the wheat yield estimation at the field scale, we examined the results based on field-measured LAI during the seven phenological stages in the 53 sample plots using the LAI during various combinations of these stages (Table 3). Considering only individual stages, the LAI at heading played the most critical role in wheat yield estimation (i.e., it had the highest R^2 and lowest RMSE of all the individual phenological stages), followed by LAI during the booting, anthesis, elongation, jointing, green-up, and maturity stages. Heading was the period with the maximum LAI value, reflecting the canopy vigor of winter wheat, and LAI at this stage is closely related to the final winter wheat yields. The lower importance of the LAI values during the green-up and maturity stages is because these two stages do not adequately represent the temporal variability of the LAI trajectory. When combining the stages, assimilating

LAI during the pre-heading stage achieved better accuracy than assimilating LAI during the post-heading stage (Table 3). This may be because our selection of pixels with at least 50% wheat fraction 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 our study area). In contrast, the LAI trajectory of the post-heading period tends to be influenced by other simultaneously emerging summer crops (e.g., cotton, soybean). Overall, assimilating the field-measured LAI or scale-adjusted LAI from the seven phenological stages during the growing season achieved the best accuracy ($R^2 = 0.77$ and 0.72, respectively; RMSE = 373 and 410 kg ha^{-1}, respectively) compared with using LAI from any single phenological stage or from the pre-heading or post-heading stages. This demonstrated the importance of assimilating a time series of LAI to improve assimilation accuracy.

4.5. Assimilation of TM LAI at the regional scale

The 1-km grid cells that had at least 50% winter wheat were used in the 4DVar assimilation procedure, and the simulated wheat yield for each cell in the grid was aggregated at a county level so that the results could be validated using official regional yield statistics, which are compiled at a county level. All winter wheat TM LAI pixels that fell within a 1-km grid cell were averaged to allowing scaling up to the 1-km spatial resolution. Data from the three dates available for the 1-km TM LAI (at the jointing stage, the anthesis stage, and the maturity stage) were assimilated into the WOFOST model and the regional wheat yield was estimated. The validation based on the official regional statistical data indicated that the coefficient of determination (R^2) increased slightly, from 0.23 without assimilation (Fig. 11a) to 0.27 with assimilation of the TM LAI data (Fig. 11b), and RMSE decreased from 373.61 to 191.61 kg ha^{-1}. In total, 76% of the estimated yields fell within a realistic range, from 6116.4 to 6311.7 kg ha^{-1}. Two main reasons can explain this result. First, using LAI from only three dates

is insufficient to account for the requirements of the temporal variational assimilation. Second, LAI during other important phenological stages must be included in the assimilation procedure to improve the performance of data assimilation, as illustrated at the field scale. In particular, LAI at the heading stage, when LAI reaches its maximum, plays a dominant role in determining the final assimilated yield.

4.6. Assimilation of the scale-adjusted LAI at the regional scale

When the 45 dates for the scale-adjusted LAI from DOY 1 to DOY 177 were assimilated into the WOFOST model to estimate the regional wheat yield, the estimation accuracy improved greatly, with a much higher coefficient of determination ($R^2 = 0.48$) and a much lower RMSE (151.92 kg ha^{-1}) (Fig. 11c). Furthermore, the simulated wheat yield showed more realistic spatial variability throughout the study area. This can be explained by details of the spatial variations of TDWI and SPAN after data assimilation. Fig. 12 shows the wheat yield maps obtained for the study area based on analyses with and without assimilation. For the scale-adjusted LAI, 80% of the estimated yields ranged from 6054.5 to 6417.9 kg ha^{-1}. In particular, the northern part of the study area showed lower average values than the southern regions. In general, the yield variability resulted from differences in solar radiation, temperature, and farming management (irrigation and fertilization). The spatial pattern of the estimated wheat yields agreed well with official yield statistics at the county level (Fig. 12c and d).

5. Discussion

Crop growth models often oversimplify actual crop growth conditions. Some uncertainty is introduced by the model's architecture, and especially by uncertainties in the crop and meteorological parameters, resulting in a biased simulation of crop growth and yield. Uncertainty in crop input parameters can be a major source of uncertainty as well. In this study, we tried to decrease these
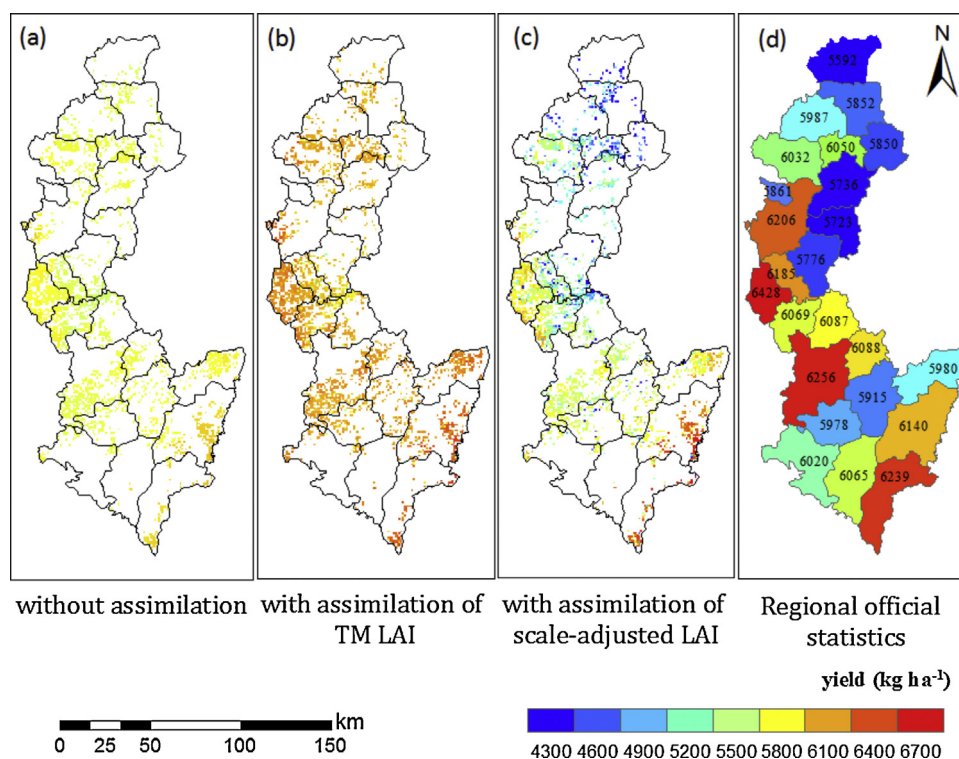

Fig. 12. Estimated winter wheat yield at a regional scale without assimilation and with assimilation using the TM LAI and the scale-adjusted LAI.

uncertainties by calibrating two key parameters (TDWI and SPAN) to account for regional crop characteristics. To reduce the uncertainties in the model parameters and structure, we constructed a cost function using the 4DVar strategy to calibrate and optimize the two parameters. Then, the cost function was optimized using the SCE–UA algorithm to minimize the difference between the observed and modeled LAI. The optimization results indicated that considerable spatial variability existed in TDWI and SPAN, and that simulating the LAI trajectory using the optimized TDWI and SPAN parameters improved the correlation between the model predictions and the field-measured LAI. Finally, we applied the scheme to estimate the regional-scale winter wheat yield.

Crop-type masking is an important aspect of developing a yield estimation approach based on data assimilation. Such masks enable isolation of the remotely sensed crop-specific signal throughout the growing season, thereby reducing the noise in the signal that is created by the presence of other land cover or crop types, and this decreases the intra-pixel heterogeneity during data assimilation operations. One of the difficulties in predicting crop yields using remotely sensed imagery is the availability of timely and annual crop masks for identification of a specific crop. Moreover, in order to predict yields during the growing season, a crop-specific mask is required prior to the end of the growing season, which can present a significant logistical challenge (Kastens et al., 2005); thus, this critical information is seldom available during the growing season. The rapid expansion of constellations of instruments with high spatial resolution could resolve this problem in the near future (de Wit et al., 2012).

In our analysis, we chose a dominant cultivar to calibrate the growth characteristics of winter wheat in our study area. However, farmers may choose different varieties in different regions of northern China, and these varieties may differ in their spectral characteristics, phenological stages, crop characteristics, and potential grain yield. Therefore, for regions that cultivate multiple varieties of wheat, obtaining data on the spatial distribution of the dominant wheat varieties and calibrating the genotype parameters for these varieties in the crop growth model would also improve the accuracy of crop yield estimation at a regional scale.

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 more heterogeneous surfaces, leading to greater scale errors than would occur with higher resolution data such as TM or ASTER data. Therefore, using low-resolution remote-sensing data in the data assimilation procedure introduces a scale disparity between the remotely sensed data and the field scale that must be accounted for. Due to the complexity and variability of the planting structure in agricultural landscapes (Zhao et al., 2010), performing the scale transformation between remotely sensed observations and crop models remains a challenging task. The scale effect and transformation models in data assimilation systems have been considered in several previous studies (e.g., de Wit et al., 2012; Montzka et al., 2012). Charoenhirunyingyos et al. (2011) established a regression relationship to convert between the MODIS LAI and field-measured LAI and to reduce the scale mismatch in the data assimilation process. Ma et al. (2013a) also adjusted the MODIS LAI to the field-measured LAI with a logistic function. However, these methods seem to be inappropriate for directly adjusting the MODIS LAI with respect to the field-measured LAI. In the present study, we proposed a nonlinear scale-discrepancy adjustment method by introducing intermediate-scale remotely sensed data to solve the problem of a scale mismatch and generate a more accurate LAI trajectory, which has the advantage of more general applicability in other agricultural regions. Validation of the results using field measurements showed that the scale-adjusted LAI was a promising way to represent actual crop growth and final yields. However, one challenge for this approach is that large amounts of high-quality ancillary data

are needed over a regional scale. Several previous studies demonstrated the great potential for using remotely sensed data with high spatial and temporal resolution in data assimilation for large-scale crop yield estimates (e.g., Claverie et al., 2012; Ma et al., 2013b). Future satellite missions, such as Venμs and Sentinel-2, will provide images with higher spatial resolution (20 m) and potentially higher temporal resolution (a revisiting period of 4–5 days). To increase the effective temporal resolution, combining optical data with synthetic aperture radar data would mitigate the effect of cloud cover in cloudy and rainy regions and increase the number of days with usable data. Accounting for the scale effect in heterogeneous pixels will be less of a problem with this improved spatial resolution, more accurate modeling of the LAI trajectory will be permitted throughout the growing season over large areas, and this will significantly improve regional crop yield estimation, particularly if the model is combined with improved data assimilation techniques.

In this study, we selected LAI as the unique state variable in the cost function used in the 4DVar assimilation. Although the temporal evolution of crop LAI is an important indicator of canopy light interception and carbon assimilation, LAI alone does not accurately represent the comprehensive impacts of solar radiation, air and soil temperatures, and soil water on crop yield. Remote sensing has been employed to estimate crop and soil characteristics such as LAI, ET, and soil moisture. Various algorithms have been developed to retrieve biophysical and biochemical variables from remotely sensed reflectance data (Dorigo et al., 2007). Other important variables such as ET and soil moisture derived from the remotely sensed data are also closely related to the crop yield and need to be included in the data assimilation framework. In addition, a more robust approach is needed to simultaneously assimilate multiple biophysical variables (e.g., LAI, ET, soil moisture), and hybrid approaches, such as combining the use of an ensemble Kalman filter with 4DVar would allow simultaneous estimates and updating of the model parameters and state variables to further improve regional crop yield prediction under water stress. Furthermore, an extension of the analysis to multiple years is needed to further validate the approach discussed in this paper and determine how well it accounts for the spatial and inter-annual variability in crop yield prediction.

6. Conclusions

In this study, we used the WOFOST process-based growth model to estimate winter wheat yield at a regional level, and enhanced the model's simulation accuracy by incorporating assimilated remotely sensed LAI data from Landsat TM and MODIS data using the 4DVar cost function combined with the SCE–UA optimization algorithm.

With assimilation of the TM LAI from three dates (i.e., a limited number of measurements but with more accurate LAI), the yield estimates improved further (higher R^2) and the RMSE decreased. The best simulation results (highest R^2 and lowest RMSE) were obtained with the scale-adjusted LAI values, demonstrating the importance of scale correction in the data assimilation procedure. Furthermore, the uncertainty related to the remotely sensed data and the time step between the assimilated LAI observations had different influences on the accuracy of the optimized parameters and on the final estimated crop yield. These results revealed that reducing the errors in the observations seems to be an effective way to improve the performance of the data assimilation. These results demonstrated that improving LAI retrieval accuracy during the crop's different phenological stages will be more effective than improving the temporal availability of the observations in the crop data assimilation system.

Our results showed that the current 1-km MODIS LAI products are not suitable for assimilation into the WOFOST crop growth

model because they tended to force the WOFOST model to reach unrealistically low crop LAI and yield values. Assimilating TM LAI data from three phenological stages had a limited ability to improve the model's performance. Heading LAI played the dominant role in improving assimilation accuracy compared with LAI during other phenological stages. In addition, using pre-heading LAI was more effective for improving the model's performance than using post-heading LAI. The nonlinear adjustment method we developed to account for the scale discrepancy between LAI simulated using the crop model and the MODIS LAI, combined with a pixel purity map, improved the ability to account for spatial heterogeneity in the 1-km wheat pixels. Our validation results showed that the scale adjustment approach generated an accurate LAI trajectory throughout the growing season and improved the agreement between the scale-adjusted LAI and the field-measured LAI. Assimilating the time series of scale-adjusted LAI greatly improved the estimates of wheat yield at both field and regional scales. These results indicated that the proposed wheat yield estimation method, based on the 4DVar strategy, is a promising way to estimate wheat yield at a regional scale, and this study provides an approach that can improve crop yield estimation in other agricultural regions of the world.

Acknowledgements

References(23)

Bastiaanssen, W., 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agric. Ecosyst. Environ. 94 (3), 321–340.

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.L., van Diepen, C.A., Rötter, R.P., Cabrera, J.M.C.A., van Laar, H.H., 1998. WOFOST 7.1; User's Guide for the WOFOST 7.1 Crop Growth Simulation Model and WOFOST Control Center 1.5. DLO Winand Staring Centre, Wageningen.

Charoenhirunyingyos, S., Honda, K., Kamthonkiat, D., Ines, A.V.M., 2011. Soil hydraulic parameters estimated from satellite information through data assimilation. Int. J. Remote Sens. 32 (23), 8033–8051.

Chen, J., Jonsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sens. Environ. 91 (3–4), 332–344.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.F., Huc, M., Keravec, P., Béziat, P., Fieuzal, R., Ceschia, E., Dedieu, G., 2012. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. Remote Sens. Environ. 124, 844–857.

Colombo, R., Bellingeri, D., Fasolini, D., Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sens. Environ. 86, 120–131.

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.

Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop yield assessment from remote sensing. Photogramm. Eng. Remote Sens. 69 (6), 665–674.

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.

Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158 (3–4), 265–284.

Duveiller, G., Baret, F., Defourny, P., 2011a. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. Remote Sens. Environ. 115 (10), 2686–2701.

Duveiller, G., Baret, F., Defourny, P., 2012. Remotely sensed green area index for winter wheat crop monitoring: 10-year assessment at regional scale over a fragmented landscape. Agric. For. Meteorol. 166, 156–168.

Duveiller, G., Baret, F., Defourny, P., 2013. Using thermal time and pixel purity for enhancing biophysical variable time series: an interproduct comparison. IEEE Trans. Geosci. Remote Sens. 51 (4), 2119–2127.

Duveiller, G., Weiss, M., Baret, F., Defourny, P., 2011b. Retrieving wheat green area index during the growing season from optical time series measurements based on neural network radiative transfer inversion. Remote Sens. Environ. 115 (3), 887–896.

Fang, H.L., Liang, S.L., Hoogenboom, G., 2011. Integration of MODIS LAI and vegetation index products with the CSM-CERES-Maize model for corn yield estimation. Int. J. Remote Sens. 32 (4), 1039–1065.

Fang, H.L., Liang, S.L., Hoogenboom, G., Teasdale, J., Cavigelli, M., 2008. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. Int. J. Remote Sens. 29 (10), 3011–3032.

Fang, H.L., Wei, S.S., Liang, S.L., 2012. Validation of MODIS and CYCLOPES LAI products using global field measurement data. Remote Sens. Environ. 119, 43–54.

FAO, 2012. http://en.wikipedia.org/wiki/International_wheat_production_statistics

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25 (3), 295–309.

Jégo, G., Pattey, E., Liu, J., 2012. Using leaf area index retrieved from optical imagery, in the STICS crop model for predicting yield and biomass of field crops. Field Crops Res. 131, 63–74.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J Agron. 18 (3–4), 235–265.

Kastens, J.H., Kastens, T.L., Kastens, D.L.A., Price, K.P., Martinko, E.E., Lee, R.Y., 2005. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. Remote Sens. Environ. 99 (3), 341–356.

Liang, S.L., 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.

Ma, G.N., Huang, J.X., Wu, W.B., Fan, J.L., Zou, J.Q., Wu, S.J., 2013a. Assimilation of MODIS-LAI into WOFOST model for forecasting regional winter wheat yield. Math. Comp. Modell. 58 (3–4), 634–643.

Ma, H.Y., Huang, J.X., Zhu, D.H., Liu, J.M., Zhang, C., Su, W., Fan, J.L., 2013b. Estimating regional winter wheat yield by assimilation of time series of HJ-1CCD into WOFOST–ACRM model. Math. Comp. Modell. 58 (3–4), 753–764.

Ma, Y.P., Wang, S.L., Zhang, L., How, Y.Y., Zhang, L.W., He, Y.B., Wang, F.T., 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.

Mo, X., Liu, S., Lin, Z., Xu, Y., Xiang, Y., McVicar, T.R., 2005. Prediction of crop yield, water consumption and water use efficiency with a SVAT-crop growth model using remotely sensed data on the North China Plain. Ecol. Model. 183 (2–3), 301–322.

Montzka, C., Pauwels, V.R.N., Hendricks Franssen, H.J., Han, X.J., Vereecken, H., 2012. Multivariate and multiscale data assimilation in terrestrial systems: a review. Sensors 12 (12), 16291–16333.

Moulin, S., Bondeau, A., Delécolle, R., 1998. Combining agricultural crop models and satellite observations: from field to regional scale. Int. J. Remote Sens. 19 (6), 1021–1036.

Peterson, D.L., Spanner, M.A., Running, S.W., Teuber, K.B., 1987. Relationship of thematic mapper simulator data to leaf area index of temperate coniferous forests. Remote Sens. Environ. 22 (3), 323–341.

RSI, 2001. ENVI User's Guide. September 2001 edition. Research Systems.

Sellers, P.J., 1985. Canopy reflectance: photosynthesis and transpiration. Int. J. Remote Sens. 6 (8), 1335–1372.

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron. J. 101 (3), 426–437.

Tian, L.Y., Li, Z.X., Huang, J.X., Wang, L.M., Su, W., Zhang, C., Liu, J.M., 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.

Van Diepen, C.A., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manage. 5 (1), 16–24.

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.

Williams, J.R., Singh, V., 1995. The EPIC model. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO, pp. 909–1000.

Xu, W.B., Jiang, H., Huang, J.X., 2011. Regional crop yield assessment by combination of a crop growth model and phenology information derived from MODIS. Sensor Lett. 9 (3), 981–989.

Zhao, F., Gu, X., Verhoef, W., Wang, Q., Yu, T., Liu, Q., Huang, H., Qin, W., Chen, L., Zhao, H., 2010. A spectral directional reflectance model of row crops. Remote Sens. Environ. 114 (2), 265–285.