Agricultural and Forest Meteorology 204 (2015) 106–121

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

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

ARTICLE INFO

Article history:
Received 13 February 2014
Received in revised form 24 January 2015
Accepted 2 February 2015
Available online 16 February 2015

Keywords:
Data assimilation
Leaf area index
Scale adjustment
Wheat yield estimation
WOFOST

ABSTRACT

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.

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

* Corresponding author at: College of Information and Electrical Engineering, China Agriculture University, No. 17 Qinghua East Road, Haidian District, Beijing 100083, China. Tel.: +86 10 6273 7628; fax: +86 10 6273 7855.
E-mail address: jxhuang@cau.edu.cn (J. Huang).

http://dx.doi.org/10.1016/j.agrformet.2015.02.001
0168-1923/

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