A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Regional corn yield prediction using Moderate Resolution Imaging Spectroradiometer (MODIS) data and crop growth model

BY

HO-YOUNG BAN

FEBRUARY, 2017

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Regional corn yield prediction using Moderate Resolution Imaging Spectroradiometer (MODIS) data and crop growth model

HO-YOUNG BAN

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

GENERAL ABSTRACT

Crop yield was commonly estimated by sample survey, and the method would require considerable costs and labor. However, remote sensing data would help reliable crop yield prediction with minimal costs, and also help to acquire and monitor timely the crop growth conditions. Two approaches employed for predicting crop growth and yield based on remote sensing. One approach is to use empirical model which represents the direct relationship between remote sensing data and observed yields, and another approach is to assimilate remote sensing data into crop growth model to improve corn yield prediction. In this study, a simple model for each approach was developed to predict regional corn yield using a minimum dataset and examined for the feasibility of regional corn yield.

A simple model was developed to predict corn yields using the MoDerate Resolution Imaging Spectroradiometer (MODIS) data product from two geographically separate major corn crop production regions: Illinois, USA and Heilongjiang Province, China. Corn yields and phenology data were collected by agricultural district (AD) in Illinois from 2000 to 2013. Corn yields were also compiled by county in Heilongjiang Province from 2002 to 2012. Data from the three years were selected to validate the model by state, and 70 and 30% of the data from the other years were used to calibrate and validate the model by district, respectively. The MOD09A1 data product, which are 8-day interval surface reflectance data, were obtained from day of the year (DOY) 89 to 337 to calculate the leaf area index (LAI). The sum of the LAI from early in the season to a given date in the season [end of DOY (EOD)] was well fitted to a logistic function and represented seasonal change of leaf area duration (LAD), which is the integral of LAI over a specific season. A simple phenology model was derived to estimate the dates of emergence and maturity using the logistic function parameters b_1 and b_2, which represented the rate of increase in LAI and the date of maximum LAI at a given site, respectively. The phenology model predicted emergence and maturity dates fairly well, with root mean square error (RMSE) values of 6.3 and 4.9 days for the validation dataset, respectively. Two simple linear regression models (Y_P and Y_F) were established using LAD as the variable to predict corn yield; the phenology

model (Y_P) model used LAD from emergence to maturity, and the yield model (Y_F) model used LAD for a predetermined period from DOY 89 to a particular EOD. When state/province corn yields for the validation dataset were predicted at DOY 321, near completion of the corn harvest, the Y_P model performed much better than the Y_F model, with RMSE values of 0.68 and 0.66 t/ha for Illinois and Heilongjiang, respectively. The Y_P model showed a similar or better performance, even for the much earlier yield prediction at DOY 257. In addition, the model performance showed no difference between the two study regions with very different climates and cultivation methods, including cultivar and irrigation management.

Crop growth models and remote sensing are useful tools for predicting crop growth and yield, but each tool has inherent drawbacks when predicting crop growth and yield at a regional scale. To improve the accuracy and precision of regional corn yield predictions, a simple approach for assimilating MODIS product into a crop growth model was developed, and regional yield prediction performance was evaluated in a major corn-producing region in Illinois, USA. Corn yields and phenology data were collected at state and AD levels from 2000 to 2013. Corn growth and yield were simulated using the Crop Environment Resource Synthesis (CERES)-Maize model with a minimum input dataset comprising planting date, fertilizer amount, genetic coefficients, soil, and weather data. Planting date for each grid was estimated

using a phenology model with a LAD logistic function that describes the seasonal evolution of MODIS-derived LAD. Genetic coefficients of the maize cultivar for each grid were determined to be the genetic coefficients of the mature group [included in Decision Support System for Agrotechnology Transfer (DSSAT) 4.6], which shows the minimum difference between the maximum LAI value derived from the LAD logistic function and that simulated by the CERES-Maize model. In addition, the daily water stress factors employed in CERES-Maize model were estimated from the ratio of daily leaf area/weight growth rate estimated from the LAD logistic function to the daily leaf area/weight growth rate estimated by simulating CERES-Maize model under an auto-irrigation condition. Corn yield predictions using only the estimated planting date and maturity group were very poor under rain-fed conditions at both the AD and state levels, whereas corn yield predictions improved under the auto-irrigation condition, indicating that irrigation has been applied in a considerable portion of cornfields in Illinois. In addition to assimilation of the estimated planting date and maturity group, further assimilation of the estimated daily LAI and water stress factors also improved the corn yield prediction considerably, increasing the R^2 value from 0.72 to 0.78 and decreasing the RMSE from 1.47 to 0.75 t/ha for the yearly corn yield prediction. In addition, an earlier corn yield prediction at DOY 257 was possible without decreased accuracy.

In conclusion, simple corn yield prediction model for each approach was developed using remote sensing data, and had considerable accuracy and precision for predicting the corn yield in study regions. However, these models and method need to be examined for spatial portability in more diverse agro-climatic and agro- technology regions.

Keywords: MODIS; corn yield; phenology; LAD; logistic function; crop growth model; water stress

Regional corn yield prediction using Moderate Resolution Imaging Spectroradiometer (MODIS) data and crop growth model

CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTON

Among major crops, corn ranks top in the world production (FAOSTAT, 2015) and the domestic consumption as food and livestock feed in South Korea. Corn which is a cereal crop grown over the world, is produced for use in food (e.g., organic cereals, tortillas, corn chips, snack foods, and cornmeal), animal feed, and industrial products such as ethanol (Ranum et al., 2014). Corn is representative C_4 plant, and has higher water use and photosynthetic efficiency, and better adaptation in harsh environmental conditions (e.g., high temperature and dry) than C_3 plants (Ashraf and Harris 2013).

Global climate change caused by emissions of greenhouse gases is accelerating, and more frequently generating extreme weather events (e.g., heavy rainfall, high temperature, and droughts) (Rosenzweig et al., 2001; Trenberth, 2008). These events can negatively affect crop production for food, feed, or fodder (Kumar and Tuti, 2016; Sinha et al., 1988), and will change pattern and balance of trade for food and food products (Wheeler and Braun, 2013). If adaptation according to climate change over the years is not, the impacts will become worse (Gbegbelegbe et al., 2014). Particularly, South Korea which has very low self-sufficiency and depend on imports, will be more affected on food security than the other country. Thus, predicting crop production and yield of major producing countries in advance is essential for

these countries. Early crop production and yield prediction prior to harvest can help to play important role in decision making for food pricing and trading policies (Hayes and Decker, 1996; Kouadio et al., 2012).

Crop yield was commonly estimated by sample survey, and the method would require considerable costs and labor (Guannan et al., 2013). However, remote sensing products would help reliable prediction of crop production and yield with minimal costs because large datasets are open to public (Mumby et al., 1999), and also help to acquire and monitor timely the crop growth conditions (Cheng et al, 2016). Two approaches are employed for predicting crop growth and yield based on remote sensing (Shao et al., 2015).

The first approach is to use empirical model which represents the direct relationship between remote sensing data and observed yields at selected region (Morel et al., 2014). Various vegetation indices (VIs) [e.g., normalized difference vegetation index (NDVI), wide dynamic range vegetation index (WDRVI), and enhanced vegetation index (EVI)] have been developed and evaluated with surface reflectance remote sensing product (Clevers, 1988; Huete et al., 2002; Nguy-Robertson el al., 2012; Rouse et al., 1973), and the VIs have a strong correlation with crop growth and yield. Crop growth and yield have been predicted with VI data using simple empirical model (Dadhwal, 2003; Dash and Curran, 2007; Jaafar and Ahmad, 2015; Kolotii et al., 2015; Lopresti et al., 2015; Rembold et al., 2013), and sophisticated

empirical model requiring data such as rainfall and temperature have also been developed to improve accuracy for predicting crop growth and yield (Prasad et al., 2006). However, these methods are likely that the spatial portability of the model would be limited. In addition to VIs, remote sensing derived LAI has also been used to predict crop biomass and yield (Casanova et al., 1997; Maki and Homma, 2014; Son et al., 2013). LAI, the one side total leaf area per unit of ground area (Wasseige et al., 2003), is a key biophysical variable to determine crop growth and yield (Bach, 1998; Noureldin et al., 2013). LAD is the integrated value of LAI over time, and important crop parameter that has a strong correlation with dry matter production and grain yield (Liu et al., 2005; López-Bellido et al., 2008).

The second approach is to assimilate remote sensing products into crop growth models (Ma et al., 2013). Crop growth models can simulate daily crop state variables (e.g., LAI and biomass), nitrogen, carbon, and water cycles in response to cultivar characteristics, environment (e.g., solar radiation, temperature, and precipitation), and management practices (Huang et al, 2013; Yang et al., 2004). Crop growth models require many input data (e.g., management practices, cultivar parameters, soil properties, and weather data) to predict crop growth and yield (Basso et al., 2013; Machwitz et al., 2014), and the these constraints make it difficult to predict crop growth and yield in a region where cultivation information is not easily obtained (Jégo et al., 2015;

Motha et al., 2011). However, integration of remote sensing data and crop growth models made it possible to predict crop growth and yield at regional scale (Doraiswamy et al., 2003; Wu et al., 2011), and two strategies are mainly used for integrating remote sensing data into crop growth model (Moulin et al., 1998). The first strategy is "forcing" that updates state variables derived from remote sensing data into crop growth model (Dadhwal, 2003). The second strategy is "recalibration" that adjusts initial conditions and parameters of crop growth model using remote sensing data (Yuping et al., 2007). Ensemble Kalman Filter (EnKF) which is a representative "recalibration" method, has been widely used to predict crop yield by assimilating remote sensing data into crop growth model (Ines et al., 2013; Li et al., 2014; Machwitz et al., 2014; Wu et al., 2012; Zhao et al., 2013; Zhu et al., 2013). These methods use repetitive process which adjusts initial conditions (e.g., physical attributes of soil profile) and parameters of crop growth model (e.g., cultivar characteristics) by minimizing the difference between remote sensing-derived value and simulated value by crop growth model (Huang et al., 2015; Ines et al. 2013; Jiang et al., 2014). Therefore, these methods require high computational cost to predict crop yield at large scale (Biniaz Delijani et al., 2014; Lei et al., 2012) because of the repetitive process to find optimum value, and it would be spatially limited due to the localization in EnKF using calibration dataset (Anderson, 2012).

This dissertation focuses on the following objectives. The first objectives were to develop a simple approach to predict crop phenology and yield using a minimum set of remote sensing data products and to examine the spatial portability of the simple method. Reliable corn yield prediction could help assess the socioeconomic impact on food production at regional and global scales. The first study focused on predicting corn yield in major production areas of the USA and China. The second objectives were to develop a new approach to predict corn yields at large area by assimilating remote sensing data into crop growth model without re-initialize and re-parameterize processes, and to evaluate possibility of the approach by applying in a major production area of the USA.

LITERATURE REVIEW

Traditional crop growth and yield prediction have relied on sampling of a number of fields (Guannan et al., 2013; Mosleh et al., 2015), and this method requires considerable labor, costs and time. However, usage of remote sensing data can be help a reliable and timely crop growth and yield prediction without sizable costs (Sharifi, 2000). Two approaches have been used for predicting crop growth and yield based on remote sensing (Shao et al., 2015).

1. Empirical model using remote sensing data

Various vegetation indices (VIs) [e.g., normalized difference vegetation index (NDVI), wide dynamic range vegetation index (WDRVI), and enhanced vegetation index (EVI)] have been developed and evaluated with surface reflectance remote sensing product (Clevers, 1988; Huete et al., 2002; Nguy-Robertson el al., 2012; Rouse et al., 1973).

Predicting crop yield using remote sensing data products often depends on an empirical approach that relates VIs alone or in combination with remote sensing derived meteorological variables to the reported crop yields (Bolton and Friedl, 2013; Li et al., 2007; Johnson, 2014; Shao et al., 2015). Bolton and Friedl (2013) used a simple linear regression model to predict corn and

soybean yield using VIs (i.e., NDVI, EVI2, and NDWI) from MODIS in the Central USA and concluded that the EVI2 index exhibited better ability to predict maize yield than the NDVI and that the use of crop phenology information from MODIS improved the model predictability. Johnson (2014) developed a regression tree model using the linear and/or exponential relationship of MODIS-derived NDVI and daytime land surface temperature with county-level yield statistics to predict corn and soybean yield for 12 states in central and northern USA. Shao et al. (2015) developed simple linear regression model using multi-temporal NDVI from MODIS to predict county-level corn yields for the entire Midwestern USA and confirmed that using a digital elevation model climate data as additional model inputs slightly improved the performance of the regional corn yield model. Although simple models to predict crop yield can be developed (Huang et al., 2013; Huang et al., 2011), it is likely that the spatial portability of the model would be limited because the parameters of the empirical equation that were estimated in the study region would not be applicable to the other regions with different agroclimate and agrotechnologies. Additionally, VIs such as NDVI can detect inter-annual fluctuation of crop yield due to weather conditions while they cannot detect human-induced factors that result in increased crop yield (Huang et al., 2013), making the NDVI-crop yield regression model difficult to extend to other regions (Huang et al., 2013; Mkhabela and Mashinini, 2005).

Sophisticated approaches requiring data in addition to remote sensing data have also been developed to improve predictions of crop growth and yield (Huang et al., 2013; Lobell, 2013; Prasad et al., 2006). For instance, Prasad et al. (2006) used a piecewise linear regression model with a break point to predict corn and soybean yield using monthly NDVI, soil moisture, surface temperature, and total rainfall. Nevertheless, it is preferable to develop a crop yield prediction model with wide spatial portability using a small dataset (Zhang et al., 2012).

In addition to VIs, remote sensing derived LAI has also been used to predict crop biomass and yield (Casanova et al., 1997; Maki and Homma, 2014; Son et al., 2013). LAI, the one side total leaf area per unit of ground area (Wasseige et al., 2003), is a key biophysical variable to determine crop growth and yield (Bach, 1998; Noureldin et al., 2013). LAD is the integrated value of LAI over time. Although LAD is an important crop parameter that has a strong correlation with dry matter production and grain yield (Liu et al., 2005; López-Bellido et al., 2008), it has not been used in a yield prediction model using remote sensing data. LAD has been reported to have positive correlation with corn yield under water and nitrogen stress conditions imposed at different growth stages (Wolfe et al., 1988) and under varying planting densities of three maize hybrids (Alias et al., 2011), and genetic differences in photosynthetic duration (longer LAD) were reported to be associated with a

longer grain filling duration and higher yield (Russell, 1991). These findings suggest that LAD would have greater potential to represent corn yield variability in regions with diverse agroclimate and agrotechnologies compared to VIs and LAI at a particular crop growth stage.

2. Assimilating remote sensing data into crop growth model

Crop growth models can simulate daily crop state variables (e.g., LAI and biomass), nitrogen, carbon, and water cycles in response to cultivar characteristics, environment (e.g., solar radiation, temperature, and precipitation), and management practices (Huang et al, 2013; Yang et al., 2004). Crop growth models require many input data (e.g., management practices, cultivar parameters, soil properties, and weather data) to predict crop growth and yield (Basso et al., 2013; Machwitz et al., 2014), and these constraints make it difficult to predict crop growth and yield in a region where cultivation information is not easily obtained (Jégo et al., 2015; Motha et al., 2011). However, integration of remote sensing data and crop growth models is possible to predict crop growth and yield at regional scale (Doraiswamy et al., 2003; Wu et al., 2011), and two strategies are mainly used for integrating remote sensing data into crop growth model (Moulin et al., 1998).

The first strategy is “forcing” that updates state variables derived from

remote sensing data into crop growth model (Dadhwal, 2003). The state variables derived from remote sensing data were interpolated to convert daily time series due to temporal characteristic of remote sensing data and atmospheric effects (Delecolle and Guerif, 1998). Delecolle and Guerif (1988) estimated wheat yield by updating interpolated LAI derived from SOPT/HRV into AFRCWHEAT model. Bouman (1995) estimated biomass of winter wheat at harvest by updating LAI derived from radar remote sensing into SUCROS model. Although forcing strategy is simple, the initial conditions or parameters of crop growth model should be estimated to improve prediction performance (Moulin et al., 1998).

The second strategy is “recalibration” that adjusts initial conditions and parameters of crop growth model using remote sensing data (Yuping et al., 2007). Ensemble Kalman Filter (EnKF) which is a representative “recalibration” method, has been widely used to predict crop yield by assimilating remote sensing data into crop growth model (Ines et al., 2013; Li et al., 2014; Machwitz et al., 2014; Wu et al., 2012; Zhao et al., 2013; Zhu et al., 2013). For example, Li et al. (2014) assimilated LAI retrieved from ETM+ data into hydrology-crop growth model which links World food studies (WOFOST) model to better predict corn yields in study region which located in the middle reaches of the Heihe River basin, northwest China, and parameters related to maintenance respiration, rooting depth, and soil

hydraulic properties were adjusted using EnKF. Wu et al. (2011) used EnKF to assimilate MODIS-LAI into World Food Studies (WOFOST) model to estimate winter wheat yield in Hengshui district, Hebei Province, China. Ines et al. (2013) used EnKF to assimilate soil moisture and/or MODIS-LAI into CERES-Maize model to estimate corn yield from 2003 to 2009 in Story County, Iowa, USA. These methods use repetitive process which adjusts initial conditions (e.g., physical attributes of soil profile) and parameters of crop growth model (e.g., cultivar characteristics) by minimizing the difference between remote sensing-derived value and simulated value by crop growth model (Huang et al., 2015; Ines et al. 2013; Jiang et al., 2014). Therefore, these methods require high computational cost to predict crop yield at large scale (Biniaz Delijani et al., 2014; Lei et al., 2012) because of the repetitive process to find optimum value, and it would be spatially limited due to the localization in EnKF using calibration dataset (Anderson, 2012).

REFERENCES

Alias, M.A., Bukhsh, H.A., Ahmad, R., Malik, A.U., Hussain, S., Ishaque, M. (2011). Profitability of three maize hybrids as influenced by varying plant density and potassium application. J. Anim. Plant Sci. 21, 42-47.

Anderson, J.L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. Mon. Weather Rev. 140(7), 2359-2371.

Ashraf, M. and Harris, P.J.C. (2013). Photosynthesis under stressful environments: an overview. Photosynthetica, 51(2), 163-190.

Bach, H. (1998). Yield estimation of corn based on multitemporal LANDSAT-TM data as input for agrometeorological model. Pure and Applied Optics. 7, 809–825.

Basso, B., Cammarano, D., Carfagna, E.s (2013). Review of crop yield forecasting methods and early warning systems. In Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics, FAO Headquarters, Rome, Italy (July, pp. 18-19).

Biniaz Delijani, E., Pishvaie, M.R., Bozorgmehry Boozarjomehry, R. (2014). Distance Dependent Localization Approach in Oil Reservoir History Matching: A Comparative Study. Iran. J. Chem. Chem. Eng. 33(1), 75-91.

Bolton, D.K. and Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agr. Forest Meteorol. 173, 74-84.

Bouman, B.A.M. (1995). Crop modelling and remote sensing for yield prediction. Neth. J. Agr. Sci. 43, 143-143.

Casanova, D., Epema, G.F., Goudriaan, J. (1997). Monitoring rice reflectance at field level for estimating biomass and LAI. Field Crops Res. 55, 83-92.

Cheng, T., Yang, Z., Inoue, Y., Zhu, Y., Cao, W. (2016). Preface: Recent Advances in Remote Sensing for Crop Growth Monitoring. Remote Sens. 8(2), 116.

Clevers, J.G.P.W. (1988). The derivation of a simplified reflectance model for the estimation of leaf area index. Remote Sens. Environ. 25(1), 53-69.

Dadhwal, V.K. (2003). Crop growth and productivity monitoring and simulation using remote sensing and GIS. Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, 263-289.

Dash, J., Curran, P.J. (2007). Relationship between the MERIS vegetation indices and crop yield for the state of South Dakota, USA. In Proc. Envisat Symposium.

Delecolle, R., Maas, S.J., Guerif., M., Baret, F. 1992. Remote sensing and crop production models: present trends. ISPRS J. Photogramm. Remote Sens. 47:145-161.

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

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J. (2005). Application of MODIS derived parameters for regional crop yield assessment. Remote Sens. Environ. 97(2), 192-202.

Faostat.fao.org. (2015). FAOSTAT. [online] Available at: http://faostat3.fao.org/home/E [Accessed 27 Dec. 2016].

Gbegbelegbe S., Chung U., Shiferaw B., Msangi S., Tesfaye K. (2014). Quantifying the impact of weather extremes on global food security: A spatial bio-economic approach. Weather Clim. Extrem. 4, 96–108.

Guannan, M., Jianxi, H., Wenbin, W., Jinlong, F., Jinqiu, Z., Sijie, W. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. Math Comput. Model. 58(3), 634-643.

Hayes, M.J. and Decker, W.L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. Remote Sens. 17(16), 3189-3200.

Huang, L., Yang, Q., Liang, D., Dong, Y., Xu, X., Huang, W. (2011). The estimation of winter wheat yield based on MODIS remote sensing data. Computer and Computing Technologies in Agriculture V. IFIP Advances in Information and Communication Technology, Springer, October, Berlin Heidelberg 2011, pp. 496-503.

Huang, Y., Zhu, Y., Li, W., Cao, W., Tian, Y. (2013). Assimilating Remotely Sensed Information with the Wheat Grow Model Based on the Ensemble Square Root Filter for Improving Regional Wheat Yield Forecasts. Plant Prod. Sci. 16(4), 352-364.

Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., Wu, W. (2015). Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. IEEE J. Sel. Top. Appl. 8(8), 4060-4071.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83(1), 195-213.

Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149-164.

Jaafar, H.H. and Ahmad, F.A. (2015). Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. Int. J. Remote Sens. 36(18), 4570-4589.

Jiang, Z., Chen, Z., Chen, J., Liu, J., Ren, J., Li, Z., Li, H. (2014). Application of crop model data assimilation with a particle filter for estimating regional winter wheat yields. IEEE J. Sel. Top. Appl. 7(11), 4422-4431.

Jégo, G., Pattey, E., Mesbah, S.M., Liu, J., Duchesne, I. (2015). Impact of the spatial resolution of climatic data and soil physical properties on regional corn yield predictions using the STICS crop model. Int. J. Appl. Earth Obs. 41, 11-22.

Johnson, D.M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116-128.

Kolotii, A., Kussul, N., Shelestov, A., Skakun, S., Yailymov, B., Basarab, R., Lavreniuk, M., Oliinyk, T., Ostapenko, V. (2015). Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine. Int. Arch. Photogramm. Remote Sens. XL-7/W3, 39-44.

Kouadio, A.L., Duveiller, G., Djaby, B., Tychon, B., Defourny, P. (2010). Wheat yield estimates at NUTS-3 level using MODIS data: an approach based on the decreasing curves of green area index temporal profiles. In Proceedings of RSPSoc2010 Annual Conference. 1st-3rd September 2010, Cork, Ireland (Nottingham: RSPSoc) (pp. 214-221).

Kumar B. and Tuti A. (2016). Effect and adaptation of climate change on fodder and livestock management. International journal of science, environment and technology 5(3): 1638-1645.

Kussul, N., Shelestov A., Skakun S. (2009). Grid and sensor web technologies for environmental monitoring. Earth Sci. Inform. 2(1): 37–51.

Lei, L., Stauffer, D.R., Deng, A. (2012). A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. Q. J. Roy. Meteor. Soc. 138(669), 2066-2078.

Li, A., Liang, S., Wang, A., Qin, J. (2007). Estimasting crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogramm. Eng. Rem. S. 73, 1149-1157.

Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., Wang, J. (2014). Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. Ecol. Model. 291, 15–27.

Liu, X., Jin, J., Herbert, S.J., Zhang, Q., Wang, G. (2005). Yield components, dry matter, LAI and LAD of soybeans in Northeast China. Field Crops Res. 93, 85-93.

López-Bellido, F.J., López-Bellido, R.J., Khalil, S.K., López-Bellido, L. (2008). Effect of Planting Date on Winter Kabuli Chickpea Growth and Yield under Rainfed Mediterranean Conditions. Agron. J. 4, 957-964.

Lobell, D.B. (2013). The use of satellite data for crop yield gap analysis. Field Crops Res. 143, 56-64.

Lopresti, Mariano F., Carlos M. Di Bella, Américo J. Degioanni. (2015). Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. Information Processing in Agriculture 2(2), 73-84.

Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., Wu, S. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. Math. Comput. Model. 58(3), 634-643.

Machwitz, M., Giustarini, L., Bossung, C., Frantz, D., Schlerf, M., Lilienthal, H., Udelhoven, T. (2014). Enhanced biomass prediction by assimilating satellite data into a crop growth model. Environ. Modell. Softw. 62, 437-453.

Maki, M. and Homma, K. (2014). Empirical Regression Models for Estimating Multiyear Leaf Area Index of Rice from Several Vegetation Indices at the Field Scale. Remote Sens. 6, 4764-4779.

Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J. F., Petit, M. (2014). Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. Remote Sens. 6(7), 6620-6635.

Mosleh, M.K., Hassan, Q.K., Chowdhury, E.H. (2015). Application of remote sensors in mapping rice area and forecasting its production: A review. Sensors 15(1), 769-791.

Motha, R.P. (2011). Use of Crop Models for Drought Analysis. Droughttigation Center Faculty Publications. p 58.

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

Mumby, P.J., Green, E.P., Edwards, A.J., Clark, C.D. (1999). The cost-effectiveness of remote sensing for tropical coastal resources assessment and management. J. Environ. Manage. 55(3), 157-166.

Nguy-Robertson, A.L., Gitelson, A.A., Peng, Y., Viña, A., Arkebauer, T.J., Rundquist, D.C. (2012). Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. Agron. J. 104, 1336-1347.

Noureldin, N.A., Aboelghar, M.A., Saudy, H.S., Ali, A.M. (2013). Rice yield forecasting models using satellite imagery in Egypt. Egypt. J. Remote Sens. Space Sci. 16,125–131.

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26-33.

Ranum P., Pena-Ross J.P., Garcia-Casal M.N. (2014). Global maize production, utilization and consumption. Annals of the New York Academy of Sciences 1312: 105–112.

Rembold, F., Atzberger, C., Savin, I., Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sens. 5, 1704-1733.

Rosenzweig, C., Iglesias, A., Yang, X.B., Epstein, P.R., Chivian, E. (2001). Climate change and extreme weather events; implications for food production, plant diseases, and pests. Global change Hum. health 2, 90-104.

Rouse, J., Hass, R., Schell, J., Deering, D. (1973). Monitoring vegetation systems in the great plains with ERTS. In *Third ERTS Symposium*, NASA: SP-351 I, p. 309–317.

Russell W.A**.** (1991). Genetic improvement of maize yields. Adv. Agron. 46, 245–298.

Sharifi, M.A. (2000). Crop Inventory and production forecasting using remote sensing and agrometorological models: the case of major agricultural commodities in Hamadan Province, Iran. Int. Arch. Photogramm. Remote Sens. 33, 1364-1372.

Shao, Y., Campbell, J.B., Taff, G.N., Zheng, B. (2015). An analysis of cropland mask choice and ancillary data for annual corn yield forecasting using MODIS data. Int. J. Appl. Earth Obs. Geoinf. 38, 78-87.

Sinha, S.K., Rao, N.H., Swaminathan, M.S. Food security in the changing global climate. In the Conference Proceedings for The Changing Atmosphere: Implications for Global Security, Toronto, Canada, 27-30 June 1988.

Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y., Duc, H.N., Nguyen, L.D. (2013). Prediction of rice crop yield using MODIS EVI− LAI data in the Mekong Delta, Vietnam. Int. J. Remote Sens. 34, 7275-7292.

Trenberth, K.E. (2008). Climate change and extreme weather events. In Procs. of Catastrophe Modeling Forum: Chaging Climatic Dyn. and Catast. Model Projections, P Epstein, Ed.; New York, NY, USA, May 2008, p. 7.

Wasseige, C., Bastin, D., Defourny, P. (2003). Seasonal variation of tropical forest LAI based on field measurements in Central African Republic. Agric. For. Meteorol. 119, 181-194.

Wheeler T. and von Braun J. (2013). Climate change impacts on global food security. Science. 341(6145), 508-513.

Wolfe, D.W., Henderson, D.W., Hsiao, T.C., Alvino, A. (1988). Interactive Water and Nitrogen Effects on Senescence of Maize. I. Leaf Area Duration, Nitrogen Distribution, and Yield. Agron. J. 80, 859-864

Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J. (2011). Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield. International Conference on Computer and Computing Technologies in Agriculture. Springer Berlin Heidelberg.

Yang, P., Tan, G. X., Zha, Y., Shibasaki, R. (2004). Integrating remotely sensed data with an ecosystem model to estimate crop yield in north China. In Proceedings of XXth ISPRS Congress Proceedings Commission VII, WG VII (July, Vol. 2, pp. 150-156).

Yuping, M., Shili, W., Li, Z., Yingyu, H., Liwei, Z., Yanbo, H., Futang, W. (2008). Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. Int. J. Appl. Earth Obs. 10(4), 426-437.

Zhao, Y., Chen, S., Sheng, S. (2013). Assimilating remote sensing information with crop model using ensemble Kalman filter for improving LAI monitoring and yield estimation. Ecol. Model. 270, 30–42.

Zhu, X., Zhao, Y., Feng, X. (2013). A methodology for estimating Leaf Area Index by assimilating remote sensing data into crop model based on temporal and spatial knowledge. Chinese Geogr. Sci. 23(5), 550-561.

Zhang, H., Chen, H., Zhou, G. (September 2012). The Model of Wheat Yield Forecast based on MODIS-NDVI: A Case Study of XinXiang. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sci., I-7, XXII ISPRS Congress. Melbourne, Australia, 25 August – 01 September 2012.

Chapter I.

Predicting regional corn yields with MODIS data

ABSTRACT

A simple approach was developed to predict corn yields using the MoDerate Resolution Imaging Spectroradiometer (MODIS) data product from two geographically separate major corn crop production regions: Illinois, USA and Heilongjiang, China. The MOD09A1 data, which are 8-day interval surface reflectance data, were obtained from day of the year (DOY) 89 to 337 to calculate the leaf area index (LAI). The sum of the LAI from early in the season to a given date in the season [end of DOY (EOD)] was well fitted to a logistic function and represented seasonal changes in leaf area duration (LAD). A simple phenology model was derived to estimate the dates of emergence and maturity using the LAD-logistic function parameters b_1 and b_2, which represented the rate of increase in LAI and the date of maximum LAI, respectively. The phenology model predicted emergence and maturity dates fairly well, with root mean square error (RMSE) values of 6.3 and 4.9 days for the validation dataset, respectively. Two simple linear regression models (Y_P and Y_F) were established using LAD as the variable to predict corn yield. The

yield model Y_P used LAD from predicted emergence to maturity, and the yield model Y_F used LAD for a predetermined period from DOY 89 to a particular EOD. When state/province corn yields for the validation dataset were predicted at DOY 321, near completion of the corn harvest, the Y_P model, including the predicted phenology, performed much better than the Y_F model, with RMSE values of 0.68 t/ha and 0.66 t/ha for Illinois and Heilongjiang, respectively. The Y_P model showed similar or better performance, even for the much earlier pre-harvest yield prediction at DOY 257. In addition, the model performance showed no difference between the two study regions with very different climates and cultivation methods, including cultivar and irrigation management. These results suggested that the approach described in this paper has potential for application to relatively wide agroclimatic regions with different cultivation methods and for extension to the other crops. However, it needs to be examined further in tropical and sub-tropical regions, which are very different from the two study regions with respect to agroclimatic constraints and agrotechnologies.

Keywords: MODIS, yield, phenology, LAD, logistic function

INTRODUCTION

Global warming is projected to accompany more frequent extreme weather events, such as heavy rainfall, high temperature, and drought (Rosenzweig et al., 2001; Trenberth, 2008). Although these climate changes will positively affect crop production in some regions, and crop production for food, feed, and fodder in other regions will be negatively affected by climate changes (Kumar and Tuti, 2016; Sinha et al., 1988), aggravating all dimensions of food security. Timely and reliable information on crop growth and yield at the regional, national, and global scales is essential for food security and trade policies (Hutchinson et al., 1991; Kouadio et al., 2012; Macdonald and Hall, 1980).

Notable advances in remote sensing have enabled reliable and timely prediction of crop yields (Huang et al., 2013; Kussul et al., 2009). Vegetation indices (VIs), which are calculated by combining the reflectance values of several spectral bands from multi spectral satellite systems have been used directly as crop yield estimators (Figueiredo et al., 2016; Huang et al., 2013; Li et al., 2007) and/or to estimate intermediate crop growth variables, such as LAI and biomass, for yield prediction (Jaafar and Ahmad, 2015; Kolotii et al., 2015; Rembold et al., 2013; Sibley et al., 2014). In addition to VIs and crop

growth variables, crop phenology information is required for reliable crop yield prediction because the effects of environmental conditions on crop yield differ by growth stage (Prasad et al., 2008). Crop phenology has also been estimated using remote sensing data products (Islam and Bala, 2008; Qiu et al., 2016; Sakamoto et al., 2010). For example, Islam and Bala (2008) used NDVI and LAI derived from remote sensing data to identify the planting and ending dates of potato, Qiu et al. (2016) estimated phenological dates in multiple cropping regions using EVI2 of the MODIS data, and Sakamoto et al. (2010) applied a two-step filtering approach to detect maize and soybean phenology with time-series MODIS data.

Predicting crop yield using remote sensing data products often depends on an empirical approach that relates VIs alone or in combination with remote sensing derived meteorological variables to the reported crop yields (Bolton and Friedl, 2013; Li et al., 2007; Johnson, 2014; Shao et al., 2015). Bolton and Friedl (2013) used a simple linear regression model to predict corn and soybean yield using VIs (i.e., NDVI, EVI2, and NDWI) from MODIS in the Central USA and concluded that the EVI2 index exhibited better ability to predict maize yield than the NDVI and that the use of crop phenology information from MODIS improved the model predictability. Johnson (2014) developed a regression tree model using the linear and/or exponential relationship of MODIS-derived NDVI and daytime land surface temperature

with county-level yield statistics to predict corn and soybean yield for 12 states in central and northern USA. Shao et al. (2015) developed simple linear regression model using multi-temporal NDVI from MODIS to predict county-level corn yields for the entire Midwestern USA and confirmed that using a digital elevation model climate data as additional model inputs slightly improved the performance of the regional corn yield model. Although simple models to predict crop yield can be developed (Huang et al., 2013; Huang et al., 2011), it is likely that the spatial portability of the model would be limited because the parameters of the empirical equation that were estimated in the study region would not be applicable to the other regions with different agroclimate and agrotechnologies. Additionally, VIs such as NDVI can detect inter-annual fluctuation of crop yield due to weather conditions while they cannot detect human-induced factors that result in increased crop yield (Huang et al., 2013), making the NDVI-crop yield regression model difficult to extend to other regions (Huang et al., 2013; Mkhabela and Mashinini, 2005). Sophisticated approaches requiring data in addition to remote sensing data have also been developed to improve predictions of crop growth and yield (Huang et al., 2013; Lobell, 2013; Prasad et al., 2006). For instance, Prasad et al. (2006) used a piecewise linear regression model with a break point to predict corn and soybean yield using monthly NDVI, soil moisture, surface temperature, and total rainfall. Nevertheless, it is preferable to develop a crop

yield prediction model with wide spatial portability using a small dataset (Zhang et al., 2012).

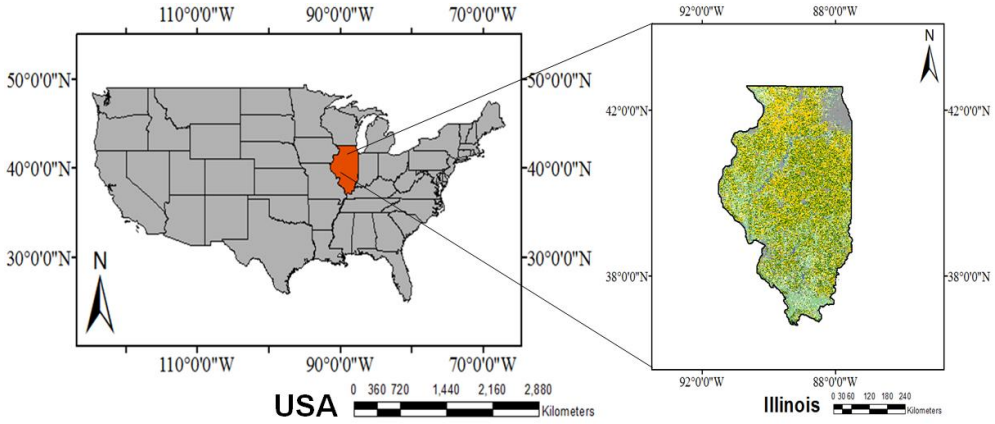
In addition to VIs, remote sensing derived LAI has also been used to predict crop biomass and yield (Casanova et al., 1997; Maki and Homma, 2014; Son et al., 2013). LAI, the one side total leaf area per unit of ground area (Wasseige et al., 2003), is a key biophysical variable to determine crop growth and yield (Bach, 1998; Noureldin et al., 2013). Son et al. (2013) developed a regression model using MODIS-LAI and EVI as input variables used to predict rice yield in the Mekong delta, Vietnam. LAD is the integrated value of LAI over time. Although LAD is an important crop parameter that has a strong correlation with dry matter production and grain yield (Liu et al., 2005; López-Bellido et al., 2008), it has not been used in a yield prediction model using remote sensing data. LAD has been reported to have positive correlation with corn yield under water and nitrogen stress conditions imposed at different growth stages (Wolfe et al., 1988) and under varying planting densities of three maize hybrids (Alias et al., 2011), and genetic differences in photosynthetic duration (longer LAD) were reported to be associated with a longer grain filling duration and higher yield (Russell, 1991). These findings suggest that LAD would have greater potential to represent corn yield variability in regions with diverse agroclimate and agrotechnologies compared to VIs and LAI at a particular crop growth stage.

The objectives of this study were to develop a simple approach to predict crop phenology and corn yield using a minimum set of remote sensing data products and to examine the spatial portability of the simple method. This study focused on predicting corn yield in major production areas of the USA and China with different agroclimates and agrotechnologies.

MATERIALS AND METHODS

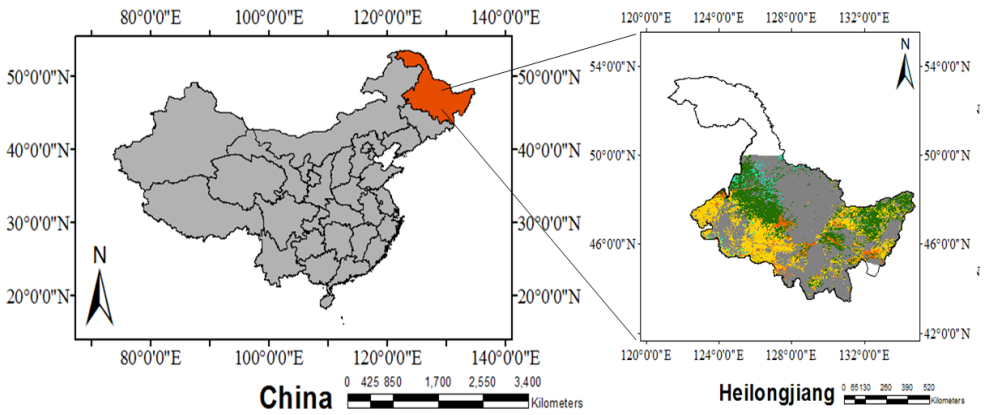
1. Study area

A simple approach using remote sensing data products was developed to predict crop yields in major crop production areas in the USA and China, where varieties and cultivation methods differ considerably. Illinois, USA (40°0′0″N, 89°0′0″W) (Figure I–1a) and Heilongjiang Province, China (47°50′0″N, 127°40′0″E) (Figure I–2a) were selected as the regions of interest. Illinois and Heilongjiang are located on opposite sides of the Earth. The latitude of Heilongjiang is higher than that of Illinois. Accordingly, the environmental conditions (e.g., temperature, precipitation, and solar radiation) differ between sites. The annual mean temperature in Illinois is approximately 10 °C, while in Heilongjiang, it is −4 °C to 4 °C. These different environmental conditions are why different crop varieties and cultivation methods are used at the two sites (Eberhart and Russell, 1966).



(a) (b)

Figure I–1. Map of the USA showing the location of Illinois (a) and crop cover data for Illinois in 2013 (b) (Corn is indicated by yellow).



(a) (b)

Figure I–2. Map of the China showing the location of Heilongjiang (a) and crop cover data for Heilongjiang in 2012 (b) (Corn is indicated by yellow).

The corn production area and the quantity of corn produced as a proportion of the national total in Illinois were approximately 32% and 15% in 2013, respectively. The corresponding values in Heilongjiang were approximately 9% and 12%.

2. Data and processing

2.1. Crop yield and phenology data

Corn yields in Illinois and Heilongjiang (Figure I–3) were obtained from official reports provided by statistical services in each country to evaluate the reliability of a model for predicting corn yield. In Illinois, crop yields from 2000 to 2013 were gathered from the National Agricultural Statistics Service (NASS) by agricultural district (AD) (available at https://quickstats.nass.usda.gov/). Crop area and total production of each crop were also obtained from the NASS by county and state. The unit system for crop yield in Illinois was converted from “bushels per acre” to “kilograms per hectare” to fit the unit system used in Heilongjiang.

The yields and planted area for 2002 to 2012 in Heilongjiang Province were collected from the Heilongjiang Statistical Yearbook by prefecture and province. The corn yields in Heihe and Daxinganling prefectures in Heilongjiang Province were excluded from the calculation to minimize the computing resources because mean corn production was considerably lower

than in other areas of the province.

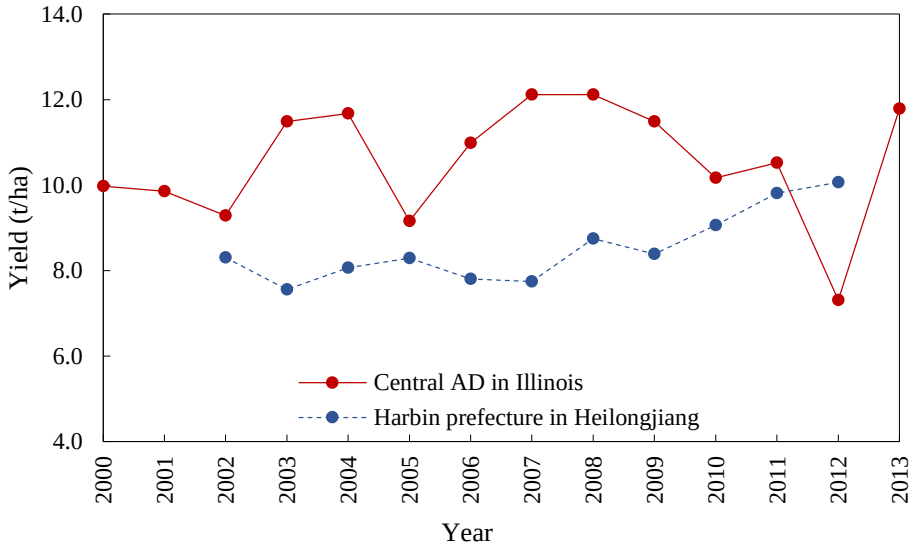


Figure I–3. Reported corn yields from 2000 to 2013 in the central AD, Illinois and from 2002 to 2012 in Harbin Prefecture, Heilongjiang.

The data of corn emergence and maturity date for Illinois were obtained from weekly crop progress reports by the NASS-Illinois Field Office (IFO). These data which were provided by AD were only available for five ADs (northwest, northeast, central, western, and eastern districts) between 2003 and 2012 due to missing data in some ADs. The dates of emergence and maturity were obtained. Because of the limited availability of phenology data in Heilongjiang Province, a model was established and validated to predict the

phenological stages using data only for the ADs in Illinois, USA and was applied to predict the phenological stages in Heilongjiang, China.

The phenology data were used to evaluate the reliability of a model to predict corn phenological dates using remote sensing data. The date on which a given phenological stage reached 50% in the study area, e.g., an AD, was used to represent when a phenological stage occurred. It was assumed that the proportion of area in a given stage would increase linearly over a two-week period. The weekly data for the proportion of area in which the crop was at the phenological stage of interest were collected from the NASS-IFO. Those weekly data were used to compare the estimated dates on which a phenological stage occurred in the AD. For example, the day on which the corn crop reached a phenological stage was determined by a linear interpolation between the weekly proportion of area for the given phenological stage.

2.2. Crop cover data

Crop cover data for Illinois (Figure I–1b) were obtained from cropland data layers (CDL) provided by NASS to identify the area where a given crop was grown (available at https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php). Crop cover data were obtained from 2000 to 2013. The crop cover data from 2002 to 2012 in China (Figure I–2b) were generated using remote sensing

(Kim et al., 2016). Crop areas were identified from the MODIS land cover type product (MCD12Q1). Major crops, including corn, within the crop areas were classified using a maximum likelihood classifier and time-series MODIS 16-day NDVI dataset (MOD13Q1). Crop cover data, which have a spatial resolution of 250 m, were subjected to c using ENVI (ExelisVIS: Exelis Visual Information Solutions, Boulder, CO, USA). The projections of the crop cover maps in both regions were converted to a Universal Transverse Mercator (UTM) projection and WGS-84 coordinates at 1 km spatial resolution.

2.3. Remote sensing data

The MODIS surface reflectance data for 2000–2013 were obtained from the MOD09A1 data which have a spatial resolution of 500 m. These data were downloaded from Reverb, which is a web-based remote sensing data portal operated by the National Aeronautics and Space Administration (available at http://modis.gsfc.nasa.gov/). The h10v04, h10v05, h11v04, and h11v05 tiled grid data for Illinois were collected from DOY 89 to 329. The h26v04 and h27v04 tiled grid data were obtained for the same period in Heilongjiang.

The near-infrared (NIR; band 2) and red (band 1) band of reflectance data were prepared to estimate LAI using a series of data tools (Figure I–4). First, all tiled reflectance data were mosaicked into a single dataset using interface description language (IDL; ExelisVIS). The projection of mosaic data was

converted to UTM projection at 1 km spatial resolution and WGS-84 geographic latitude and longitude coordinates using IDL, which applies the triangulation wrap and nearest neighbor resampling methods. Reflectance data were resized to fit the size and georeference of crop cover data using FWTools, which is a collection of open-source GIS applications. Gridded data for the area where a given crop was grown by year, were prepared by extracting the reflectance data for the given extent to compare grids belonging to the crop of interest in the crop cover data using MATLAB (MathWorks Inc., Natick, MA, USA).

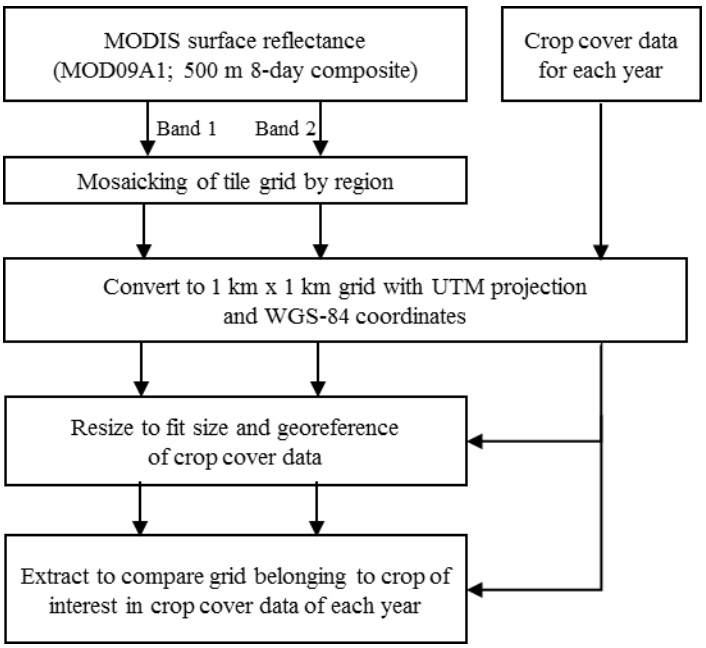


Figure I–4. Data-flow diagram of the surface reflectance and crop cover data used to standardize the data.

Cropland was identified using the proportion of crop cover in each pixel of the MODIS product using a mixed problem and low resolution (Bolton and Friedl, 2013; Shao et al., 2015). Crop cover data were overlapped with the MODIS data to calculate the percentage of pixels where a given crop was grown in the research region. It was assumed that the crop of interest would be grown within pixels where the percentage of corn was > 60% and > 90%. Although identifying a crop using the percentage of crop cover could be used for pure growing regions of the crop, a percentage value was not set for the identification method, so the mixed problem differed by region and crop. This method can cause problems when integrating from a smaller to a larger scale due to the pixels excluded below the threshold percentage. In this study, the pixels for corn growing in Illinois and Heilongjiang were identified according to the data standardization procedure described in Figure I–4, and the number of identified pixel in Illinois and Heilongjiang were approximately 610,000 and 860,000, respectively.

3. Estimation of LAI

3.1. Estimation of LAI using remote sensing data

LAI was calculated using the NIR and red bands of the reflectance data. Nguy-Robertson et al. (2012) suggested that the combined vegetation index (CVI) is more accurate than a single vegetation index set to estimate LAI. Instead of using the LAI products from the MODIS data, the LAI value was calculated as follows (Nguy-Robertson et al., 2012):

$$\mathrm{LAI} = \begin{cases} (\mathrm{NDVI} - 0.28)/0.28\,, & \mathrm{NDVI} \le 0.7 \\ (\mathrm{SR} - 1.0)/3.5\,, & \mathrm{NDVI} > 0.7 \end{cases} \quad \text{----- Equation (I–1)}$$

where NDVI and SR are the normalized difference vegetation index (Rouse et al., 1973) and simple ratio (Jordan, 1969), respectively. NDVI and SR were determined as follows:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{red})/(\mathrm{NIR} + \mathrm{red}) \quad \text{----- Equation (I–2)}$$

$$\mathrm{SR} = \mathrm{NIR}/\mathrm{red} \quad \text{----- Equation (I–3)}$$

where NIR and red indicate the near-infrared and red band spectra.

3.2. Estimation of daily LAI using a logistic function

Various smoothing algorithms were used to reduce noise in remote sensing time-series data (Shao et al., 2016). The Savitzky-Golay, asymmetric Gaussian, double-logistic, Whittaker smoother, and discrete Fourier transformation smoothing algorithms have been applied to the NDVI data in the MODIS product. In this study, MODIS-derived LAI was smoothed using a simple method. It was assumed that the total sum of the LAI over time would be shaped like a logistic function. For example, the LAI would be negligible early in the season until the emergence date. However, the daily LAI sum would increase rapidly during the vegetative growth stage, but the sum of the LAI would become negative after flowering. Finally, the sum of the LAI over a season would remain relatively constant until harvest after physiological maturity (Figure I–5).

A logistic function was used to represent the temporal change in leaf area duration (LAD), which is the integral of LAI over time as follows (Equation (I–4)):

$$\int LAI(t)\,dt = \frac{b_3}{1.0+\exp[-b_1(t-b_2)]} \quad \text{----- Equation (I–4)}$$

where b_1 represents the rate of LAI growth, b_2 represents the date of the

maximum LAI, b_3 represents the cumulative LAI at physiological maturity, and t indicates days after planting.

Although it would be preferable to determine the t as the date after the exact planting date, it was challenging to obtain this date in each grid cell. Because a logistic function was used in Equation (I–4), it was assumed that the t could be determined using a date earlier than the actual planting date. In this study, DOY 89 was used as the beginning of the cropping season, which was the a prior to the actual planting for the regions of interest. The order of the remote sensing data products since DOY 89 was used to determine the t and conveniently estimate the logistic function parameters. For example, t was 2 when data production on DOY 97 was used for Equation (I–4). The simplex method (Nelder and Mead, 1965) was used to determine the parameter values of Equation (I–4) for each grid cell. The sum of the LAI derived from the MODIS product until a given date was compared with that obtained from Equation (I–4). The LAI values from 89 to 337 DOY were accumulated at eight-day intervals for each grid cell at a 1-km resolution. The sum of the square error between the observed and simulated values of cumulative LAI was minimized to obtain parameter values for b_1, b_2, and b_3 in the simplex algorithm.

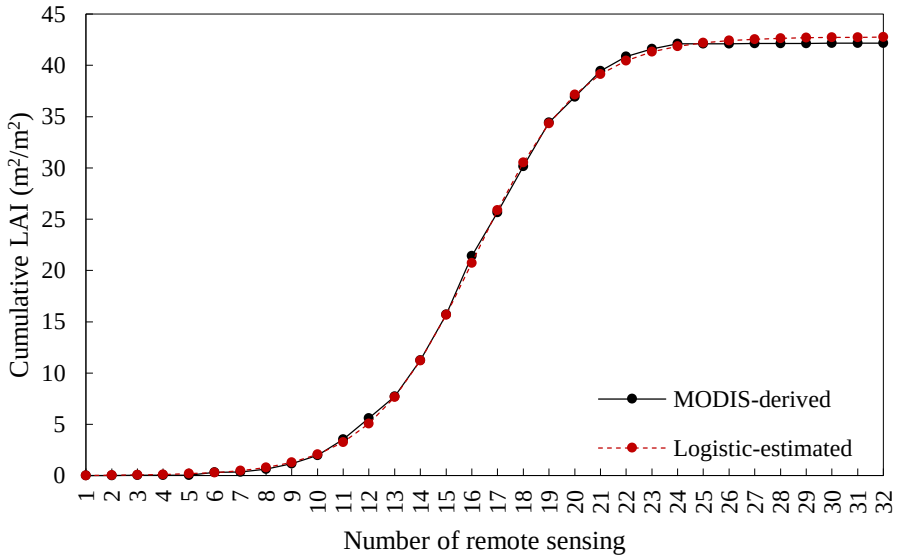


Figure I–5. MODIS-derived and logistic-estimated cumulative LAI for corn in Illinois. The X-axis denotes the order of the dates for the remote sensing data product. For example, products for DOY 89 and 97 are indicated by 1 and 2, respectively.

The derivative of Equation (I–4) on a given date represented the LAI on that date.

$$\text{LAI} = \frac{b_3 \cdot b_1 \cdot \exp[-b_1(d-b_2)]}{\{1.0+\exp[-b_1(d-b_2)]\}^2} \quad \text{----- Equation (I–5)}$$

Daily changes in LAI (Figure I–6) were determined for each grid cell.

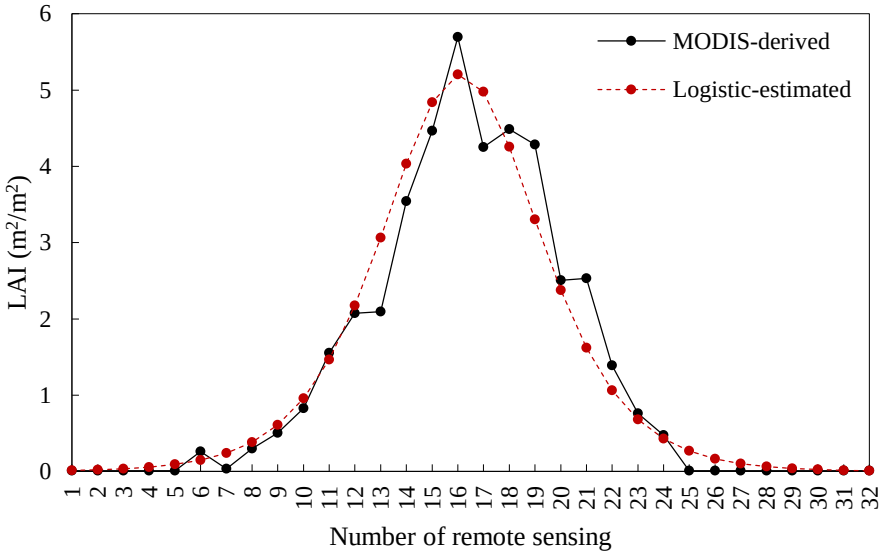


Figure I–6. MODIS-derived and logistic-estimated LAI for corn in Illinois. X-axis denotes the order of the dates for the remote sensing data product. For example, the products for DOY 89 and 97 are indicated by 1 and 2, respectively.

4. Prediction of crop phenological dates

The dates of emergence and maturity were estimated using a simple empirical equation (Equation (I–6)). It was assumed that date D_P for a given phenology P could be determined using b_1 and b_2 from the LAD-logistic function (Equation I–5), which represent the rate of LAI increase and the date of maximum LAI at a given site, respectively, as follows:

$$D_P = b_2 + \tau_P + \rho_P/b_1 \quad \text{----- Equation (I–6)}$$

where τ_P and ρ_P are empirical coefficients, τ_P represents the overall time difference between the date of the maximum LAI and a given phenological stage P, and ρ_P represents the impact of the increase in LAI on the phenological change over time. τ_P and ρ_P were estimated using the simplex method. The NASS-derived dates for a given phenology were compared with those obtained from Equation (I–5).

5. Prediction of crop yield

Daily LAD was also determined from the date of emergence to that of maturity for each grid cell, as follows (Power et al., 1967):

$$\mathrm{LAD} = \frac{(LAI_{d+1}+LAI_d)}{2} \quad \text{----- Equation (I–7)}$$

where LAI_d is the LAI value on d. The sum of the LAD values was obtained for the growing periods, e.g., from emergence to maturity, for each grid cell. Then, the sum of those values was averaged by region, e.g., AD and prefecture in a season as follows:

$$\mathrm{ALAD} = \frac{1}{n}\sum_{c=0}^{n}\sum_{d=0}^{m} LAD_{dc} \quad \text{----- Equation (I–8)}$$

where c and d represent the grid cell index in the region in which the corn was grown and the date index from emergence to maturity at c, respectively.

Huang et al. (2011) reported that crop yield can be determined with the LAI using simple linear regression. In this study, ALAD was used as an independent variable in the simple linear regression to predict grain yield. The coefficients of the regression line were obtained for the reported yields and yield predictions for a given district D in a season as follows:

$$\text{Yield}_D = \alpha \text{ALAD}_D + \beta \quad \text{----- Equation (I–9)}$$

where α and β are coefficients estimated by the least square difference method.

5.1. Y_P model using LAD accumulated from the estimated emergence date

The Y_P model used the predicted phenological stages and smoothed LAI values to calculate LAD. LAI values were accumulated from the emergence to maturity dates predicted using Equations (I–4) and (I–6). The last date on which the remote sensing data products were used was denoted EOD. The EOD value was the date elapsed from the initial date of analysis when the approach was used in other regions, e.g., in the southern hemisphere. However, DOY was used to indicate the EOD for simplicity. The coefficients of the

simple linear regression for the Y_P model were obtained for the relations between the reported yields and yield estimates. No remote sensing data was available from EOD to maturity if the EOD was earlier than the maturity date. In such cases, Equation (I–5) was used to estimate daily LAI until maturity.

5.2. Y_F model using LAD accumulated from an arbitrarily fixed date

To estimate the dates of emergence and maturity, phenology data in the region of interest must be available to determine τ_P and ρ_P for Equation (I–6). Therefore, the Yp model would not be applicable to a region where few phenology data are available. Instead of using the smoothed LAD values from the predicted emergence to maturity date, the Y_F model used the LAD accumulated with LAI values calculated from Equation (I–1) for the period representing the entire growing season, e.g., from DOY 89 to 337. The ALAD value was calculated to determine yields using Equation (I–9) between DOY 89 and a particular EOD or between DOY 89 and 337. The coefficients of the simple linear regression for the Y_F model were obtained for the relations between the reported yields and yield estimates.

5.3. Comparison between the Y_P and Y_F models

The Y_P and Y_F models were different with respect to the period and the LAI resource data used to calculate LAD. The Y_P model used LAD from the emergence to maturity dates predicted using Equations (I–4) and (I–6) and LAD was calculated from the LAI values smoothed using Equations (I–4 and I–5). The Y_F model used LAD from DOY 89 to a particular EOD, and LAD was calculated with raw LAI values derived from Equation (I–1).

6. Classification of the calibration and validation datasets

Yield data were classified into two groups for the calibration and validation of phenology and yield. Three years in which yield data were available for both Illinois and Heilongjiang were chosen randomly for validation of corn yield prediction at the state scale. Calibration datasets for predicting phenology and yield at the district scale were selected randomly from approximately 70% of the remaining datasets, including data for the 2000, 2001, and 2013 seasons, during which yield data were available in either China or the USA. In total, 69 and 62 sets of yield data were used as the calibration datasets for Illinois and Heilongjiang, respectively (Figure I–7). Yield data at the district scale, e.g., AD and prefecture for the USA and China, respectively, were pooled to determine the empirical parameter values, including τ_P, ρ_P, α, and β. The other datasets (i.e., 30 and 26 sets in Illinois and Heilongjiang,

respectively) were used to validate the yield models.

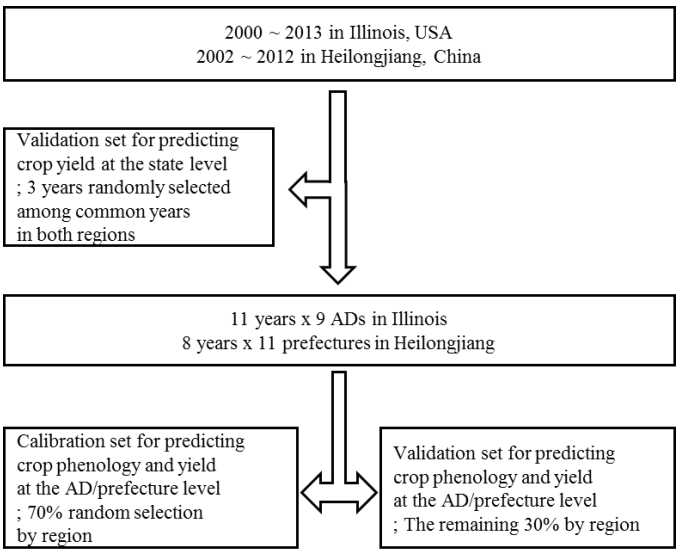


Figure I–7. Data-flow diagram showing the process used to classify the calibration and validation datasets.

τ_P and ρ_P were determined for districts where phenology data were available in the calibration datasets. For example, all phenological dates were available in only five ADs in the USA for 2003 to 2012. The simplex method was used to determine τ_P and ρ_P for emergence and maturity, respectively. The crop phenology prediction model was validated with districts where phenology data were available in the validation datasets. Because no phenological data were available in China, τ_P and ρ_P obtained from the USA

were used to determine whether Equation (I–5) could be used to predict crop yield.

7. Degree of agreement analysis

The degree of agreement statistics was determined by spatial scale, season, and variable. The RMSE value was determined to compare the observed and estimated phenology dates, as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad \text{----- Equation (I–10)}$$

where n represents the number of comparisons, and P_i and O_i are the estimated and observed data.

Four statistics, including correlation (R^2), normalized root mean square error (NRMSE), the concordance correlation coefficient (CCC), and RMSE were determined for crop yield. The yield of each grid cell was summarized by individual season and district, e.g., AD and prefecture to compare with the reported yields at the regional scale. Yields by region were aggregated to compare the predicted and reported yields at the state scale, e.g., Illinois and Heilongjiang, by season. The NRMSE was determined as follows (Soler et al., 2007):

$$\mathrm{NRMSE} = \mathrm{RMSE} \times \frac{100}{\mathrm{M}} \quad \text{----- Equation (I–11)}$$

where M is the mean yield reported by the statistical agencies in China and the USA. The simulated results were considered to be either excellent (NRMSE < 10%), good (10% < NRMSE < 20%), fair (20% < NRMSE < 30%), or poor (NRMSE > 30%). The CCC value, which was used to represent precision and accuracy, was determined as follows (Lin, 1989):

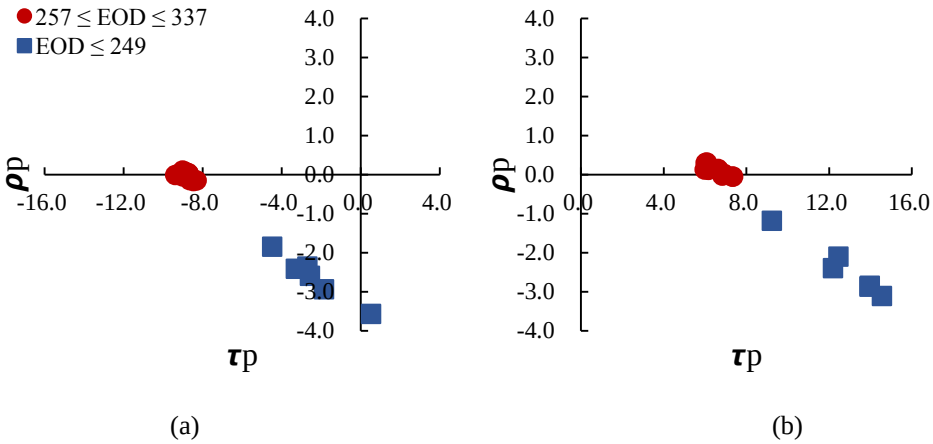
$$\mathrm{CCC} = \frac{2\rho\sigma_s\sigma_y}{\sigma_{\mathrm{x}}^2+\sigma_{\mathrm{y}}^2+(\mu_{\mathrm{x}}-\mu_{\mathrm{y}})} \quad \text{----- Equation (I–12)}$$

where ρ is the correlation coefficient between the estimated and reported data, and σ and μ are the standard deviations and means of the estimated and observed data, respectively. CCC ranges from -1 to 1, where -1 and 1 represent perfect disagreement and agreement, respectively, and 0 represents independence between the estimated and reported data (Carrasco et al., 2009).

RESULTS

1. Crop phenology

The estimated τ_P and ρ_P in Equation (I–6) for the calibration datasets differed by the last date on which the remote sensing product was used (EOD) (Figure I–8). τ_P and ρ_P tended to group together by EOD. For example, τ_P and ρ_P for the emergence and maturity dates were similar after EOD 257.



(a) (b)

Figure I–8. Estimated τ_P and ρ_P values by EOD of emergence (a) and maturity (b). The X-axis denotes the τ_P values and the Y-axis denotes the $\rho_P$values.

The RMSEs for the occurrence date of the phenological stage for the calibration datasets differed by phenological stage (Figure I–9). For example, the RMSE for maturity (<7 days) were relatively higher than those for

emergence (<5 days). Temporal changes in RMSE differed by phenological stage. Although RMSE decreased with increasing EOD until EOD 257, the error values after EOD 257 were relatively similar.

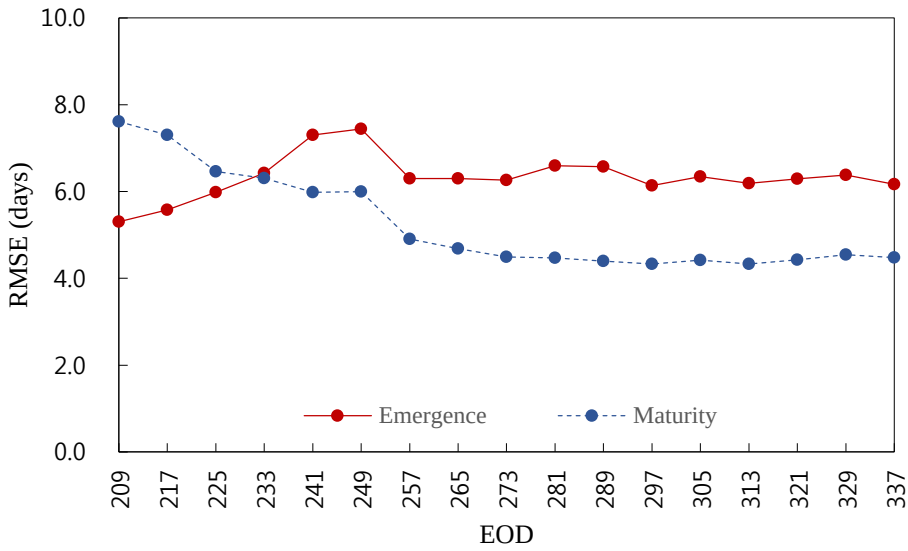


Figure I–**9.** RMSE values by EOD for the phenology model estimates using a logistic function and the calibration datasets.

The crop phenology prediction model was validated for EODs 209, 257, and 321. Corn flowering occurred on approximately DOY 209 in Illinois, and the corn harvest was completed near DOY 321. DOY 257 was selected to examine the reliability of the yield predictions in advance because the date was one of the earliest EODs on which maturity dates for corn were reliably

predicted in the calibration dataset.

The occurrence dates of a given phenology stages were estimated within a reasonable range of error (Figure I–10). For example, emergence and maturity dates were estimated within seven days for most districts, e.g., 67% and 75%, respectively, when these phenological stages were estimated on EOD 257.

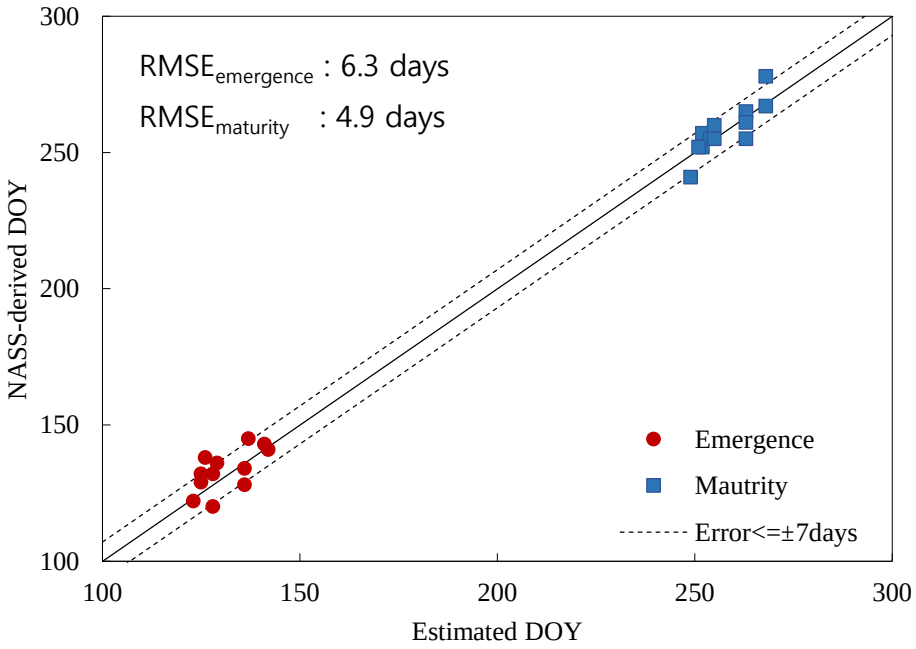


Figure I–10. Comparison between NASS-derived and estimated phenological stages at EOD 257.

The degree of agreement tended to be higher for the estimates of the timing of maturity than those of emergence (Table I–1). For example, the RMSE values for maturity were lower than those for emergence. Temporal changes in the degree of agreement also differed.

Table I–1. Statistical indices for the phenological stage estimate model in the validation dataset.

EOD	Emergence			Maturity		
	R^2	RMSE (days)	NRMSE (%)	R^2	RMSE (days)	NRMSE (%)
209	0.51	5.31	3.98	0.27	7.61	2.95
257	0.41	6.30	4.72	0.70	4.91	1.90
321	0.35	6.29	4.72	0.78	4.43	1.71

2. Crop yield at the district level

Estimated α and β values of Equation (I-9) for the calibration datasets changed with EOD in the Y_P model (Figure I–11) and grouped together after EOD 257 because they depended on an estimated date for a phenological event, e.g., emergence or maturity. The α and β values changed over the EODs in the Y_F model using an arbitrarily fixed starting date of DOY 89

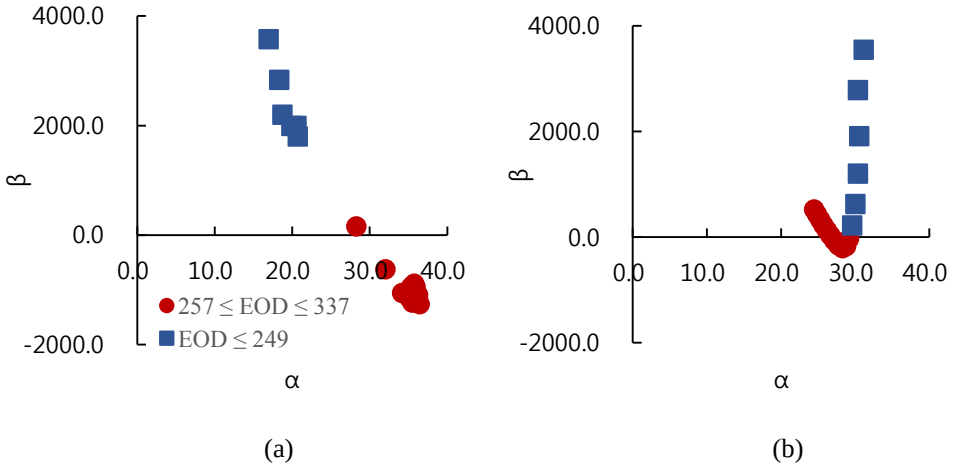


Figure I–11. Estimated α and β values for the Y_P (a) and Y_F (b) corn yield models. The X-axis denotes the α values and the Y-axis denotes the β values.

During calibration, the Y_P model had a greater degree of agreement for predicting crop yield than the Y_F model (Figure I–12). The R^2 values for all EODs of the Y_P model were higher than those for the Y_F model when predicting crop yield by district. Although both models predicted differences in crop yield between Illinois and Heilongjiang, the Y_P model was more precise than the Y_F model (Figure I–13).

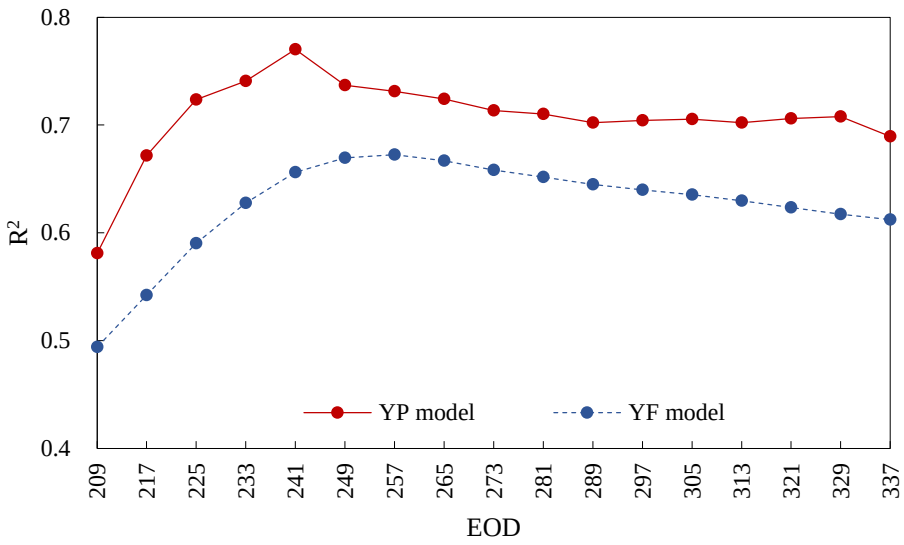
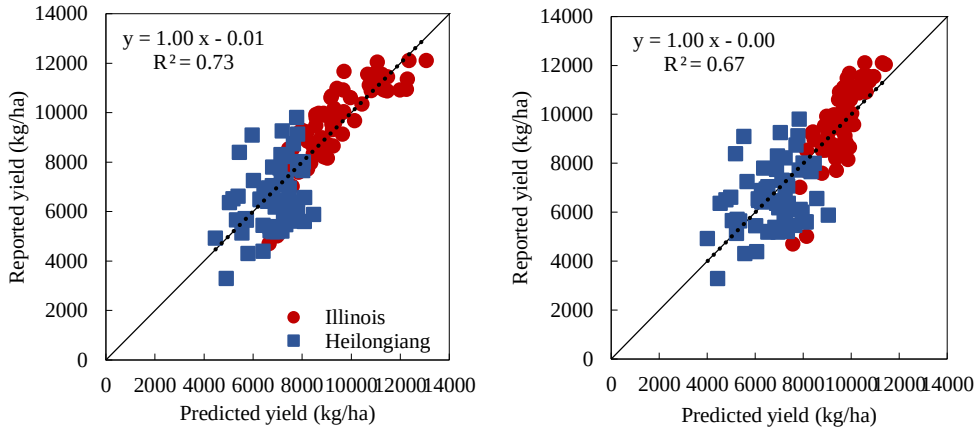


Figure I–12. R^2 values by EOD for the corn yield predictive models at agricultural the district/prefecture level.



(a) (b)

Figure I–13. Comparison of the reported and predicted agricultural district/prefecture-level corn yields at EOD 257 from the Y_P (a) and Y_F (b) models in Illinois and Heilongjiang.

The crop yield prediction model was validated for EODs 209, 257, and 321. The corn flowering date occurred on approximately DOY 209 in Illinois, and the corn harvest was completed near DOY 321. DOY 257 was selected to examine the reliability of the yield predictions in advance because the date was one of the earliest EODs on which maturity dates for corn were reliably predicted in the calibration dataset.

The Y_P model based on the phenology dates identified from the logistic function had a lower error than that of the Y_F model based on fixed dates (Table I–2). The Y_P and Y_F models predicted differences in yield between the two regions, and corn yield was considerably lower in Heilongjiang than in Illinois. Although the Y_P and Y_F models for EOD 257 had similar R^2, the Y_P model had greater R^2 than the Y_F model for EODs 209 and 321. The Y_F model always had a higher RMSE than the Y_P model.

Table I–2. Statistical indices for the corn yield prediction models at the agricultural district/prefecture level in Illinois and Heilongjiang.

EOD	Y_P model			Y_F model		
	R^2	RMSE (kg/ha)	NRMSE (%)	R^2	RMSE (kg/ha)	NRMSE (%)
209	0.65	1158.82	13.20	0.57	1283.03	14.62
257	0.68	1083.74	12.35	0.68	1086.66	12.38
321	0.70	1042.43	11.88	0.66	1127.67	12.85

The errors in yield prediction were similar for EODs 257 and 321 in the Y_P model. The NRMSE of the crop yield prediction on EOD 257 was slightly greater than that on EOD 321. However, the Y_F model had the lowest error on EOD 257.

3. Crop yield at the state/province level

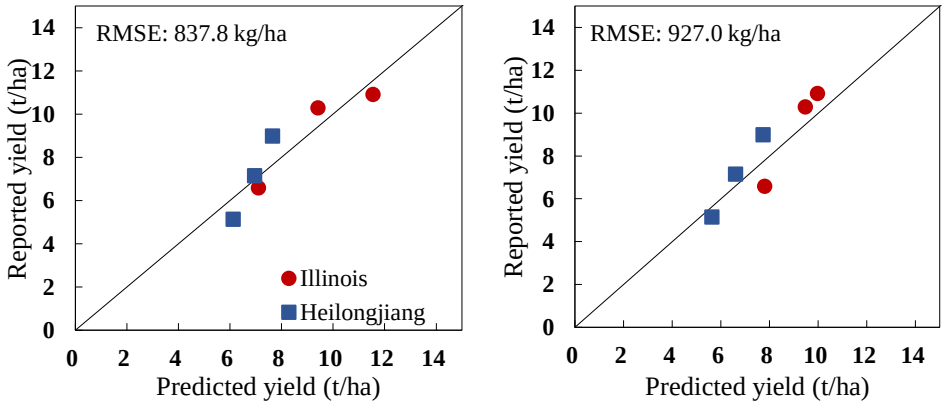
The statistical indices for the corn yield prediction model over 3 randomly selected years (2003, 2009, and 2012), are shown in Table I–3. The degree of agreement statistics for the Y_P model were high for corn yield prediction in Illinois and Heilongjiang. In particular, the Y_P model performed the best in Illinois, while the Y_F model had a similar performance in both regions. The R^2 and CCC values of both prediction models and regions, except for EOD 209, were > 0.87 and 0.68, respectively. The NRMSE values were 7.39–13.59. Both prediction models exhibited good corn yield prediction performance in the two regions.

Table I–3. Statistical indices for the yield prediction models at the state level in Illinois and at the province level in Heilongjiang.

Region	EOD	Y_P model				Y_F model			
		R^2	RMSE (kg/ha)	NRMSE (%)	CCC	R^2	RMSE (kg/ha)	NRMSE (%)	CCC
IL	209	0.43	1785.62	19.27	0.21	0.18	2137.85	23.07	-0.12
	257	0.87	687.68	7.42	0.93	0.99	1006.67	10.86	0.78
	321	0.95	684.72	7.39	0.91	0.94	1068.36	11.53	0.74
HE	209	0.99	1115.68	15.72	0.59	0.96	1008.13	14.20	0.68
	257	0.99	964.88	13.59	0.68	0.99	839.75	11.83	0.79
	321	0.99	664.07	9.36	0.87	0.99	816.82	11.51	0.79

IL: Illinois, HE: Heilongjiang

The reported and predicted corn yields by state/province for the prediction models at EOD 257 in Illinois and Heilongjiang are compared in Figure I–14.



(a) (b)

Figure I–14. Comparison of the reported and predicted state/province-level corn yields at EOD 257 for the Y_P (a) and Y_F (b) models in Illinois and Heilongjiang (validation dataset: 2003, 2009, and 2012).

DISCUSSION

Simple models were developed to predict corn phenological stages and yield using only the red and NIR band surface reflectance data of the MODIS products. Sakamoto et al. (2005) suggested that the temporal discontinuity of remote sensing data makes it difficult to estimate crop properties, e.g., phenological dates; therefore different approaches have been developed. In this study, the combination of Equations (I–4) and (I–5) allowed us to overcome data discontinuity to estimate the phenological date. The remote sensing data revealed fluctuations in LAI during a season, as shown in Figure 6. However, the integrative approach using Equation (I–4) made it possible to identify the date of maximum LAI accurately when EOD was later than 249.

The phenology prediction model performance for predicting emergence and maturity dates differed depending on EOD. The errors for predicting emergence and maturity were greater when an EOD before the maximum LAI was used. For example, a reliable and consistent value of b_2 of Equation (I–4) could be obtained using remote sensing data only after the flowering period until near harvest. Because τ_P and ρ_P depend on b_2, long-term data were used to represent the growth conditions and to obtain reliable estimates of τ_P and ρ_P. The phenology model accuracy for maturity date prediction was relatively higher than for the emergence date prediction. Although the

phenology including maturity date depends on the characteristics of the cultivar such as the physiological responses to environmental conditions like photoperiod, and temperature (Cutforth and Shaykewich, 1990), changes in phenology are closely related to changes in LAI (Biswal et al., 2014). For example, flowering occurs near the time of maximum LAI, which is b_2 in Equation (I–4). Once the time of maximum LAI is predicted accurately, the maturity date can be reliably identified. Estimating the maximum LAI using remote sensing products is more accurate than measuring the LAI early in the season when LAI is < 1 (Heiskanen et al., 2012). Emergence dates are related to planting dates, and the planting date is influenced by the soil temperature and soil moisture and varies widely by district and season. For example, 33.6 ℃ is the optimal soil temperature for corn germination (Itabari et al., 1993). Thus, the planting date is delayed until the temperature condition is met in a specific field (Chen and Wiatrak, 2010; Thomison and Nielson, 2002), which results in large variability in the length of the effective growing season in a particular region. As a result, the errors in the emergence dates were relatively higher than the errors in the maturity dates. The present crop phenology prediction model based on MODIS-derived LAD-logistic function provided more reliable and accurate maturity date predictions than those of previous studies. Sakamoto et al. (2011) reported that the RMSE value of the maturity date for corn in Illinois was 5.9 days when using a two-step filtering

method and WDRVI derived from MODIS data. The RMSE value of the maturity date on EOD 257 was approximately 17% less than in a previous study. In contrast, the model resulted in greater prediction error for the emergence date. For example, the RMSE of the emergence date predicted on EOD 257 was 6.3 days, whereas the value in a previous study (Sakamoto et al., 2011) was 4.9 days. The phenological dates of the previous study were used to predict specific shape by crop. The present phenology prediction model, which is based on MODIS-derived LAD-logistic function and does not need specific shape by crop, can be applied to other crops.

The α and β values in the Y_P model using the predicted phenological dates grouped together after EOD 257 (Figure I–11). The α and β values of the model depended on the reliability of b_3, which represents the maximum cumulative LAI, LAD. When an EOD before the maximum LAI was used, b_3 was less reliable because the maximum LAI had not been reached in the field; thus, data up to a specific period after the maximum LAI are essential to reliably predict b_3. Due to errors in remote sensing data, e.g., caused by clouds and rainfall, maximum LAI estimates can have considerable errors, even after the maximum LAI. Nevertheless, α and β were similar after a set of EOD, suggesting that a small range of α and β values could be obtained to predict corn yield. The α and β values for the Y_F model using arbitrarily fixed starting date depended on the last DOY (i.e., EOD) used for the LAD calculation, and

the error for the predicted corn yield also depended on EOD because LAI decreased after the maximum LAI. This model also required data up to a specific period (i.e., EOD 257) after maximum LAI to predict corn yield reliably.

The errors for predicting corn yield in the Y_P and Y_F models at the AD level were relatively small. The RMSE for the Y_P model was 1.08 t/ha and 1.09 t/ha at EOD 257 for 20 districts (Table I–2). In the previous studies, the RMSE of the predicted corn yield ranged from 1.2 t/ha to 1.7 t/ha, depending on the region of interest and season (Sakamoto et al., 2013; Doraiswamy et al., 2007; Johnson, 2014). Doraiswamy et al. (2007) reported an RMSE of 1.21 t/ha for corn yield in several Illinois counties. Because of differences in spatial extent and scale, e.g., district or county and extent, caution is needed when interpreting the differences in RMSE values between the current and previous studies. Nevertheless, the RMSE of both the Y_P and Y_F models tended to be smaller than those of the previous studies, which merits further validation in a variety of additional regions.

For the validation datasets for each region, the errors for predicting corn yield at the state/province scale were relatively small in both the Y_P and Y_F models. When yield prediction was performed after EOD 257, the Y_P model exhibited RMSE values of 0.69–0.68 t/ha in Illinois and 0.66–0.96 t/ha in Heilongjiang, while the Y_F model had RMSE values of 1.0–1.1 t/ha in Illinois

and 0.82–0.84 t/ha in Heilongjiang (Table I–3). Even for the validation data set, the two models revealed little decay in corn yield prediction performance compared with the calibration data set, with RMSEs of 0.84 and 0.93 t/ha across the two regions for Y_P and Y_F, respectively. In addition the current model performances were comparable to previous studies: the RMSE values for predicting corn yield ranged from 0.62 to 1.45 t/ha in major USA production regions (Prasad et al., 2006; Shao et al., 2015), and Sakamoto et al. (2013) reported that the RMSE of the corn yield prediction was 0.83 t/ha in Illinois. The corn yield prediction errors using only LAD result from changes in specific leaf area (SLA) in a given season and region. SLA is the ratio of leaf area to leaf biomass (Setiyono et al., 2008) and is affected by environmental conditions (Gunn et al., 1999), such as weather and disease (Kim et al., 2012). As a result, SLA varies by season, even when the same crops are cultivated at a given site (Maki and Homma, 2014). Because remote sensing products represent LAI instead of biomass, the change in SLA affects the accuracy of the biomass and yield estimation when using LAI. Therefore, the inclusion of other factors affecting SLA would improve the performance yield prediction model using LAI and LAD as yield estimator.

Corn management practices were different between Illinois and Heilongjiang. Irrigation practices have been widely adopted in Illinois, while rainfed cultivation is a common practice in Heilongjiang. Corn cultivars were

also different by latitude and temperature, and corn yield in Illinois was higher than corn yields in Heilongjiang. Nevertheless, the current yield prediction models using only LAD showed similar high performances high performance for the two study regions, Illinois and Heilongjiang, which have different crop cultivar characteristics, crop management, and environments. This result is due to the nature of LAD used as predictor; the fluctuation in LAD accurately represents the yield variation caused by environmental stress, such as drought, cultivation management, including nitrogen fertilization and planting density, and the genetic improvement of corn hybrids (Wolfe et al., 1988; Alias et al., 2011; Russell, 1991).

REFERENCES

Alias, M.A., Bukhsh, H.A., Ahmad, R., Malik, A.U., Hussain, S., Ishaque, M. (2011). Profitability of three maize hybrids as influenced by varying plant density and potassium application. J. Anim. Plant Sci. 21, 42-47.

Bach, H. (1998). Yield estimation of corn based on multitemporal LANDSAT-TM data as input for agrometeorological model. Pure and Applied Optics. 7, 809–825.

Biswal, A.N.I.M.A., Sahay, B., Ramana, K.V., Rao, S.V.C.K., Sai, M.S. (2014). Relationship between AWiFS derived Spectral Vegetation Indices with Simulated Wheat Yield Attributes in Sirsa district of Haryana. Int. Arch. Photogramm. Remote Sens. 40, 689.

Bolton, D.K. and Friedl, M.A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agr. Forest Meteorol. 173, 74-84.

Carrasco, J.L., King, T.S., Chinchilli, V.M. (2009). The concordance correlation coefficient for repeated measures estimated by variance components. J. Biopharm. Stat. 19, 90-105.

Casanova, D., Epema, G.F., Goudriaan, J. (1997). Monitoring rice reflectance at field level for estimating biomass and LAI. Field Crops Res. 55, 83-92.

Chen, G.H. and Wiatrak, P. (2010). Soybean development and yield are influenced by planting date and environmental conditions in the southeastern coastal plain, United States. Agron. J. 102, 1731-1737.

Cutforth, H.W. and Shaykewich, C.F. (1990). A temperature response function for corn development. Agr. Forest Meteorol. 50, 159-171.

Doraiswamy, P.C., Akhmedov, B., Beard, L., Stern, A., Mueller, R. (2007). Operational prediction of crop yields using MODIS data and products. International archives of photogrammetry, remote sensing and spatial information, Sciences Special Publications, Commission Working Group VIII WG VIII/10. Ispra, Italy: European Commission DG JRC-Institute for the Protection and Security of the Citizen, 1-5.

Eberhart, S.T. (1966). Russell, W.A. Stability parameters for comparing varieties. Crop Sci. 6, 36-40.

Figueiredo, G.K., Brunsell, N.A., Rocha, J.V., Lamparelli, R.A., Picoli, M.C. (2016). Using temporal stability to estimate soya bean yield: a case study in Paraná state, Brazil. Int. J. Remote Sens. 37, 1223-1242.

Gunn, S., Farrar, J.F., Collis, B.E., Nason, M. (1999). Specific leaf weight in barley: individual leaves versus whole plants. New Phytol. 413, 45–51.

Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., Vesanto, V. H., Korhonen, L., Majasalmi, T. (2012). Seasonal variation in MODIS LAI for a boreal forest area in Finland. Remote Sens. Environ. 126, 104-115.

Huang, L., Yang, Q., Liang, D., Dong, Y., Xu, X., Huang, W. (2011). The estimation of winter wheat yield based on MODIS remote sensing data. Computer and Computing Technologies in Agriculture V. IFIP Advances in Information and Communication Technology, Springer, October, Berlin Heidelberg 2011, pp. 496-503.

Huang, J., Wang, X., Li, X., Tian, H., Pan, Z. (2013). Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. PLoS ONE 8, e70816.

Hutchinson, C.F. (1991). Uses of satellite data for famine early warning in sub-Saharan Africa. Int. J. Remote Sens. 12, 1405–1421.

Islam, A.S. and Bala, S.K. (2008). Assessment of Potato Phenological Characteristics Using MODIS-Derived NDVI and LAI Information. Gisci. Remote Sens. 45, 454-470.

Itabari, J.K., Gregory, P.J., Jones, R.K. (1993). Effects of temperature soil water status and depth of planting on germination and emergence of maize (Zea mays) adapted to semi-arid eastern Kenya. Expl. Agric. 29, 351-364.

Jaafar, H.H. and Ahmad, F.A. (2015). Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. Int. J. Remote Sens. 36, 4570-4589.

Jordan, C.F. (1969). Derivation of leaf-area index from quality of light on the forest floor. Ecology 50, 663-666.

Johnson, D.M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116-128.

Kim S.H., Hong S.Y., Sudduth K.A., Kim Y., Lee K. (2012). Comparing LAI estimates of corn and soybean from vegetation indices of multi-resolution satellite images. Korean J. Remote Sens. 28, 597–609.

Kim, Y., Lee, K.D., Na, S.I., Hong, S.Y., Park, N.W., Yoo, H.Y. (2016). MODIS data-based crop classification using selective hierarchical classification. Korean J. Remote Sens. 32, 235-244.

Kolotii, A., Kussul, N., Shelestov, A., Skakun, S., Yailymov, B., Basarab, R., Lavreniuk, M., Oliinyk, T., Ostapenko, V. (2015). Comparison of biophysical and satellite predictors for wheat yield forecasting in Ukraine. Int. Arch. Photogramm. Remote Sens. XL-7/W3, 39-44.

Kouadio, L., Duveiller, G., Djaby, B., Jarroudi, M.E., Defourny, P. (2012). Estimating regional wheat yield from the shape of decreasing curves of green area index temporal profiles retrieved from MODIS data. Int. J. Appl. Earth Obs. Geoinf. 18, 111-118.

Kumar B. and Tuti A. (2016). Effect and adaptation of climate change on fodder and livestock management. Intl. J. Sci. Environ. Technol. 5, 1638-1645.

Kussul, N., Shelestov, A., Skakun, S. (2009). Grid and sensor web technologies for environmental monitoring. Earth Sci. Inform. 2, 37–51.

Li, A., Liang, S., Wang, A., Qin, J. (2007). Estimasting crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. Photogramm. Eng. Rem. S. 73, 1149-1157.

Lin, L. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics. 45, 255–268.

Liu, X., Jin, J., Herbert, S.J., Zhang, Q., Wang, G. (2005). Yield components, dry matter, LAI and LAD of soybeans in Northeast China. Field Crops Res. 93, 85-93.

López-Bellido, F.J., López-Bellido, R.J., Khalil, S.K., López-Bellido, L. (2008). Effect of Planting Date on Winter Kabuli Chickpea Growth and Yield under Rainfed Mediterranean Conditions. Agron. J. 4, 957-964.

Lobell, D.B. (2013). The use of satellite data for crop yield gap analysis. Field Crops Res. 143, 56-64.

Macdonald, R.B. and Hall, F.G. (1980). Global crop forecasting. Science 208, 670–679.

Maki, M. and Homma, K. (2014). Empirical Regression Models for Estimating Multiyear Leaf Area Index of Rice from Several Vegetation Indices at the Field Scale. Remote Sens. 6, 4764-4779.

Mkhabela, M.S. and Mashinini, N.N. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAAs-AVHRR. Agric. For. Meteorol. 129: 1–9.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. The Computer J. 7, 308-313.

Nguy-Robertson, A.L., Gitelson, A.A., Peng, Y., Viña, A., Arkebauer, T.J., Rundquist, D.C. (2012). Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. Agron. J. 104, 1336-1347.

Noureldin, N.A., Aboelghar, M.A., Saudy, H.S., Ali, A.M. (2013). Rice yield forecasting models using satellite imagery in Egypt. Egypt. J. Remote Sens. Space Sci. 16,125–131.

Power, J.F., Willis, W.O., Reichman, G.A. (1967). Effect of soil temperature, phosphorus and plant age on growth analysis of barley. Agron. J. 59, 231-234.

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26-33.

Prasad, P.V.V., Staggenborg, S.A., Ristic, Z. (2008)/ Impacts of drought and/or heat stress on physiological, developmental, growth, and yield processes of crop plants. In Response of Crops to Limited Water: Understanding and Modeling Water Stress Effects on Plant Growth Processes; Ahuja, L.H., Saseendran, S.A., Eds.; Advances in Agricultural Systems Modeling Series 1. ASA-CSSA: Madison, WI, USA, pp. 301–355.

Qiu, B., Feng, M., Tang, Z. (2016). A simple smoother based on continuous wavelet transform: Comparative evaluation based on the fidelity, smoothness and efficiency in phenological estimation. Int. J. Appl. Earth Obs., 47, 91-101.

Rembold, F., Atzberger, C., Savin, I., Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sens. 5, 1704-1733..

Rosenzweig, C., Iglesias, A., Yang, X.B., Epstein, P.R., Chivian, E. (2001). Climate change and extreme weather events; implications for food production, plant diseases, and pests. Global change hum. health 2, 90-104.

Rouse, J., Hass, R., Schell, J., Deering, D. (1973). Monitoring vegetation systems in the great plains with ERTS. In *Third ERTS Symposium*; NASA: SP-351 I; pp. 309–317.

Russell, W.A. (1991). Genetic improvement of maize yields. Adv. Agron. 46, 245–298.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96, 366-374.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., Arkebauer, T.J. (2010). A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. Remote Sens. Environ. 114, 2146-2159.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A. (2011). Detecting spatiotemporal changes of corn developmental stages in the U.S. corn belt using MODIS WDRVI data. IEEE Trans. Geosci. Remote Sens. 49, 1926-1936.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J. (2013). MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sens. Environ. 131, 215-231.

Setiyono, T.D., Weiss, A., Specht, J.E., Cassman, K.G., Dobermann, A. (2008). Leaf area index simulation in soybean grown under near-optimal condition. Field Crops Res. 108, 82-92.

Shao, Y., Campbell, J.B., Taff, G.N., Zheng, B. (2015). An analysis of cropland mask choice and ancillary data for annual corn yield forecasting using MODIS data. Int. J. Appl. Earth Obs. Geoinf. 38, 78-87.

Shao, Y., Lunetta, R.S., Wheeler, B., Iiames, J.S., Campbell, J.B. (2016). An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. Remote Sens. Environ. 174, 258-265.

Sibley, A.M., Grassini, P., Thomas, N.E., Cassman, K.G., Lobell, D.B. (2014). Testing remote sensing approaches for assessing yield variability among maize fields. Agron. J. 106, 24-32.

Sinha, S.K., Rao, N.H., Swaminathan, M.S. Food security in the changing global climate. In the Conference Proceedings for The Changing Atmosphere: Implications for Global Security, Toronto, Canada, 27-30 June 1988.

Soler, C., Sentelhas, P., Hoogenboom, G. (2007). Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. Eur. J. Agron. 27, 165-177.

Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y., Duc, H.N., Nguyen, L.D. (2013). Prediction of rice crop yield using MODIS EVI− LAI data in the Mekong Delta, Vietnam. Int. J. Remote Sens. 34, 7275-7292.

Thomison, P. and Nielson, R. (2002). Impact of Delayed Planting on Heat Unit Requirements for Seed Maturation in Maize, vol. 15. Pontificia Universidad Católica deChile. Departamento de Ciencias Vegetales. Seminario Internacional Semillas:comercialización producción y tecnología, Santiago, pp. 140–164.

Trenberth, K.E. (2008). Climate change and extreme weather events. In Procs. of Catastrophe Modeling Forum: Chaging Climatic Dyn. and Catast. Model Projections, P Epstein, Ed.; New York, NY, USA, May 2008; p. 7.

Wasseige, C., Bastin, D., Defourny, P. (2003). Seasonal variation of tropical forest LAI based on field measurements in Central African Republic. Agric. For. Meteorol. 119, 181-194.

Wolfe, D.W., Henderson, D.W., Hsiao, T.C., Alvino, A. (1988). Interactive Water and Nitrogen Effects on Senescence of Maize. I. Leaf Area Duration, Nitrogen Distribution, and Yield. Agron. J. 80, 859-864.

Zhang, H., Chen, H., Zhou, G. (September 2012). The Model of Wheat Yield Forecast based on MODIS-NDVI: A Case Study of XinXiang. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sci., I-7, XXII ISPRS Congress. Melbourne, Australia, 25 August – 01 September 2012.

Chapter II
Assimilating MODIS data into a crop growth model improves regional corn yield predictions

ABSTRACT

Crop growth models and remote sensing are useful tools for predicting crop growth and yield, but each tool has inherent drawbacks when predicting crop growth and yield at a regional scale. To improve the accuracy and precision of regional corn yield predictions, a simple approach for assimilating Moderate Resolution Imaging Spectroradiometer (MODIS) product into a crop growth model was developed, and regional yield prediction performance was evaluated in a major corn-producing region in Illinois, USA. Corn yields and phenology data were collected at state and agricultural district (AD) levels from 2000 to 2013. Corn growth and yield were simulated using the Crop Environment Resource Synthesis (CERES)-Maize model with a minimum input dataset comprising planting date, fertilizer amount, genetic coefficients, soil, and weather data. Planting date for each grid was estimated using a phenology model with a leaf area duration (LAD) logistic function that describes the seasonal evolution of MODIS-derived LAD. Genetic coefficients of the maize cultivar for each grid were determined to be the

genetic coefficients of the mature group [included in Decision Support System for Agrotechnology Transfer (DSSAT) 4.6], which shows the minimum difference between the maximum leaf area index (LAI) value derived from the LAD logistic function and that simulated by the CERES-Maize model. In addition, the daily water stress factors employed in CERES-Maize model were estimated from the ratio of daily leaf area/weight growth rate estimated from the LAD logistic function to the daily leaf area/weight growth rate estimated by simulating CERES-Maize model under an auto-irrigation condition. Corn yield predictions using only the estimated planting date and maturity group were very poor under rain-fed conditions at both the AD and state levels, whereas corn yield predictions improved under the auto-irrigation condition, indicating that irrigation has been applied in a considerable portion of cornfields in Illinois. In addition to assimilation of the estimated planting date and maturity group, further assimilation of the estimated daily LAI and water stress factors also improved the corn yield prediction considerably, increasing the R^2 value from 0.72 to 0.78 and decreasing the root mean square error (RMSE) from 1.47 to 0.75 t/ha for the yearly corn yield prediction. In addition, an earlier corn yield prediction at day of the year (DOY) 257 was possible without decreased accuracy. In conclusion, the present strategy for assimilating MODIS data into a crop growth model using a minimum dataset was successful for predicting regional yields, and it should be examined for

spatial portability to diverse agro-climatic and agro-technology regions.

Keywords: Crop growth model, MODIS, Data assimilation, LAD, Water stress, Regional corn yield

INTRODUCTION

Monitoring crop growth and predicting yield are essential for proper crop management, agricultural operation improvement, and food-security policy decision making (Li et al., 2014; Zhao et al., 2013). Crop growth modeling and remote sensing have been useful tools for monitoring and predicting crop growth and yield (Dadhwal, 2003; Yuping et al., 2007). However, each tool has inherent drawbacks for predicting crop growth and yield at a regional scale (Jeong et al., 2016; Rauff and Bella, 2015).

Crop growth models have supported simulations of crop growth and development, physiological processes, and yield at the field scale since the late 1960s (Todorovic et al., 2009), and advanced computer technology allows simulations close to actual crop growth, which is regulated by the complex interaction of many factors (Oteng-Darko et al., 2012). Despite the noticeable improvement in crop growth model performance, regional prediction of crop growth and yield using crop growth models remains challenging due the difficulty of obtaining many of the model input parameters at a regional scale and uncertainties in the parameters due to spatial variability (Grassini et al., 2015; Paul et al., 2003; Imak et al., 2005).

Remote sensing data provide information related to crop growth status (Leon et al., 2003; Xiong, 2014), and various state variables associated with

crop growth have been estimated using vegetation indices derived from remote sensing data. For example, leaf area index (LAI) is estimated with vegetation indices such as the simple ratio index, normalized difference vegetation index (NDVI), and triangular vegetation index (Nguy-Robertson et al., 2012), and biomass was estimated using NDVI (Kryvobok, 2000). Although remote sensing data provide spatial information for a specific region (Ozdogan et al., 2010), the data are not consecutive due to temporal characteristics and atmospheric effects (Hadjimitsis et al., 2010; Weng, 2012). Remote sensing data only show symptoms; they cannot explain the cause of the spectral expression of a crop (Lilienthal and Schnug, 2007).

These constraints inherent in crop growth modeling and remote sensing can be overcome by integrating remote sensing data into a crop growth model (Sehgal, 2013); "forcing" and "recalibration" strategies have been used to perform this integration (Moulin et al., 1998). The forcing strategy involves updating state variables derived from remote sensing data into a crop growth model (Dadhwal, 2003). The state variables derived from remote sensing data are interpolated to obtain daily time series data due to the temporal characteristic of remote sensing data and atmospheric effects (Delecolle and Guerif, 1998). Delecolle and Guerif (1988) estimated wheat yield by updating interpolated LAI derived from SOPT/HRV into the AFRCWHEAT model. Bouman (1995) estimated biomass of winter wheat at harvest by updating the

LAI derived from radar remote sensing into the SUCROS model. Although the forcing strategy is simple, the initial conditions and/or parameters of the crop growth model should be estimated to improve prediction performance (Moulin et al., 1998). The recalibration strategy is used to adjust the initial conditions and parameters of the crop growth model using remote sensing data (Yuping et al., 2007). The ensemble Kalman filter (EnKF), a representative recalibration method, has been widely used to predict crop yield by assimilating remote sensing data into crop growth models (Ines et al., 2013; Li et al., 2014; Machwitz et al., 2014; Wu et al., 2012; Zhao et al., 2013; Zhu et al., 2013). For example, Li et al. (2014) assimilated LAI retrieved from ETM+ data into a hydrology crop growth model, which links the World Food Studies (WOFOST) model to better predict corn yields in a study region located in the middle reaches of the Heihe River basin, northwest China; parameters related to maintenance respiration, rooting depth, and soil hydraulic properties were adjusted using EnKF. Wu et al. (2011) used EnKF to assimilate MODIS-LAI into the WOFOST model to estimate winter wheat yield in Hengshui district, Hebei Province, China. Ines et al. (2013) used EnKF to assimilate soil moisture and/or MODIS-LAI into the Crop Environment Resource Synthesis (CERES)-Maize model to estimate corn yields from 2003 to 2009 in Story County, Iowa, USA. These methods use a repetitive process that adjusts initial conditions (e.g., physical attributes of soil profile) and

parameters of the crop growth model (e.g., cultivar characteristics) by minimizing the difference between remote sensing-derived values and simulated values by the crop growth model (Huang et al., 2015; Ines et al. 2013; Jiang et al., 2014). Therefore, these methods require high computational cost to predict crop yield at a large scale (Biniaz Delijani et al., 2014; Lei et al., 2012) because of the repetitive process employed to find the optimum value. Furthermore, this approach would be spatially limited due to EnKF localization using the calibration dataset (Anderson, 2012).

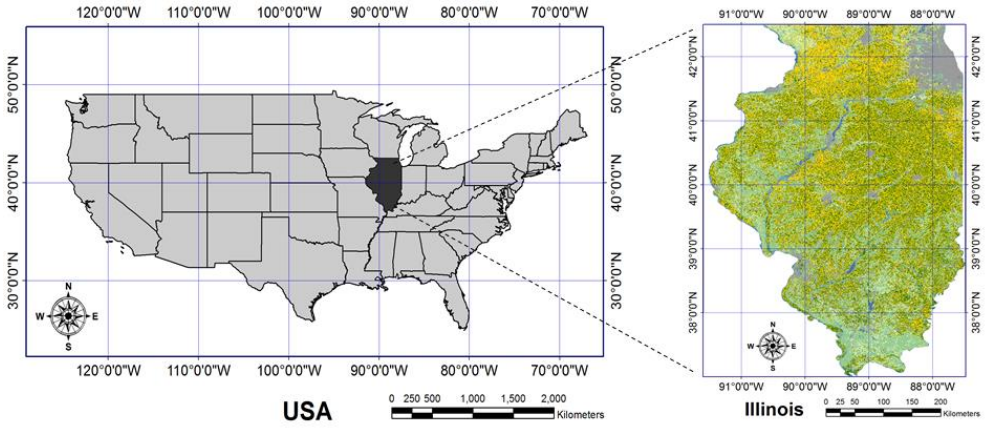
The objectives of this study were to develop a simple strategy for assimilating MODIS data into a crop growth model without re-initializing and re-parameterizing processes, and to evaluate the regional crop yield prediction performance in a major corn production region, Illinois, USA.

MATERIALS AND METHODS

1. Study area

Illinois (Figure II-1a), USA, was selected as the region of interest because this state belongs to a major corn-belt region, so corn production statistics at the county and AD levels are easily accessible. In 2013, Illinois occupied about 32% and 15% of the national total corn production area and amount,

respectively. The annual mean temperature in Illinois is approximately 11℃, and annual precipitation varies from approximately 800 to 1,200 mm according to location. Growing degree days (GDD) of corn hybrids ranges from 2,200 (northern Illinois) to 2,900℃·day (southern Illinois). Corn is planted from mid-April to late June and harvested from early September to late November (Nafziger, 2009). The irrigation system in Illinois has increased gradually, rising to approximately 625,000 acres in 2014 (Bridges et al., 2015).



(a) (b)

Figure II-1. Map of USA showing the location of Illinois (a) and corn crop cover data for Illinois in 2013 (b) (Corn is indicated by yellow).

2. Data and data processing

2.1. Corn yield and phenology data

Corn yields from 2000 to 2013 in Illinois were obtained from the National Agricultural Statistics Service (NASS) by AD and state to evaluate the reliability of assimilation strategies for predicting regional corn yields. Planted and harvested area in acres and production in bushels were available at the national, state, and county levels. Corn yields, which are measured in bushel per acre in Illinois, were converted to kilogram per hectare.

Corn phenology data in Illinois, which were provided weekly by AD and state, were obtained from the NASS-Illinois Field Office (IFO) to estimate corn planting date. Those data were available in only five ADs, including the Northwest, Northeast, Central, West, and East districts between 2003 and 2012 because phenology data and ADs were not available for several years.

The median DOY on which a given planting stage reached 50% was calculated using linear interpolation because the planting data by AD were surveyed as planted proportion by week, and the calculated median DOY parameter was used for comparisons with the estimated dates on which a certain phenological stage occurred in the AD.

2.2. Crop cover data

Corn crop cover data (Figure II-1b) were obtained from cropland data layers used by NASS to identify a region where a given crop was grown (https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php). Crop cover data in Illinois were obtained from 2000 to 2013. The projection of crop cover data was converted to a Universal Transverse Mercator projection and WGS-84 coordinates at 1-km spatial resolution using ENVI (Exelis VIS; Exelis Visual Information Solutions, Boulder, CO, USA).

2.3. Weather and soil data

Weather and soil data were generated and obtained to use as input data for the crop growth model.

Weather data including daily solar radiation ($MJ/m^2/day$), maximum and minimum temperature (℃), and rainfall (mm) at 10-km spatial resolution from 2000 to 2013 in Illinois were hind-casted using the PNU CGCM model and downscaled using the dynamic downscaling method.

Soil data were obtained from Web Soil Survey (http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx) operated by the United States Department of Agriculture, Natural Resources Conservation Service of the USA, and the data were produced by the National

Cooperative Soil Survey. Representative soils by county were selected based on the map unit symbol, which accounts for the largest area of the county and data of representative soil related to chemical and physical properties were obtained. The soil data were processed using Sbuild program within DSSAT 4.6 for subsequent use in the crop growth model. Variables related to soil water contents (e.g., saturated water content, drained upper limit, lower limit of plant extractable soil water, and root growth factor), which are dependent on physical soil properties, were calculated by soil layer using Sbuild. Soil organic carbon (OC) was calculated with soil organic matter (OM) using the following equation (Perie and Ouimet, 2008):

$$\text{Soil OC}\ (\%) = 0.4724\ \times \text{soil OM}\ (\%). \text{ ----- Equation (II-1)}$$

3. Data assimilation strategy for predicting regional corn yields

3.1. Crop growth model

The CERES-Maize model (in DSSAT4.6), which has been widely used to simulate maize growth and yield (Chisanga et al., 2015), was employed for this study. The CERES-Maize model simulates daily changes in physiological processes (e.g., phenological development, crop growth, biomass partitioning, nutrient uptake, and water use) in response to changes in environmental components (e.g., solar radiation, temperature, and rainfall) and management

practices (e.g., planting date and amount of fertilizer) and final yield (Cabrera el al., 2007; Charles et al. 2015; López-Cedrón et al., 2005).

3.2. MODIS data assimilation strategies

Planting date and maturity group estimated using MODIS-derived LAD logistic function for each grid were assimilated into the CERES-Maize model to predict corn yields by grid, as shown in Figure II-2. Daily LAI and water stress factors estimated using the MODIS-derived LAD logistic function for each grid were additionally assimilated into the CERES-Maize model to predict corn yields by grid, as shown in Figure II-3. The predicted corn yields by grid are aggregated to the AD and state levels.

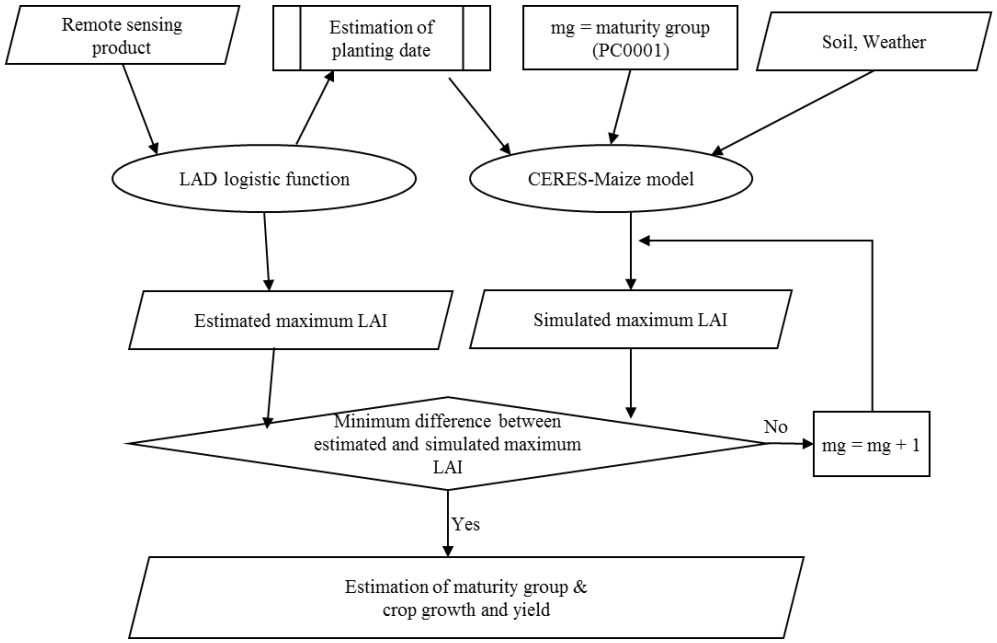


Figure II-2. Flowchart for assimilating the estimated planting date and maturity group.

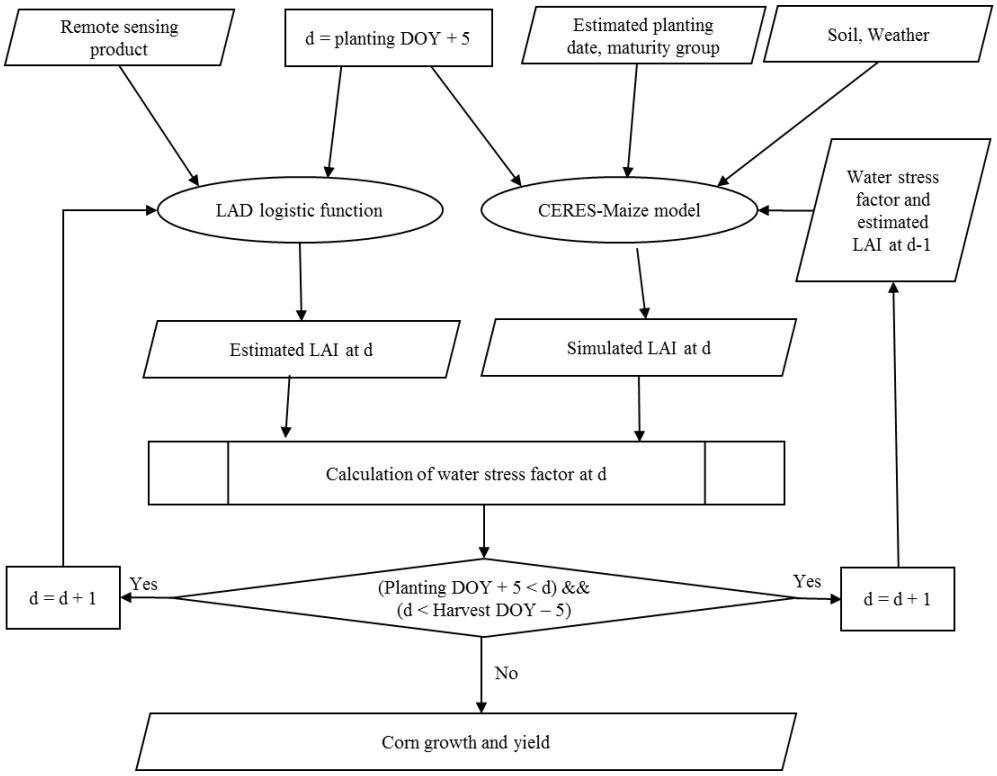


Figure II-3. Flowchart for assimilating estimated daily LAI, water stress factors, estimated planting date, and maturity group.

3.3. Estimating planting date and daily LAI

The planting date and daily LAI value were estimated via the crop phenology prediction model (Ban et al., 2016) using a logistic function describing the seasonal changes in LAD. Instead of the LAI product provided directly from MODIS, this model uses LAI values calculated from MODIS surface reflectance data (MOD09A1) of red and near infrared bands according to the equations suggested by Nguy-Roberson et al. (2012). Seasonal changes in LAD were fitted to a logistic function, and daily LAI was estimated by

differentiating the LAD logistic equation in terms of time. Ban et al. (2016) established a crop phenology model [Equation (II-2)] using the parameters (b_1 and b_2) of the LAD logistic equation as predictor variables.

$$D = b_2 + \tau + \rho/b_1, \text{ ----- Equation (II-2)}$$

where τ represents the difference between the date when LAI reaches the maximum value and the date of a given phenological stage, ρ represents the effect of an increase in LAI on phenological change over the growing season, and b_1 and b_2 represent the rate of LAI growth and the date when the LAI value reaches the maximum, respectively. Using the phenology data reported by NASS, the τ and ρ values for planting date by the end of the DOY (EOD) were estimated as represented in Table II-1. EOD denotes the last date of remote sensing data products used to fit the logistic function.

Table II-1. Estimated parameters for the crop phenology prediction model for planting date

EOD	τ	ρ
209	-4.73	0.96
257	-10.74	0.06
321	-8.77	0.36

EODs 209, 257, and 321 were selected to evaluate corn yield predictions, and EODs 209, 257, and 321 are near the usual corn flowering DOY in Illinois,

the earliest DOY when the LAD logistic function could be established reliably, and the date on which the corn harvest was completed, respectively.

3.4. Estimate of corn maturity group

The CERES-Maize model was simulated to estimate corn maturity group by grid under auto-irrigation simulation (irrigation and water management simulation options are set to automatic when required). Management practices such as planting density, depth, and amount of fertilizer that were used to simulate the CERES-Maize model are shown in Table II-2. The first and second fertilizers were applied at the planting date and 2 weeks after planting, respectively. The planting date for each grid, which was estimated using the crop phenology model [Equation (II-2)], and the estimated parameters for the planting date (Table II-1) were used for the simulation. The soil and weather data were representative soil and weather data from the grid using Arcmap (Esri, Redlands, CA, USA).

Table II-2. Management settings for the CERES-Maize model

Management	Unit	Value
Planting density	plant/m^2	7.41
Planting depth	cm	4.5
Amount of first fertilizer	kg/ha (N-P-K)	90-30-69
Amount of second fertilizer	kg/ha (N-P-K)	90-0-0

The cultivar coefficients for five generic corn hybrids, identified as PC0001–PC0005 according to growing degree days and included in DSSAT 4.6 (Table II-3), were used to identify the maturity group of the corn cultivar in a given grid. The RMSE between the maximum LAI value estimated by the LAD logistic function and that simulated by the CERES-Maize model during the growing season was calculated by maturity groups (Table II-3), and the maturity group that had the smallest RMSE was designated the mature cultivar for a given grid.

Table II-3. Genetic coefficients used to estimate corn maturity groups

Maturity group	VRNAME	P1	P2	P5	G2	G3	PHINT
PC0001	2500–2600 GDD	160.0	0.75	780.0	750.0	8.5	49.0
PC0002	2600–2650 GDD	185.0	0.75	850.0	800.0	8.5	49.0
PC0003	2650–2700 GDD	212.0	0.75	850.0	800.0	8.5	49.0
PC0004	2700–2750 GDD	240.0	0.75	850.0	800.0	8.5	49.0
PC0005	2750–2800 GDD	260.0	0.75	850.0	800.0	8.5	49.0

VRNAME: Name of cultivar, P1: Thermal time from seedling emergence to the end of the juvenile phase in degree day, P2: Photoperiod sensitivity (0–1.0) expressed in days delayed for each hour increase in photoperiod above the longest photoperiod (12.5 hours) at which development proceeds at a maximum rate, P5: Thermal time from silking to physiological maturity in degree days, G2: Potential kernel number in no. per plant, G3: Potential kernel filling rate during the linear grain filling stage in mg/kernel/day, PHINT: interval between leaf tip appearances in degree.

3.5. Estimate of daily water stress factors

The most crucial limitation for crop model-based crop yield prediction in regions where rain-fed and irrigated areas are mixed, as in Illinois, is to assess water stress as a critical factor for crop growth and yield. Leaf growth is very sensitive to inhibition by water stress (Boyer, 1968), and leaf area growth rate is a good indicator of water stress. The water stress factors (i.e., TURFAC and SWFAC) in the CERES-Maize model were estimated using the MODIS-derived LAD logistic function. TURFAC and SWFAC variables, which are water stress factors for leaf area expansion and soil water stress effect on photosynthesis, respectively, have values ranging from 0.0 to 1.0 (Singh and Helmers, 2008). In the CERES-Maize model, these variables are calculated as the ratio of total root water uptake to potential transpiration, and if the ratio is less than a specific value, the variables have values <1.0 (Tsvetsinskaya et al., 2001). The TURFAC and SWFAC variables affect the rates of crop growth and development (e.g., leaf expansion and senescence and crop phenology) (Boote et al., 2008). Finally, crop yields decrease in response to these variables (Heinemann et al., 2016).

The water stress factors were estimated differently depending on the crop growth stage (Tables II-4 and II-5) by the ratio of daily leaf area/weight growth rate estimated from MODIS-derived LAD logistic function to that estimated by the CERES-Maize simulation under the auto-irrigation condition.

Water stress factors were calculated from 5 days after planting to 5 days before harvest, which were considered the emergence date and physiological maturity date, respectively. The daily water stress factors and LAI estimated using the MODIS-derived LAD logistic function were integrated into the CERES-Maize model for predicting corn growth and yield.

Table II-4. Estimation equation of TURFAC and SWFAC variables by ISTAGE

ISTAGE	Estimation equation
1,2	$TURFAC_{est} = (LAI_d - LAI_{d-1})_{obs}/(LAI_d - LAI_{d-1})_{sim} * TURFAC_{sim}$
	$SWFAC_{est} = TURFAC_{est} * 1.5$
3	$LFWT = (LAI_{obs}/PLTPOP/0.0001/267.0)**1.25$
	$TURFAC_{est} = (LFWT_d - LFWT_{d-1})_{obs}/(LFWT_d - LFWT_{d-1})_{sim} * TURFAC_{sim}$
	$SWFAC_{est} = TURFAC_{est} * 1.5$
4,5	$SWFAC_{est} = 1.0 - (-(LAI_d - LAI_{d-1})_{obs}/PLAS*(1.0 - SWFAC_{sim}))$
	$TURFAC_{est} = SWFAC_{est}/1.5$

Obs: Value derived from MODIS data; sim: Simulated value of CERES-Maize model; est: Estimated value; LFWT: Leaf weight; TURFAC: Water stress factor for expansion; SWFAC: Effect of soil-water stress on photosynthesis; PLAS: Rate of senescence of leaf area on one plant (cm^2/day); PLTPOP: Plant population, Plants/m^2; d: Current day; d-1: Previous day.

Table II-5. Description of ISTAGE variable in CERES-Maize model

ISTAGE	Description
1	Emergence to end of juvenile stage
2	End of juvenile stage to tassel initiation
3	Tassel initiation to end of leaf growth
4	End of leaf growth to beginning effective grain filling period
5	Beginning to end of effective grain filling period

4. Degree of agreement analysis

Three types of statistics, namely R^2, RMSE, and normalized RMSE (NRMSE), were determined for crop yields. Corn yield for each grid was summarized by individual season and AD/state to compare with the reported yields at the regional scale. Corn yields were also aggregated to compare the yields predicted with those reported in Illinois by season. The RMSE value was determined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad \text{, ----- Equation (II-3)}$$

where n represents the number of comparisons, and P_i and O_i are estimated and reported data, respectively. The NRMSE was determined as follows (Soler et al. 2007):

$$\text{NRMSE} = RMSE \times \frac{100}{M}, \text{ ----- Equation (II-4)}$$

where M is the mean reported yield. Depending on the NRMSE value, the predicted results are considered excellent (NRMSE <10%), good (10% < NRMSE < 20%), fair (20% < NRMSE < 30%), and poor (NRMSE >30%).

RESULTS

1. Corn yields at the AD level

As presented in Figure II-4, corn yields simulated under three different irrigation and MODIS-derived data assimilation conditions were compared with reported corn yields. Corn yields at the AD level were simulated under two water supply conditions, i.e., "rain-fed" and "auto-irrigation," using the CERES-Maize model, which was assimilated with planting date and maturity group estimated from the LAD logistic function. In addition to the estimated planting date and maturity group, the estimated daily LAI and water stress factors were assimilated for predicting corn yields at the AD level under the auto-irrigation condition.

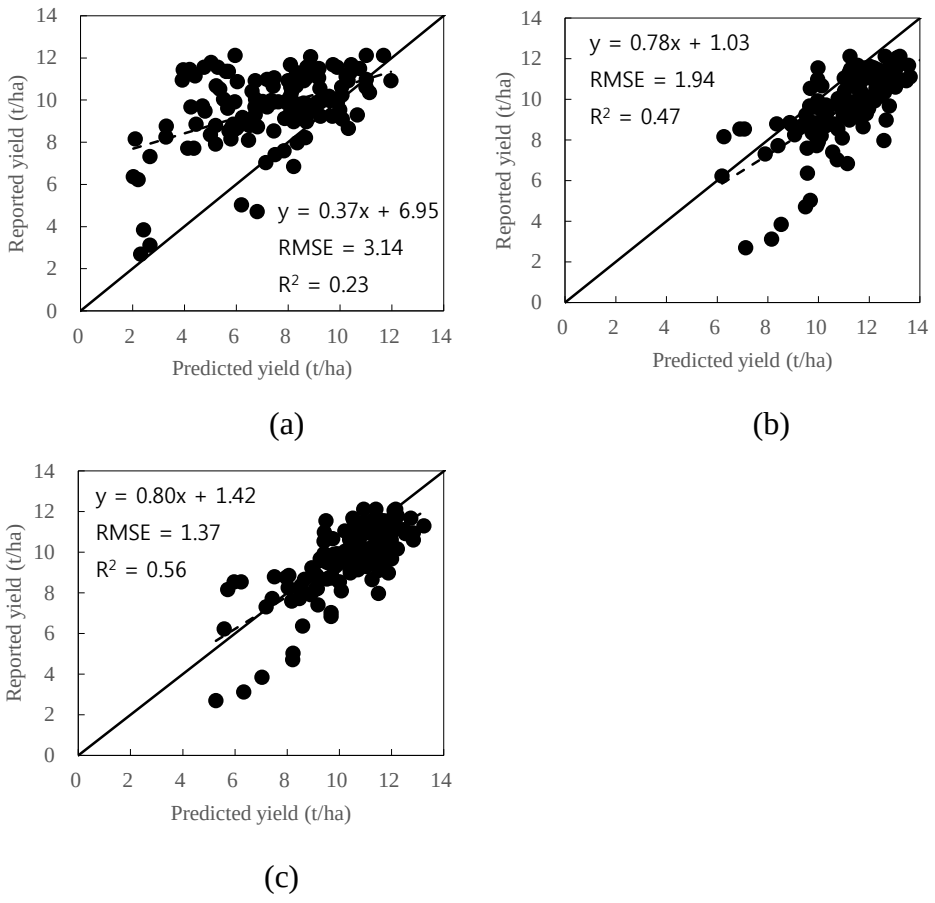


Figure II-4. Comparison of reported and predicted corn yields at the AD level with different data assimilation and simulation conditions from 2000 to 2013 in Illinois, USA, at EOD 257 [The CERES-Maize model was used for the simulation, with estimated planting date and maturity group under (a) rain-fed and (b) auto-irrigation conditions, and (c) simulated by assimilating the MODIS-derived daily LAI and water stress factors in addition to estimated planting date and maturity group under the auto-irrigation condition].

The simulation using the estimated planting date and maturity group under the rain-fed condition tended to underestimate corn yield and showed very

poor performance (Figure II-4a), whereas the simulation involving the same assimilation of the estimated planting date and maturity group under the auto-irrigation condition tended to overestimate corn yield, but the prediction performance was improved compared to that under the rain-fed condition (Figure II-4b). These results show that irrigation is practiced in a considerable portion of corn fields in Illinois. In addition, further assimilation of daily LAI and water stress factors improved the prediction performance of corn yield (Figure II-4c).

2. Corn yields at the state level

Corn yields predicted at the AD level were aggregated for comparison with the reported corn yields at the state level, as shown in Figure II-5. The overall results were similar to the predicted corn yields at the AD level. Yearly corn yields simulated with the estimated planting date and maturity group under the rain-fed condition were much lower than the reported corn yields and poorly represented the yearly variations in corn yield at the state level, whereas yearly corn yields simulated with under the auto-irrigation condition were slightly higher than the reported corn yields and represented the yearly variation in corn yield fairly well. Further assimilation of daily LAI and water stress factors with the estimated planting date and maturity group improved simulation performance by predicting corn yield and representing the yearly

yield variation better than the simulation without additional assimilation of daily LAI and water stress factors was able to do.

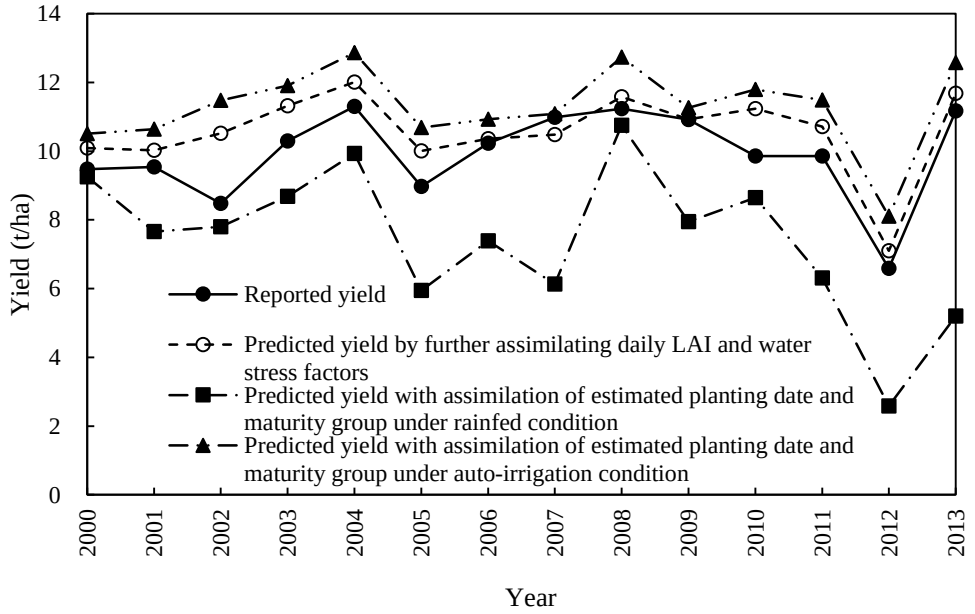


Figure II–5. Reported and predicted corn yields at the state level with different data assimilation and simulation conditions from 2000 to 2013 in Illinois, USA, at EOD 257.

The statistical indices for the corn yields predictions at the state level are shown in Table II-6. The corn yield simulation using the estimated planting date and maturity group under the rain-fed condition showed the worst performance for all EODs, whereas the corn yield simulation with the same assimilation under the auto-irrigation condition showed much better performance, increasing the R^2 value from 0.34 to 0.71 and decreasing the

RMSE from 2.98 to 1.53 at EOD 257. Additional assimilation of daily LAI and water stress factors to the estimated planting date and maturity group also resulted in further improvement of the corn yield prediction. Additional assimilation increased the R^2 value from 0.71 to 0.78 and decreased the RMSE from 1.53 to 0.88 for the EOD 257 simulation. Although corn yield simulation with additional assimilation of daily LAI and water stress factors at EOD 209 was worse than those for the other EODs, the level of agreement statistics for all EODs showed good performance, and performance improved with increasing EOD. The R^2, RMSE, and NRMSE values for predicting corn yields at all EODs were >0.57, <0.91 t/ha, and 9.19%, respectively.

Table II-6. Statistical indices for predicted corn yields at the state level with different data assimilation and simulation conditions by EOD.

EOD	R^2			RMSE (t/ha)			NRMSE (%)		
	Rain	Auto	Stress	Rain	Auto	Stress	Rain	Auto	Stress
209	0.37	0.73	0.57	2.80	1.60	0.91	28.22	16.15	9.19
257	0.34	0.71	0.78	2.98	1.53	0.88	30.02	15.42	8.91
321	0.38	0.72	0.78	3.07	1.47	0.75	30.90	14.79	7.58

Rain: Corn yield prediction with assimilation of estimated planting date and maturity group under the rain-fed condition, Auto: Corn yield prediction with additional assimilation of daily leaf area index and water stress factors under the auto-irrigation condition. Stress: corn yield prediction with assimilation of estimated planting date and maturity group under auto-irrigation condition.

DISCUSSION

Regional crop yield predictions using a crop growth model are challenging due to the large uncertainty inherent in the input data and parameters (e.g., soil properties, initial condition, crop parameters, weather, and management practices) (Hansen and Jones, 2000). Although remote sensing data provide information related to crop growth status at a regional scale, the data are not consecutive. These constraints can be overcome by assimilating remote sensing data into a crop growth model (Jiang et al., 2014). Two strategies (i.e., forcing and recalibration) were used to integrate remote sensing data into crop growth models, and prediction performance improved through use of these strategies. However, crop growth and yield predictions using these strategies were spatially limited due to estimates of the initial conditions and/or parameters for the crop growth model using a calibration dataset.

In this study, a simple data assimilation strategy was developed to improve regional corn yield prediction performance by integrating information on crop management and growth derived from MODIS data into the CERES-Maize model using a minimum input dataset. This method does not need to estimate the initial conditions and/or parameters of CERES-Maize model. Only planting date, maturity group, daily LAI, and water stress factors, which were

estimated using a MODIS-derived LAD logistic function, were assimilated into the CERES-Maize model to improve accuracy for predicting corn yield. The corn yield simulation at the AD and state levels using the estimated planting date and maturity group showed very poor performance under the rain-fed condition, whereas much improved yield prediction performance was observed under the auto-irrigation condition (Figure II-4 and Table II-6). This result suggests that irrigation has been practiced in a considerable portion of corn fields in Illinois. Bridges et al. (2015) reported that irrigation systems have increased gradually in Illinois, rising to approximately 625,000 acres in 2014. It is most important to estimate the degree of water stress directly using remote sensing and consider water stress when simulating crop growth and yield in order to improve the corn yield prediction in a region such as Illinois, where irrigation is only practiced partially and rainfall is insufficient during the growing season. Water is one of the most important factors limiting crop growth and yield (Boyer and Westgate, 2004; Davis et al., 2014; Shao et al., 2009). Leaf growth is reduced, dry matter allocation to the root is increased, and the root-to-shoot ratio decreases when water stress occurs in a plant (Guo et al., 2015; Li et al., 2009; Medeiros et al., 2012). Therefore, leaf growth rate is a good criterion to use in assessing the degree of water stress, and water stress factors can be estimated using daily crop growth rate based on the balance between soil water supply and crop water demand (Chisanga et al.,

2015). Daily water stress factors employed in the CERES-Maize model were estimated by the ratio of daily leaf area/weight growth rate estimated from the LAD logistic function to the daily leaf area/weight growth rate estimated by the CERES-Maize model under the auto-irrigation condition shown in Table II-4. In addition to the estimated planting date and maturity group, the additional assimilation of MODIS-derived daily LAI and water stress factors into the CERES-Maize model further improved yield prediction performance, as the R^2 value increased from 0.71 to 0.78, and RMSE decreased from 1.53 to 0.88 t/ha for the corn yield prediction at EOD 257. However, the corn yield simulation with the additional assimilation of daily LAI and water stress factors showed slightly poorer performance at EOD 209 than at EOD 257 and EOD 321. This may have been caused by the unreliable estimate of daily LAI and water stress factors, which was calculated from the estimated daily LAI. Ban et al. (2016) reported that the MODIS-derived LAD logistic function parameters may not have been estimated reliably at the EOD before the date of maximum daily LAI, resulting in an unreliable estimate of daily LAI.

The RMSE values for the state-level corn yields predicted with daily LAI and water stress factors estimated using the MODIS-derived LAD logistic function were 0.88 and 0.75 t/ha, respectively. Doraiswamy et al. (2005) predicted corn yield in McLean County, Illinois, USA, with a RMSE value of 0.9 t/ha using a method that adjusts for crop model parameters, and Fang et al.

(2011) predicted corn yield in several counties in Indiana, USA, with an RMSE value of 0.85 t/ha using the Markov model. Although the region and scale in the current study differed from those in previous studies, the RMSE value of the predicted corn yields achieved by additional assimilation at EOD 321 was smaller than the RMSE values reported in the previous studies. The two previous studies used a repetitive process that adjusted the environmental conditions and parameters of the crop growth model by minimizing the difference between remote sensing-derived values and simulated values in the crop growth model. For example, Fang et al. (2011) estimated planting date, population, row spacing, and quantity of nitrogen fertilizer by minimizing the difference between simulated LAI and MODIS-derived LAI. This method requires a high computational cost and a large input dataset, as well as local characteristics for the estimated parameters in the crop growth model, and would be spatially limited. However, the present assimilation strategy using minimum data (i.e., daily water stress factors, daily LAI, planting date, and maturity group) required only a few input parameters, without a re-parameterization and re-initialization process.

By assimilating daily LAI and water stress factors, the predicted yearly trend in the state corn yields was very close to the reported trend of yearly corn yields. However, corn yields were predicted to be much higher than the reported corn yield in 2002 and 2010 (Figure II-5), indicating that factors other

than water stress decreased corn yields in those years. Actual yields (i.e., reported yields) are largely affected by regional socioeconomic conditions, crop management, and disease (e.g., fertilizer and biocide use) (Marra et al., 2012; Reidsma and Ewert, 2008). Although remote sensing data were used to overcome the uncertainties caused by the large scale, not all of the information about the actual yield loss was addressed. The accuracy of corn yield predictions would improve by adding reliable information about other components (e.g., insects, pests, and extreme weather events).

REFERENCES

Anderson, J.L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. Mon. Weather Rev. 140(7), 2359-2371.

Ban, H.Y., Kim, K.S., Park, N.W., Lee, B.W. (2016). Using MODIS data to predict regional corn yields. Remote sens. 9(1), 16.

Biniaz Delijani, E., Pishvaie, M.R., Bozorgmehry Boozarjomehry, R. (2014). Distance Dependent Localization Approach in Oil Reservoir History Matching: A Comparative Study. Iran. J. Chem. Chem. Eng. 33(1), 75-91.

Boote, K.J., Sau, F., Hoogenboom, G., Jones, J.W. (2008). Experience with water balance, evapotranspiration, and predictions of water stress effects in the CROPGRO model, In: Ahuja, L.R., Reddy, V.R., Saseendran, S.A., Yu, Q. (Eds.), Response of Crops to Limited Water: Understanding and Modeling Water Stress Effect on Plant Growth Processes. Advances in Agricultural Systems Modeling Series 1. ASA, CSSA, SSSA, Madison, WI, 59–103.

Bouman, B.A.M. (1995). Crop modelling and remote sensing for yield prediction. Neth. J. Agr. Sci. 43, 143-143.

Boyer, J.S. (1968). Relationship of water potential to growth of leaves. Plant physiol. 43(7), 1056-1062.

Boyer, J.S. and Westgate, M.E. (2004). Grain yields with limited water. J. Exp. Bot. 55(407), 2385–2394.

Bridges, K., Wilson, S., Perry, R. (2015). Center Pivot Irrigation in Illinois 2012 and 2014. Illinois State Water Survey, ISWS Publications Series: Maps. http://www.isws.illinois.edu/iswsdocs/maps/ISWSMS2014-03.pdf. Accessed 05 November 2016.

Cabrera, V.E., Jagtap, S.S., Hildebrand, P.E. (2007). Strategies to limit (minimize) nitrogen leaching on dairy farms driven by seasonal climate forecasts. Agr. Ecosyst. Environ. 122(4), 479-489.

Chisanga, C.B., Phiri, E., Shepande, C., Sichingabula, H. (2015). Evaluating CERES-Maize model using planting dates and nitrogen fertilizer in Zambia. J. Agr. Sci. 7(3), 79.

Dadhwal, V.K. (2003). Crop growth and productivity monitoring and simulation using remote sensing and GIS. Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, 263-289.

Davis, R.F., Earl, H.J., Timper, P. (2014). Effect of simultaneous water deficit stress and Meloidogyne incognita infection on cotton yield and fiber quality. J. Nema. Tol. 46,108–118.

Delecolle, R., Maas, S.J., Guerif., M., Baret, F. 1992. Remote sensing and crop production models: present trends. ISPRS J. Photogramm. Remote Sens. 47:145-161.

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J. (2005). Application of MODIS derived parameters for regional crop yield assessment. Remote Sens. Environ. 97(2), 192-202.

Fang, H., Liang, S., 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.

Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Cassman, K.G. (2015). How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crop. Res. 177, 49-63.

Guo, Y., Yu, H., Kong, D., Yan, F., Liu, D., Zhang, Y. (2015). Effects of gradual soil drought stress on the growth, biomass partitioning, and chlorophyll fluorescence of Prunus mongolica seedlings. Turk. Biol. 39(4), 532-539.

Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, A., Clayton, C.R.I. (2010). Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. Nat. Hazard. Earth Sys. 10(1), 89-95.

Heinemann, A.B., Ramirez-Villegas, J., Souza, T.L.P., Didonet, A.D., Di Stefano, J.G., Boote, K.J., Jarvis, A. (2016). Drought impact on rainfed common bean production areas in Brazil. Agr. Forest Meteorol. 225, 57-74.

Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., Wu, W. (2015). Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. IEEE J. Sel. Top. Appl. 8(8), 4060-4071.

Ines, A. V., Das, N.N., Hansen, J.W., Njoku, E.G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sens. Environ. 138, 149-164.

Irmak, A., Jones, J.W., Jagtap, S.S. (2005). Evaluation of the CROPGRO-soybean model for assessing climate impacts on regional soybean yields. T. ASAE 48(6), 2343-2353.

Jiang, Z., Chen, Z., Chen, J., Liu, J., Ren, J., Li, Z., Li, H. (2014). Application of crop model data assimilation with a particle filter for estimating regional winter wheat yields. IEEE J. Sel. Top. Appl. 7(11), 4422-4431.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Kim, S.H. (2016). Random Forests for Global and Regional Crop Yield Predictions. PloS ONE, 11(6), e0156571.

Kryvobok, O. (2000). Estimation of the productivity parameters of wheat crops using high resolution satellite data, Int. Arch. Photogramm. Remote Sens. 33(B7), 717-722.

Launay, M. and Guerif, M. (2005). Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agr. Ecosyst. Environ. 111(1), 321-339.

Lei, L., Stauffer, D.R., Deng, A. (2012). A hybrid nudging-ensemble Kalman filter approach to data assimilation in WRF/DART. Q. J. Roy. Meteor. Soc. 138(669), 2066-2078.

Leon, C.T., Shaw, D.R., Cox, M.S., Abshire, M.J., Ward, B., Wardlaw III, M.C., Watson, C. (2003). Utility of remote sensing in predicting crop and soil characteristics. Precis. Agric. 4(4), 359-384.

Li, F. L., Bao, W. K., Wu, N. (2009). Effects of water stress on growth, dry matter allocation and water-use efficiency of a leguminous species, Sophora davidii. Agroforest. Syst. 77(3), 193-201.

Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., Wang, J. (2014). Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. Ecol. Model. 291, 15–27.

Lilienthal, H., Schnug, E. (2007). New issues for remote sensing in agriculture–a critical overview. Dahlia Greidinger Symposium on Advanced Technologies for Monitoring Nutrient and Water Availability to Plants, 12–13 March. Haifa, Israel, pp 87–104. Accessed at http://gwri-ic.technion.ac.il/pdf/DG/2007/6.pdf accessed 2016 December.

López-Cedrón, F.X., Boote, K.J., Ruíz-Nogueira, B., Sau, F. (2005). Testing CERES-Maize versions to estimate maize production in a cool environment. Eur. J. Agron. 23(1), 89-102.

Machwitz, M., Giustarini, L., C. Bossung, D. Frantz, M. Schlerf, H. Lilienthal, L. Wandera, P. Matgen, L. Hoffmann, T. Udelhoven. (2014). Enhanced biomass prediction by assimilating satellite data into a crop growth model. Environ. Model. Softw. 62(0), 437–453.

Marra, M.C., Piggott, N.E., Goodwin, B. K. (2012). The impact of corn rootworm protected biotechnology traits in the United States. AgBioForum 15, 217-230.

Medeiros, D. B., Silva, E. C. D., Santos, H. R. B., Pacheco, C. M., Musser, R. D. S., Nogueira, R.J.M.C. (2012). Physiological and biochemical responses to drought stress in Barbados cherry. Braz. J. Plant Physiol. 24(3), 181-192.

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

Nafziger, E.D. (2009). Corn. In: E.D. Nafziger, editor, Illinois agronomy handbook: 24th Edition. Department of Crop Science, University of Illinois at Urbana-Champaign. 13–26.

Nguy-Robertson, A.L., Gitelson, A.A., Peng, Y., Viña, A., Arkebauer, T.J., Rundquist, D.C. (2012). Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. Agron. J. 104(5), 1336-1347.

Oteng-Darko, P., Yeboah, S., Addy, S.N. T., Amponsah, S., Danquah, E.O. (2012). Crop modeling: A tool for agricultural research–A review. Journals of Agricultural Research and Development 2(1), 1-6.

Ozdogan, M., Yang, Y., Allez, G., Cervantes, C. (2010). Remote sensing of irrigated agriculture: Opportunities and challenges. Remote Sens. 2(9), 2274-2304.

Paul, C.D., Sophie, M., Paul, W.C., Alan, S. (2003). Crop Yield Assessment from Remote Sensing. Photogramm. Eng. Remote Sens. 69(6), 665–674.

Perie, C. and Ouimet, R. (2008). Organic carbon, organic matter and bulk density relationships in boreal forest soils. Can. J. Soil Sci. 88(3), 315-325.

Rauff, K. O., & Bello, R. (2015). A Review of Crop Growth Simulation Models as Tools for Agricultural Meteorology. Agric. Sci. 6(9), 1098.

Reidsma, P. and Ewert, F. (2008). Regional farm diversity can reduce vulnerability of food production to climate change. Ecol. Soc. 13(1), 38.

Sehgal, V.K. (2013). Remote sensing for crop growth and crop simulation modelling. http://www.iasri.res.in/ebook/GIS_TA/M4_4_RSCGCSM.pdf accessed 2016 July.

Shao, H.B., Chu, L.Y., Jaleel, C.A., Manivannan, P., Panneerselvam, R., Shao, M.A. (2009). Understanding water deficit stress-induced changes in the basic metabolism of higher plants—biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. Crit. Rev. Biotechnol. 29(2), 131–51.

Singh, R. and Helmers, M.J. (2008). Improving Crop Growth Simulation in the Hydrologic Model DRAINMOD to Simulate Corn Yields in Subsurface Drained Landscapes. In 2008 Providence, Rhode Island, June 29–July 2, 2008 (p. 1). American Society of Agricultural and Biological Engineers.

Soler, C., Sentelhas, P., Hoogenboom, G. (2007). Application of the CSM-CERES-Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. Eur. J. Agron. 27, 165-177.

Todorovic, M., Albrizio, R., Zivotic, L., Saab, M.T.A., Stöckle, C., Steduto, P. (2009). Assessment of AquaCrop, CropSyst, and WOFOST models in the simulation of sunflower growth under different water regimes. Agron. J. 101(3), 509-521.

Tsvetsinskaya, E.A., Mearns, L.O., Easterling, W.E. (2001). Investigating the effect of seasonal plant growth and development in three-dimensional atmospheric simulations. Part I: Simulation of surface fluxes over the growing season. J. climate 14(5), 692-709.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sens. Environ. 117, 34-49.

Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J. (2011, October). Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield. In International Conference on Computer and Computing Technologies in Agriculture (pp. 410-418). Springer Berlin Heidelberg.

Xiong, D. (2014). Crop Growth Remote Sensing Monitoring and its Application. Sensors Transducers J. 169(4), 174.

Yuping, M., Shili, W., Li, Z., Yingyu, H., Liwei, Z., Yanbo, H., Futang, W. (2008). Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. Int. J. Appl. Earth Obs. 10(4), 426-437.

Zhao, Y., Chen, S., Sheng, S. (2013). Assimilating remote sensing information with crop model using ensemble Kalman filter for improving LAI monitoring and yield estimation. Ecol. Model. 270, 30–42.

Zhu, X., Zhao, Y., Feng, X. (2013). A methodology for estimating Leaf Area Index by assimilating remote sensing data into crop model based on temporal and spatial knowledge. Chinese Geogr. Sci. 23(5), 550-561.

OVERALL CONCLUSION

This study was developed new and simple models to predict corn yield, and these models was used two approaches based on remote sensing data. One approach was used empirical model which represents the direct relationship between remote sensing data and observed yields, and another approach was assimilated remote sensing data into crop growth model to improve corn yield prediction.

In chapter I, simple approaches to predict corn phenological stages and yields were developed using a minimum MODIS product dataset. Only the red and NIR band surface reflectance data were used to estimate the LAI. Rather than using the reported techniques for filtering/smoothing the LAI data, we fitted the LAI data summed over a cropping season (LAD) to a logistic function for smoothing. A phenology prediction model was established using the MODIS-derived LAD logistic function parameters, and it was used to predict emergence and maturity dates within a reasonable range of error. Simple linear regression models were developed to predict yield using LAD over the predicted period from emergence to maturity as a predictor variable and LAD for a predetermined period from DOY 89 to a particular EOD. Our results indicates that these simple models using LAD as a predictor variable

could predict yields for the two regions of interests with considerable precision and accuracy. The model including information related to phenology exhibited slightly better performance, and could be applied from a fairly early pre-harvest stage of EOD 257. In addition, the model performance showed no difference between the two regions with very different climates and cultivation methods including cultivar and irrigation management. Irrigation practices have been widely adopted in Illinois, USA while rainfed cultivation is a common practice in Heilongjiang, China. The approach described in this paper has potential to be applied to relatively wide agroclimatic regions with different cultivation methods and to be extended to additional crops. However, it needs to be examined further in the tropical and sub-tropical regions which are very different from the two study regions with respect to agroclimatic constraints and agrotechnologies

In chapter II, a simple approach to predict regional corn yield was developed by assimilating MODIS product data into the CERES-Maize model using a minimum input dataset. This method does not require an estimate of the initial conditions and/or parameters of the CERES-Maize model. A minimum input dataset comprising planting date, fertilizer amount, genetic coefficients, soil, and weather was used to simulate corn growth and yield using CERES-Maize model. Planting date, corn maturity group, daily LAI, and daily water stress factors estimated using the MODIS-derived LAD

logistic function were directly assimilated into the CERES-Maize model to predict regional corn yield in Illinois, USA. The corn yield predictions using only estimated planting date and maturity group performed very poorly under the rain-fed condition at both the AD and state levels, whereas corn yield prediction performance improved by simulation under the auto-irrigation condition. Moreover, adding the daily LAI and water stress factors into the MODIS-derived LAD logistic function further improved corn yield prediction performance. In addition, earlier corn yield prediction at DOY 257 was possible without degrading accuracy. This simple approach was successful for predicting regional corn yield with considerable accuracy and precision in Illinois, USA. However, this method needs to be examined in regions with more diverse agro-climatic and agro-technology conditions.

In conclusion, new and simple corn yield prediction models for two approaches were developed based on remote sensing data, and had considerable accuracy and precision for study regions. However, these models and method must be examined for spatial portability in more diverse agro-climatic and agro- technology regions.

ABSTRACT IN KOREAN

Moderate Resolution Imaging Spectroradiometer (MODIS) 자료와 작물 생육 모델을 이용한 지역단위 옥수수 수량 예측

반호영
작물생명과학전공
식물생산과학부
서울대학교 농업생명과학대학

표본 조사를 하여 작물의 수량을 예측하는 데에는 상당한 비용과 노동력이 요구된다. 하지만, 원격 탐사 자료는 최소한의 비용으로 작물의 수량을 신뢰성 있게 예측하는데 도움을 줄 수 있으며, 또한 시기 적절하게 작물의 생육 상태를 감시하거나 얻는데 도움을 줄 수 있을 것이다. 작물의 생육과 수량을 예측하는데 원격 탐사 자료를 이용하는 두 가지 접근법이 있다. 첫 번째 접근법은 원격 탐사 자료와 관측 수량과의 직접적인 관계를 나타내는 경험적 모델을 사용하며, 다른 접근법은 옥수수 수량

예측력을 높이기 위하여 원격 탐사 자료를 작물 생육 모델에 동화하는 방법이다. 본 연구에서는 최소한의 데이터세트를 이용하여 각 방법에 대하여 지역단위 옥수수 수량을 예측하는 간단한 모델을 개발하고, 그 모델들의 지역단위 수량 예측력을 평가하는 것이다.

지리적으로 분리된 주요 옥수수 생산 지역인 미국의 일리노이주와 중국의 흑룡강성 지역의 옥수수 수량을 예측하는 간단한 모델을 MODIS 자료를 이용하여 개발하였으며, 일리노이주의 옥수수 수량과 페놀로지 자료는 농업지구 단위로 2000년부터 2013년까지 수집하였고, 흑룡강성의 옥수수 수량 자료는 현 단위로 2002년부터 2012년까지 수집하였다. 주/성 단위 수량 예측 모델을 검증하기 위하여 3개년을 선택하였으며, 농업지구 단위 수량 및 페놀로지 예측 모델을 개선 및 검증하기 위하여 나머지 년도에서 각각 70% 와 30%의 자료를 이용하였다. 엽면적지수를 계산하기 위하여 8일 간격의 지표 반사 자료인 MOD09A1 자료를 3월 29일 (day of year, DOY 89)부터 12월 2일 (DOY 337)까지 수집하였으며, 시즌 초기부터 주어진 날짜 [End of DOY (EOD)]까지의 엽면적지수의 합은 로지스틱함수로 잘 표현되었고, 엽면적기간의 시즌 변화를 잘 나타내었다. 출아와 성숙날짜를 예측하는 단순 페놀로지 예측 모델을 엽면적지수 증가 속도와 최대 엽면적지수의 날짜를 나타내는 로지스틱함수의

파라미터인 b_1과 b_2를 이용하여 개발하였으며, 페놀로지 예측 모델은 검증 데이터세트에 출아와 성숙 날짜를 각각 6.3과 4.9 일의 root mean square error (RMSE)로 잘 예측하였다. 옥수수 수량 예측을 위한 두 개의 단순 선형 회귀 모델들 (Y_P 와 Y_F) 을 엽면적기간을 이용하여 설정하였다; Y_P 모델은 출아부터 성숙 날짜까지의 엽면적기간을 이용하였고, Y_F 모델은 3월 28일 (DOY 89)부터 특정 EOD까지의 지정된 엽면적기간을 이용하였다. 옥수수 수확이 거의 끝나는 12월 2일 (EOD 321)에 예측된 주/성 단위 옥수수 수량의 RMSE가 일리노이주 0.68 t/ha와 흑룡강성 0.66 t/ha로 Y_P 모델이 Y_F 모델보다 훨씬 나은 예측력을 보였으며, Y_P 모델은 9월 13일 (DOY 257)에 아주 이른 옥수수 수량 예측에 대해서도 비슷하거나 더 좋은 예측력을 보였다. 추가로, 모델의 예측력은 기후와 품종과 관개를 포함한 재배 방법들이 매우 다른 두 연구 지역간 차이가 없었다.

작물 생육 모델과 원격탐사 자료는 작물 생육과 수량 예측에 유용한 도구이지만, 각각은 지역단위 작물 생육과 수량을 예측하는데 불가분의 문제점을 가지고 있다. 지역단위 옥수수 수량 예측의 정확도와 정밀도를 향상시키기 위하여 MODIS자료를 작물 생육 모델에 동화시키는 간단한 방법을 개발하였고, 주요 옥수수 생산 지역인 미국의 일리노이주에 대하여 지역단위 수량의 예측력을 평가하였다. 옥수수 수량과 페놀로지 자료는 농업지구와

주 단위로 2000년부터 2013년까지 수집하였으며, CERES-Maize 모델을 이용하여 파종일, 비료 시비량, 유전 계수들, 토양 및 기상자료로 구성된 최소한의 입력 데이터세트로 옥수수의 생육과 수량을 모의하였다. 각 격자의 파종일은 MODIS에서 파생된 엽면적기간의 시즌 변화를 나타내는 leaf area duration (LAD) logistic 함수를 이용한 페놀로지 예측 모델로 추정하였으며, 옥수수 품종의 유전 계수들은 LAD logistic 함수에서 파생된 최대 엽면적지수와 CERES-Maize 모델로 모의된 최대 엽면적지수 사이의 차이가 최소가 되는 성숙군 (Decision Support System for Agrotechnology Transfer (DSSAT) 4.6에 포함된)의 유전계수들로 결정하였다. 추가로, CERES-Maize 모델에 내재된 일별 수분 스트레스 요소들은 LAD logistic 함수로부터 추정된 일별 엽면적/중 생장 속도와 CERES-Maize 모델을 자동-관개상태로 모의하여 추정된 일별 엽면적 생장 속도의 비율로 계산하였다. 추정된 파종일과 성숙군만을 이용한 옥수수 수량의 예측력은 자연관수상태에서 농업지구와 주 단위 모두 매우 낮았다. 반면에 자동-관개상태에서 모의된 옥수수 수량의 예측력은 훨씬 향상되었으며, 이것은 일리노이주의 옥수수 농장에 상당한 비율로 관개가 적용되고 있는 것을 나타낸다. 일별 엽면적지수와 수분 스트레스 요소들을 추가로 동화함으로써, 옥수수 수량의 예측력은 또한 상당히 향상되었으며, 결정계수 (R^2)가 0.72에서 0.78로

증가하였고, RMSE가 1.47 t/ha 에서 0.75 t/ha 로 감소하였다. 추가로, 9월 13일 (DOY 257)에 이른 옥수수 수량 예측에도 정확도 감소 없이 예측이 가능하다.

본 연구는 원격탐사 자료를 이용하여 각 접근법에 대한 간단한 모델을 개발하였고, 연구 지역들에 옥수수 수량 예측력은 상당한 정확도와 정밀도를 보여주었다. 그러나 이 모델들은 다른 농업기상과 기술을 가진 지역들에 대하여 수량 예측력을 평가해야 할 것이다.

주요어: MODIS; 옥수수 수량; 페놀로지; 엽면적기간; 로지스틱함수; 작물 생육 모델; 수분 스트레스