2007 International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Special Publications (In Press)

OPERATIONAL PREDICTION OF CROP YIELDS USING MODIS DATA AND PRODUCTS

Paul C. Doraiswamy[a] , Bakhyt Akhmedov[b] , Larry Beard[c] , Alan Stern[a] and Richard Mueller[c]

[a]USDA, ARS, Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705
(paul.doraiswamy, alan.stern) @ARS.USDA.GOV
[b]Science Systems and Associates, Inc. Lanham, MD 20704; - Bakhyt.akhmedov@ARS.USDA.GOV
[c]USDA, NASS, Research and Development Division, 3251 Old Lee Highway, Fairfax, VA 22030-1504 –
(larry_beard, Rick_mueller)@nass.usda.gov

Commission IV, WG VI

KEY WORDS: Remote Sensing, Agriculture, Crop yield, MODIS algorithm, Operational, Crop Classification

ABSTRACT:

Official crop progress, condition and production estimates for the United States are responsibilities of the U.S. Department of Agriculture's, National Agricultural Statistics Service (NASS). In addition to weekly and monthly survey-based data, biweekly composite maps of the normalized difference vegetation index (NDVI) from the NOAA AVHRR sensor (1 km resolution) are produced by NASS's Research and Development Division (RDD) for monitoring vegetative change. This provides a qualitative assessment of differences in crop condition that may be an indication of potential yields. There is need for a more quantitative assessment of crop yields and spatial variability. Currently, NASS acquires crop yield indications via ground-based sample surveys (objective plant and fruit counts, fruit weights and farmer reports) which are collectively used to develop tools for its decision support system to assess weekly crop progress, monthly crop yield estimates for each state and the U.S, and annual county yield estimates. This paper describes the joint research between RDD and the Agricultural Research Service (ARS) of USDA for the development of simplified process models and algorithms to supplement the NASS field data collection. Potential advantages to using remote sensing include integration of spatial variability into county yields, enhanced timeliness, and efficient use of resources.

In the preliminary phase, MODIS data and products for the states of Iowa and Illinois were used to develop an operational assessment of crop yield forecasts for corn and soybeans. Spatial estimates of crop yields at county and sub-county levels offer a major improvement of current capabilities. The timeliness in producing these estimates is a vast improvement over the present assessment capability at the county level. Potential use of the estimates will supplement current tools and improve NASS' crop condition and yield decisions. Results of the pre-harvest forecasts developed for the 2005 and 2006 crop seasons are presented.

1. INTRODUCTION

Accurate and timely monitoring of agricultural crop conditions and estimating potential crop yields are essential processes for operational programs. Assessment of particularly decreased production caused by a natural disaster, such as drought or pest infestation, can be critical for countries or locales where the economy is dependent on the crop harvest. Early assessment of yield reductions could avert a disastrous situation and help in strategic planning to meet demands. The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) monitors crop conditions and makes the Official USDA production assessments in the U.S., providing monthly production forecasts and end-of-year estimates of crop yield and production. NASS has developed methods to assess crop growth and development from several sources of information, including several types of surveys of farm operators and field-level measurements. Field offices in each state are responsible for monitoring the progress and health of the crop and integrating crop condition with local weather information. Information on crop condition and progress is also distributed in a biweekly report on regional weather conditions. NASS offices provide monthly information to the Agriculture Statistics Board, which assesses the potential yields of all commodities based on crop condition information acquired from different sources. This research complements efforts to independently assess crop condition at the county and state levels. The timely evaluation of potential yields is increasingly important because of the huge economic impact of agricultural products on world markets and strategic planning.

County statistics are noted as a driving force for rural economic development, and are essential to proper management of USDA's many farm, education, and natural resources management programs. Many allocations of federal resources to states and counties are determined by their production of farm commodities. Demand for accurate commodity estimates at the lowest level of aggregation, and at the earliest possible time, has and continues to increase substantially. Literally millions of business decisions rely on this basic production data produced by USDA/NASS.

In the early 1960s, NASS initiated "objective yield" surveys for crops such as corn, soybeans, wheat, and cotton in States with the greatest acreages (Allen et al., 1994). These surveys establish small sample units in randomly selected fields which are visited monthly to determine maturity, numbers of plants, numbers of fruits (wheat heads, corn ears, soybean pods, etc.), and weight per fruit. Yield forecasting models are based on relationships of samples of the same maturity stage in comparable months during the past four years in each State. These indications are then compared to farmer-based survey results to produce monthly yield forecasts. Additionally, the Agency implemented a midyear Area Frame Survey that enabled creation of probabilistic based acreage estimates. For major crops, sampling errors are as low as 1 percent at the U.S.

level and 2 to 3 percent in the largest producing States. Accurate crop production forecasts require accurate estimates of acreage at harvest, its geographic distribution, and the associated crop yield determined by local growing conditions. There can be significant year-to-year variability which requires a systematic monitoring capability. To quantify the complex effects of environment, soils, and management practices, both yield and acreage must be assessed. A yield forecast within homogeneous soil type, land use, crop variety, and climate preclude the necessity for use of a complex forecast model.

Doraiswamy, et al. (1979), provided an inventory of various crop yield models, including statistical and deterministic models. The performance of deterministic models for large area forecasts depended on the availability of local climatic data with adequate spatial resolution. Use of remotely sensed data was limited to studying the temporal changes in vegetation condition such as crop growth and development.

Integration of remotely sensed data in crop yield models evolved during the next decade of research based on field experiments and advances in biophysical modeling. The need for linking real-time remotely sensed data initiated research on retrieval of biophysical parameters from satellite imagery.

The potential application of remote sensing technology for monitoring crop condition and predicting crop yields at regional scales have been studied extensively during the past several decades. Traditionally, the empirical approach based on statistical regression models was used with lower resolution data such as that from NOAA AVHRR. Studies have shown that the cumulative seasonal normalized difference vegetation index (NDVI) values were significantly correlated with reported crop yields (Groten, 1993). Doraiswamy and Cook (1995) demonstrated that cumulative NDVI values for spring wheat during the grainfill period improved estimates of crop yields in North Dakota. Combining a growth model with input parameters derived from remotely sensed data provides spatial integrity as well as a real-time ‘‘calibration’’ of model parameters. Moulin, et al., (1995) successfully used this approach using SPOT/HRV data in a radiative transfer model SAIL (Scattering by Arbitrary Inclined Leaves, Verhoef,1984) to map crop leaf area index (LAI) and also inversion models to that simulated canopy reflectance that were comparable with SPOT/HRS reflectance. These methods enabled constraining the model parameters with satellite observations to retrieve key parameters of soil moisture and above ground plant biomass. Doraiswamy et al. (2003, 2004, 2005) simulated LAI from NOAA AVHRR (1 km) two-band reflectance as well as MODIS (250 m) 2-band reflectance products, and used them in crop yield simulation models over small areas. The MODIS derived LAI seasonal profile was developed in the above studies using the SAIL model, and the crop yield model parameters were initialized. The MODIS 8-day composite reflectance product was acquired through the NASA–DAAC EROS Data Center, Sioux Falls, SD. The MODIS data has already been corrected for atmospheric effects and cloud cover.

The limitation in using satellite-based LAI methods with high (Landsat) and moderate (MODIS) spatial resolution data in crop yield simulation models at regional scales is that the canopy architectural parameters in the SAIL or other radiative transfer model for crop-specific LAI should be accurate. These parameters also change during the growing season. The reflectance data are used to derive LAI seasonal profile, which is used in initializing or constraining parameters in the crop yield simulation model. The temporal and spatial inconsistencies of the MODIS 8-day reflectance product data (Doraiswamy, 2006) limits its application in crop yield models at regional scales.

In this research we evaluated the use of the MODIS NDVI and surface temperature products to develop a multi-dimensional regression algorithm to predict the state and county level yields. The NDVI seasonal dynamics is representative of crop growth and biomass changes and thermal data is representative of the crop moisture stress condition.

The objectives of this research are to: a) develop a MODIS-based algorithm for operational classifications of corn and soybean crops in the U.S. Corn Belt; b) develop a multi-dimensional regression method to provide a consistent, timely and accurate yield prediction for potential use in NASS’s operational program.

METHODOLOGY

Study Region.

The study areas presented in this paper are the states of Iowa and Illinois, which cover a major part of the U.S. Corn Belt. The two States are intensively cultivated, with approximately 75% of the land in corn and soybean crops. Crops are grown under rain fed conditions where soil moisture is normally adequate in the growing season; however, moisture stress conditions can occur in the early stages of crop development and more often during the latter part of the season. Seasonal rainfall ranged between 800 mm in the north to 450 mm in the southern part of the region. Soil moisture is generally at field capacity at the start of the season. However, because of spatial variability in the spring rainfall, planting dates across the region were variable. Crop planting in the region was completed by mid-May, with corn generally planted about 2 weeks earlier than soybeans. Crop maturity occurred by late September.

Data and Re-processing

The MODIS 8-day composite surface reflectance product (MOD09Q1 collection 4) at 250 m resolution and land surface temperature product (MOD11A2) at 1 km resolution was acquired through the NASA-DAAC EROS Data Center, Sioux Fall, SD. Imagery data for the available 5-year period (2002-2006) was acquired for day of year (DOY) from 121 to 305. The images were re-projected from ISIN projection to the standard UTM projection for zone 15 (Iowa) and 16 (Illinois) using the MODIS re-projection tool. The data was imported to ERDAS Imagine format, and NDVI time series for the crop season were calculated.

The cloud free MODIS data product has potentially three sources of errors – georeference, atmospheric influence and BRDF. The 8-day composite images always have patches even on homogeneous areas which are caused by difference in single day images used in the composites. Besides cloud cover, errors in reflectance caused by atmospheric condition always reduced the NDVI value. To eliminate these errors we used multiple-step processing of the composite data using the Savitzky-Golay filtering technique adapted for tracking the upper envelope of the NDVI time series profile (Jonsson and Eklundh, 2004; Doraiswamy et al. 2006). The Savitzky-Golay filter uses moving 5-point window for each pixel time series profile and in each window, noisy values is approximated by polynomial to smooth NDVI values in the window. The thermal data is screened using the quality assurance data provided along with the thermal imagery, taking data with errors <= 2 degrees Kelvin.

Crop Classification

Landcover classification is an important step to assure accurate retrieval of crop specific data to monitor crop condition and predict yields. Classification has traditionally been completed at 30-m resolution using Landsat ETM+ images. In the past decade, USDA-NASS used Landsat data to develop crop classification for crop acreage estimation over selected states including Iowa and Illinois. However, the classification is usually not available until about 4-5 months after the crops are harvested. In an operational program where remote sensing data is used to predict crop yields, it is critical to have timely crop classification for assessing crop specific yields at specific time periods. Additionally, the uncertainties in the availability of Landsat data required the development of crop classification using the MODIS 8-day composite time-series data. A decision tree algorithm was developed (Doraiswamy et al, 2007) to map corn and soybean fields. In a two step process the crop area in the state is first selected using a threshold of NDVI values based on the combined crop phenology of corn and soybean crops. The next step of the decision tree algorithm separates the corn and soybean crops within the crop area. Figure 1 is the general NDVI time-series profile showing differences between corn and soybeans. The soybean crop is planted several weeks after corn and the maturity follows that of corn. The clear distinction between the corn and soybean NDVI profile occurs around day of year (DOY) 177 in Iowa and Illinois. These features are used to separate the corn and soybean crops, the predominant crops in these two states. The Landsat classification for Iowa and Illinois developed by the NASS Research and Development Division was used as the template for evaluating the accuracy of the MODIS-based classification.

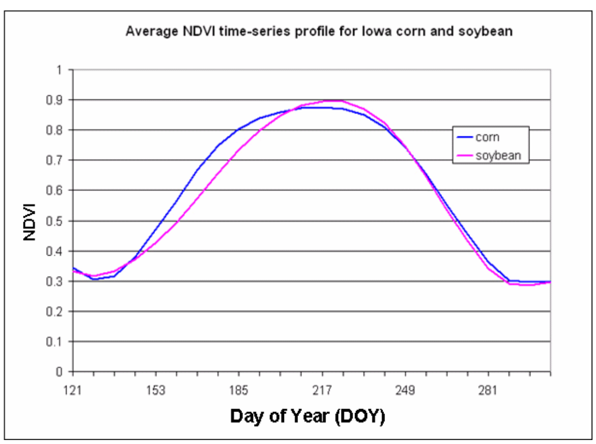

Figure 1. Example of the normalized difference vegetative index (NDVI) for corn and soybean crops using MODIS 8-day composite data (250 m) for the 2005 crop season after the data was processed with the filtering algorithm.

Crop Yield Algorithm

The NDVI and surface temperature (Ts) from MODIS imagery are parameters that are correlated with crop yields. Initial investigations suggest that these two parameters can provide spatial variability of crop growth conditions during the growing season. MODIS data shows better correlation with crop yields compared to similar analyses using data from NOAA AVHRR. The better results may be due to better spatial resolution and to the specific narrow bands in VIS and NIR of the MODIS sensors. The seasonal NDVI profile describes the crop growth and development, surface temperature provides additional information regarding potential crop stress conditions. Stress conditions may include crop water stress as well as disease and infestation. An algorithm that combines NDVI and Ts was developed that correlates with yield in a two dimensional regression equation:

$$Yield=a+b*NDVI+c*Ts \qquad (1)$$

Currently, one set of equations are developed for corn and soybeans for each state. The NDVI and Ts parameters are summed for a period between the mid-vegetative to mid-senescence period. The state yield estimate is obtained by averaging the parameters for all corn or soybean pixels in the state. The NDVI and Ts parameters extracted for each crop for the 2002-2005 crop season are then used in equation (1) with the NASS estimated yields. The coefficients **a**, **b** and **c** are derived and the regression equation is then used as a predictor for 2006 state crop yields.

The spatial variability of crop yields are assessed at the county level by extracting the mean NDVI and Ts parameters from corn and soybean pixels for each county. Equation (1) is used to determine the a, b and c coefficients using data from the 2002 crop season. Then equation (1) was applied to predict county level yields for successive years. The 2002 crop season data was selected to develop the regression algorithm because the MODIS data quality was better than other years and crop yields for corn and soybeans were spread over a wider range. County yields for the 2005 and 2006 crop seasons were predicted using the regression algorithm developed from the 2002 data sets. The initial predictions for state and county level yields are made in early September prior to crop harvest and updated after the senescence is completed in October.

RESULTS and DISCUSSIONS

The MODIS data acquired for this research covered five crop seasons when complete seasonal data was available (2002 - 2006). The results from this research are compared with USDA official estimates from NASS. Final state yield estimates are published in January of the following year. These estimates are based on objective yield and farmer-based survey data collected from well planned sampling strategies and considered to be accurate. The NASS county level estimates are generally published in March of the following year and the data used are from similar farmer-based surveys and from observed local weather and crop conditions. These estimates may not be systematic from county-to-county within the state. This research seeks to provide additional information derived from remote sensing data to supplement spatial information that would strengthen the county level estimates., at a much earlier time period.

Crop Classification

The MODIS-based classification of corn and soybean crops was developed using the NDVI time series profile, first separating the crop area from other classes by selecting pixels with an NDVI threshold of 0.4 at the beginning and end of the crop season with a value of 0.8 at mid-season. In a narrow window around DOY 177, the magnitude of NDVI for corn is greater than soybeans as shown in an example in Figure 1. The MODIS classification was compared with a NASS classification developed from Landsat ETM images for the 2005 crop season. Landsat 30 m pixels were aggregated to 250 m MODIS pixels by picking up only those pixels that contained 90% of specific (corn or soybean) crop. In Iowa 357,795 MODIS pixels had 90% of corn or soybean fields (Figure 2). These pixels were used to test the corn and soybean discrimination algorithm. The results were then compared with the Landsat classification aggregated to 250m pixel resolution. The overall accuracy was 81.7% with a kappa coefficient of 0.63. In Illinois the number of MODIS pixels with 90% crop was found to be 409,108 pixels. The comparison with Landsat classification showed that the overall classification accuracy was 75.1% with a kappa coefficient of 0.50. The accuracy for Iowa was better than Illinois perhaps because of the large variation in corn and soybean phenological stages in Illinois between the northern and southern areas of the state.

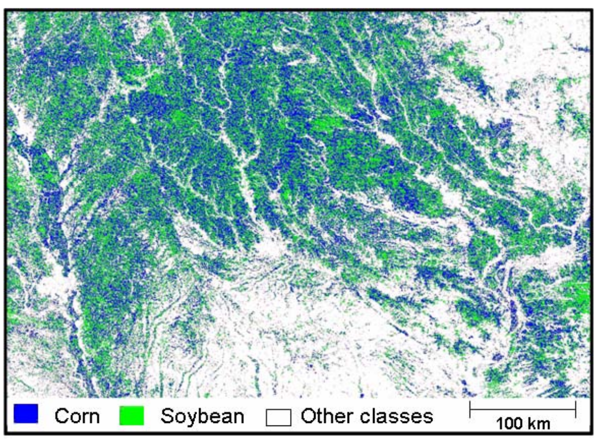

Figure 2. The 2005 classification for Iowa using the 8-day MODIS 250 m composite imagery.

Crop Yield

Using the USDA/NASS state yield estimates for the 2002-2005 crop seasons, the multi-dimensional regression equations were developed for corn and soybean yields for Iowa and Illinois. Based on the regression algorithm, the coefficients of determination for Iowa at the state level were 0.98 and 0.88 respectively for corn and soybean crops. These equations were used to predict the Iowa state yield estimates for 2006 corn and soybeans. The regression algorithm yield predictions for the 2006 season for corn was 170 bushels per acre (b/ac) and 52.2 b/ac for soybeans versus NASS's estimates of 166 b/ac and 50.5 b/ac respectively. In Illinois, the regression algorithm was developed for the same period (2002-2005), and showed coefficients of determination of 0.98 and 0.36 respectively for corn and soybeans. The low coefficient of determination for soybeans is due to disease outbreak in parts of Illinois during the latter part of the crop season in 2003 (Malvick, 2003). This occurred during the latter part of the season and the yellowing of the foliage due to disease was not observed from imagery. Nevertheless the yield predictions from the regression algorithm for 2006 were 165.2 b/ac and 47.8 b/ac versus NASS's estimate of 163 b/ac and 48 b/ac respectively for corn and soybeans.

The state level predictions for the 2007 crop season will be based on a 5-year multi-regression algorithm developed from 2002-2006 state estimates. The coefficients of determination for Iowa corn and soybeans were 0.96 and 0.88, and for Illinois they were 0.98 and 0.438, respectively. Eliminating the 2003 year from the regression algorithm for soybeans in Illinois increases the coefficient of determination to 0.97.

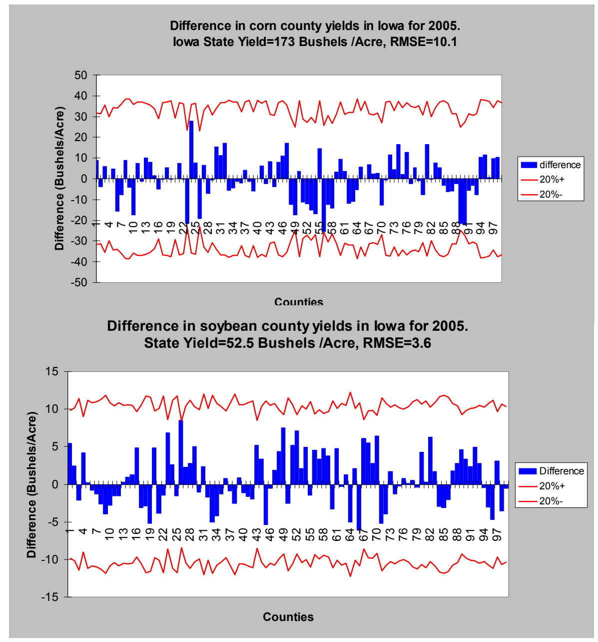

Figure 3. Difference in county yields between the regression algorithm predictions and NASS estimated yields for the 99 counties in Iowa for the 2005 crop season.

The county level yield predictions were completed at the same time as the state predictions. The regression algorithm developed from the 2002 data provided the coefficients for the multidimensional regression algorithm. The multidimensional coefficients of determination for Iowa were 0.86 and 0.71 and for Illinois they were 0.82 and 0.86 respectively for corn and soybeans. These relatively high values indicate that the MODIS data for 2002 were good and the predictions fit well with the NASS crop yield estimates. For the other crop seasons the coefficient of determination was lower. County level predictions are more sensitive to the quality of the remote sensing data, combined with the accuracy and consistency of the estimation of NASS county yields.

Using the regression model developed from 2002 crop season, the corn and soybean crop yields for 2005 were predicted and results compared with the NASS estimates. The coefficient of determination for corn is 0.73 and for soybeans is 0.59, and for Illinois these coefficients are 0.60 and 0.52 respectively. The regression coefficients are strongly dependent on the quality of satellite imagery and NASS county yield estimates for that particular year. Figure 3 shows the difference in county yields between the NASS estimate and the regression algorithm predicted yields for Iowa 2005 crop season. The red lines are the +/- 20% standard deviation from the NASS estimated yields. The regression algorithm predicted yields are all within 20% of the NASS estimates, and the great majority are within 10percent. RMSE for predicted yield for corn and soybeans in Iowa are 10.1b/ac and 3.6 b/ac and in Illinois are 19.3 b/ac and 5.6 b/ac respectively.

CONCLUSION

Timely and accurate prediction of crop yields is critical for agricultural markets, planning and development. Daily frequency of MODIS data acquisition at 250 m pixel resolution offers a great potential for use of the data and products in operational yield prediction programs. In this study, a simple algorithm that uses near-real time MODIS imagery and products was developed to predict crop yields at county and state levels. The algorithm includes crop-specific classification and yield prediction prior to crop harvest. The crop classification was developed using a decision tree algorithm that relied on the characteristics of crop growth phenology without the need for ground-based data. The classification accuracies were compared with the USDA NASS Landsat-based classification data and found to be acceptable for yield predictions. The correlation between NDVI and crop yields and between surface temperature and crop yields are integrated in a multidimensional regression model for predicting yields at the county and state levels. Differences between the NASS state level yield estimates and the regression algorithm predictions for both Iowa and Illinois for the 2006 season was less than 4 b/ac for corn and less than 2 b/ac for soybeans.

The quality of MODIS data is very critical for crop yield predictions and this paper describes some of the steps that we achieved to enhance the quality of data for cloud cover and atmospheric effects. The computational scale appeared to make a difference in the tolerance on the imagery data quality. Although the same algorithm was used for both state and county level yield predictions, the county yield predictions appeared to be more sensitive to quality of the images and the yield predictions were not as well correlated with the NASS estimated yields. Another important factor in this lower coefficient of determinations at the county level was that the NASS estimates have an error that is not reported. However, assuming an error in the NASS county yield estimates, the predictions are well within a 20% standard deviation of the estimates.

REFERENCES

Allen, R., G.A. Hanuschak, and M.E. Craig, 1994. Forecasting crop acreages and yield in the face of and in spite of floods, Crop Yield Forecasting Methods, Proceedings of the Seminar Villefranche- sur-Mer, 24–27 October, Villefranche-sur-Mer, France, pp. 87–110.

Doraiswamy, P.C., T. Hodges, and D.E. Phinney, 1979. Crop Yield Literature for AgRISTARS Crops Corn, Soybeans, Wheat, Barley, Sorghum, Rice, Cotton and Sunflowers, AgRISTARS Technical Report SR-L9-00405, Lockheed Electronics Company, Inc., Houston, Texas, 105 p.

Doraiswamy, P.C. and P.W. Cook. 1995. Spring Wheat Yield assessment using NOAA AVHRR data. Canadian J Remote Sens. 21:43-51.

Doraiswamy, P.C., P.Zara, and A. Stern. 2000. Satellite remotely sensed data application in estimating crop condition and yields. In – Remote Sensing Applications. Edited by M.S. Srinivas. Narosa Publishing House, New Delhi. Chapter 24: pp 229-240.

Doraiswamy, P.C., S. Moulin, P.W. Cook, and A. Stern. 2003. Crop yield assessment from remote sensing, Photogrammetric Engineering and Remote Sensing, 69, 665– 674.

Doraiswamy, P.C., J.L. Hatfield, T.J. Jackson, J.H., B. Akhmedov, and A.J. Stern. 2004. Crop condition and yield simulations using Landsat and MODIS imagery, Remote Sensing of Environment, 92: 548– 559.

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

Doraiswamy, P.C., B. Akhmedov and A.J. Stern. 2006a. Improved techniques for crop classification using MODIS Imagery. Proceedings of the International Geoscience and Remote Sensing Symposium, July 31 – August 4, 2006, Denver, Colorado. CD ROM.

Doraiswamy, P.C., B. Akhmedov and A.J. Stern. 2006b. MODIS time Series data applications in Agriculture. A CEOS Land Product Validation topical workshop: Validation of global vegetation indices and their time series. University of Montana Missoula, MT. August 7, 2006. (Abstract) http://www.ntsg.umt.edu/VEGMTG/val-home.html

Doraiswamy, P.C. and B. Akhmedov. 2007. Crop specific classification for yield predictions using MODIS imagery. (Submitted for publication).

Groten, S.M.E. 1993. NDVI- crop monitoring and early yield assessment of Brukina Faso. International Journal of Remote Sensing 14, 1495-1515.

Jonsson, P. and Eklundh, L., 2004. TIMESAT—A program for analyzing time-series of satellite sensor data, Computer & Geosciences. 30, 833– 845.

Malvick, D. 2003. Charcoal Rot of Soybeans in Illinois: Primary or Secondary Disease? Bulletin of Pest Management and Crop Development information for Illinois. University of Illinois at Urbana- Champaign. University of Illinois Extension.http://cptc.ipm.uiuc.edu/bulletin/pastpest/articles/200323g.html

Moulin, S., Fisher, A., Dedieu, G and Delcolle, R. 1995. Temporal variation in satellite reflectances at field and regional scales compared with values simulated by linking crop growth and SAIL models. Remote Sensing of Environment 54, 261-272.

Verhoef, V. 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. Remote Sensing of Environment. 16, 125-145.

ACKNOWLEDGEMENTS

Funding for this research was provided in part by the National Agricultural Statistics Service and the Agricultural Research Service of the U.S. Department of Agriculture. MODIS data was provided free of charge from the NASA–DAAC EROS Data Center, Sioux Falls, SD.