

CONSIDERATION AND COMPARISON OF DIFFERENT REMOTE SENSING INPUTS FOR REGIONAL CROP YIELD PREDICTION MODEL

Preeti Mali
Charles O' Hara
Valentine Anantharaj
Geo Resources Institute
Mississippi State University
#2 Research Boulevard
Starkville, MS 39759
preeti@gri.msstate.edu
cgohara@gri.msstate.edu
val@gri.msstate.edu

ABSTRACT

Regional crop yield prediction methods can be enhanced by the use of remote sensing based inputs to obtain an efficient and timely prediction capability. Inputs from remote sensing usually include vegetation indices and climatic information such as temperature, precipitation, solar radiation etc. The crop selected for this study is soybean. This study focuses on investigating and comparing a combination of satellite sensor characteristics and data products derived from satellite data stream inputs, with crop modeling input data requirements. The factors to be considered include the spatial, spectral and temporal characteristics of sensor characteristics and derived data products to determine objective methods for selecting model inputs that offer the most promise to improve regional soybean yield prediction.

INTRODUCTION

Crop models have been used for predicting crop yield before harvest. The benefits of such predictions have potential effects from local to regional to global. Such predictions warn the decision makers about potential reductions in crop yields and allow timely import and export decisions. These pre-harvest crop yield estimations also help in regional and global crop prices and trade policies. Thus, reliable yield prediction methods are highly important for national and global food security.

Remote sensing is used in a number of crop prediction models. These range from simple regression-based models to very complicated models based on a number of inputs. Although remote sensing is beneficial for local uses such as precision agriculture, remote sensing is increasingly being used in regional predictions due to the ability to efficiently provide spatially based results for larger areas. Most regional level yield prediction methods consider using Moderate Resolution Imaging Spectro-radiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) due to their wide swath width. Since crop yield models are usually developed from field-based experiments, regional prediction models based on remote sensing are usually adapted from crop models developed from field-level experimentations. Therefore, in order to obtain as much accuracy as possible in predicting yield, the spatially-based input variables should be able to represent field level conditions as much as possible.

Different types of models for field-level crop yield predictions for various crop types are available. The adaptation of crop models to regional-level predictions of yield are lacking validated mechanisms for their application. The major challenge in such adaptation lies in the area of scaling or substituting model inputs to obtain representative estimates that extend capabilities to a regional or national level. In scaling models to regional-level analysis, field-level conditions such as row spacing, amount of fertilizer per plot, and other field-level details cannot be used. Difficulties arise when field-specific input variables to the models are replaced by information extracted from satellite image based observations. The probability that the inaccuracy of the model output would increase cannot be neglected because on one hand a field level based model is being used for regional level estimates, and on the other hand the input parameters are estimated from remote sensing. Even then, it may not be a deterrent for using such crop models for regional predictions using remote sensing based inputs. In fact researchers agree that remote-sensing technologies can help to reduce the costs, time, and money to effectively predict crop yield (Reynolds *et al.*, 2000; Wiegand *et al.*, 1991).

Doraiswamy *et al.* (2005) used remote sensing inputs in a crop model for regional yield assessment and found that use of information from remote sensing observation may effectively be integrated into crop modeling methodologies. Doraiswamy's findings indicate that with the selection of an appropriate crop model and careful application of input information derived from satellite-based observations, regional crop yield assessments using remote sensing can be highly beneficial. Again, for regional-level predictions, the use of satellite based inputs highly simplifies the process considering the amount of time and labor that regional level data collection requires. This paper is based on the assumption that the use of correct model and remote sensing based data sets would help in obtaining desired estimated yield accuracies and to provide a basis whereby regional soybean yield prediction methodology based on remote sensing can be validated.

Some studies have been done using remote sensing based regression models for regional or country wide predictions (Rasmussen, 1998; Dabrawoska *et al.*, 2002). However, the regression based models are highly variable and do not consistently provide adequate accuracy for larger areas since they are empirical in nature (Moulin *et al.*, 1998). The variables used in the regression models to increase the accuracy (or the R-square) cannot be of global application as such variables differ from region to region. Other methods used by some researchers are based on application of Montieth-based models that employ Photosynthetically Active Radiation (PAR) and Absorbed Photosynthetically Active Radiation (APAR) parameters to estimate yield (Bastiaanssen and Ali, 2003; Lobell *et al.*, 2003). These methods do not use meteorological inputs and are mostly based upon the ability of plants to utilize the solar radiation for photosynthesis.

The variables used in these methods (regression and montieith based) are derived from field-level observations and parameters without regional applicability; therefore, these models are not appropriate for and have not been adapted to generalized regional application. Hence, the best crop prediction models for regional application which may be adapted to ingesting information from remote-sensing based observations are mechanistic models that relate physiological growth stage to environmental variables to obtain the model output. Moulin *et al.* (1998) correctly stated that, "more mechanistic and physiologically sound models are necessary to assimilate remote sensing data and to predict production of major crops". However, some mechanistic models also require very detailed field level information that cannot be estimated for regional level modeling purposes.

In this light, Sinclair model holds promise as it is a mechanistic model but the inputs required for running the model are very simple (Sinclair, 1986). Since, this model was developed from field-level experimentations, it becomes highly important that the remotely-sensed data sources used to derive information inputs to the model such as temperature, precipitation, and vegetation indices provide accurate and regionally representative measurements that can be efficiently modeled in a spatial environment.

In short, regional-level crop monitoring as well as yield predictions for soybean can be made operational by using remote sensing inputs and using a GIS framework for analysis. However, the processes are usually too complex to exactly replace the input variable with the remote sensing based parameters, and the results that are obtained may not be as accurate as a field-level simulation.

This concept paper considers the different inputs required by the Sinclair model and explores the application of information from remote sensing observations and other spatial dataset inputs to enable the model to operate in a spatial environment.

SINCLAIR MODEL

Sinclair model has also been used operationally by Production Estimates and Crop Assessment Division (PECAD) of Foreign Agricultural Service (FAS) to provide estimation on global agricultural production (Reynolds, 2001). This model has been described as "semi-mechanistic" and is considered as "a compromise between completely empirical approaches and extremely detailed mechanistic approaches" (Speath *et al.*, 1987). According to Speath *et al.* (1987) this model "uses five major relationships: 1) leaf emergence as a function of temperature, 2) leaf area index as a function of leaf number and plant population, 3) interception of solar radiation as a function of leaf area, 4) biomass accumulation proportional to intercepted radiation, and 5) seed yield proportional to biomass". The model has several sub-modules for various physiological processes required for soybean growth simulation and yield assessment: a. Leaf Growth b. Carbon Budget calculation c. Vegetative Growth d. Nitrogen Budget calculation e. Seed Growth f. Water Budget calculation g. Calculation of Development rate (Sinclair, 1986).

The model uses daily inputs of temperature (daily minimum and maximum), precipitation, solar radiation, day length and planting date. The temperature, precipitation, solar radiation data are usually obtained from meteorological stations. The planting date is estimated from reports and local knowledge. The day length is calculated based on latitude.

Based on the study of how these inputs as well as various physiological processes are modeled in this particular model, the use of remote sensing based inputs to adapt this model into an efficient operational regional prediction methodology is discussed in the following paragraphs.

THE LAI FACTOR

The Sinclair model simulates Leaf Area Index (LAI) by using the temperature values to simulate the Daily Increase in Phyllochron Index (DIPI) which is used to obtain the increase in plant leaf area to calculate the LAI (Sinclair, 1986). The vegetation vigor that can be detected from remote sensing can be related to the leaf area index of the Sinclair model. For regional-yield predictions low resolution satellite images are usually employed. Both Moderate Resolution Imaging Spectro-radiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) provide vegetation index products such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) due to the availability of Infrared and Red channels.

But, due to the large area coverage, the images from these sensors usually have some amount of cloud cover present; therefore, compositing processes are needed to remove the cloud cover. Currently available temporal data composites provide a representative data product for particular sequential time periods such as 7, 10, 14, or 16 consecutive days. Hence, currently available temporal composite datasets preclude daily simulations and require that simulations be performed in a basis depending upon the compositing time frame.

In the process of using remote sensing to estimate LAI for the soybean yield model, Doraiswamy *et al.* (2005) successfully used simulated LAI from a reverse Scattering by Arbitrary Inclined Leaves (SAIL) model to estimate soybean and wheat yield. The MODIS reflectance values were used to simulate LAI. In the study, the simulated LAI was also used to reset the model parameters.

Another possible source of the LAI is to use NDVI to estimate LAI. Studies have claimed that such estimates are not very useful as NDVI value saturates at high LAI (Huete *et al.*, 2002; Huete *et al.*, 1999). A possible remedy for this shortcoming in NDVI would be to use EVI. EVI is considered to have “improved sensitivity to high biomass and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmospheric influences” (Huete *et al.*, 1999).

MOD 15 LAI grid data product is also available for daily and 8-day basis with resolution of 1 km (NASA, 2006). The use of MODIS derived LAI (MOD 15) product has not been validated for the soybean predictions. Since LAI products from MODIS are already available and ready for use, it would be good idea to research the use of MODIS LAI or LAI from EVI as an input to the soybean model.

METEOROLOGICAL INPUTS TO THE MODEL

Precipitation is one of the important input parameter for running the model. The model is highly sensitive to soil water as soil moisture is a highly limiting factor for different physiological processes for soybean growth. Therefore, daily accurate precipitation data is important to obtain correct yield predictions. Similarly, temperature is also an important factor for soybean growth model. Solar radiation data is also another input that is required daily for the model.

For regional yield prediction various meteorological inputs have been used by various researchers. Reynolds *et al.* (2000) used Rainfall Estimation images which had resolution of 7.6 km obtained from geostationary Meteosat - 5 satellite for Africa. Liang *et al.* (2004) used North American land data assimilation system (NLDAS) forcing data of about 1/8 degree for coupling with the crop model.

In cases of both local and regional predictions, the most used source of meteorological datasets has been the use of data from meteorological stations. One source of meteorological data is from National Climatic Data Center (NCDC) which provides daily meteorological data sets that includes precipitation, daily min and max temperature among other weather parameters. Doraiswamy *et al.* (2005), Bastainssen and Ali (2003), and Carbone *et al.* (1996) and others have used local weather station data to input precipitation and temperature.

PECAD's operational global crop assessment method is based on an automated decision support system called Crop Condition Data Retrieval and Evaluation (CADRE). The agro-meteorological data to CADRE is provided by Agricultural Meteorological Model (AGRMET) and World Meteorological Organization (WMO) network of weather stations (Reynolds, 2001). AGRMET provides precipitation, minimum and maximum temperature, snow depth, solar and long wave radiation and potential and actual evapo-transpiration. The AGRMET data has the resolution of 40km. The vegetation datasets can be obtained in 250 m for MODIS and 1km for AVHRR, but it is almost impossible to obtain the same resolution data for meteorological datasets. The use of 1km dataset with 40

km AGRMET dataset may not produce the desired results. Geostationary Operational Environmental Satellite (GOES) which has 4 km resolution covers US in which precipitation dataset is the main component that it provides from its distribution website. There are few other products that integrate various sources of weather information to provide weather information like the one used by Liang *et al.* (2004).

NASA has sponsored the development of the NASA Land Information System (LIS), “a functional Land Data Assimilation System (LDAS).” It currently is comprised of a LIS core, three land models, data servers, and visualization systems – integrated in a high-performance computing environment (Peters-Lidard *et al.*, 2004; Kumar *et al.*, 2005; and Tian *et al.*, 2005). LIS has been developed using an “ensemble physics land surface modeling philosophy” in order to enable interactions with other earth system model and decision support systems (DSS). LIS is implemented under the Earth Systems Modeling Framework (ESMF), an interoperable computational framework that facilitates the coupled interactions between other computational models of weather, climate and the environment. The computational resources available to LIS support the global modeling of land-atmosphere studies at 1km spatial resolutions. The land models in LIS incorporate surface parameters of temperature, snow/water, vegetation, albedo, soil conditions, topography, and radiation. Many of these parameters are available from NASA platforms at various spatial and temporal resolutions. For example, the MODIS sensor onboard the Terra and Aqua platforms provide vegetation parameters of Leaf Area Index (LAI), surface albedo, surface temperature, evapotranspiration, and radiation.

PLANTING DATE ESTIMATION

Planting date is one of the most sensitive inputs to the model. Planting date of the crop may not be of any importance for field level analysis, but for a regional level, planting date estimation becomes very important because it is almost impossible to correctly obtain the exact planting dates from all the areas. Usually the planting date is calculated based on crop reports which are subjective and can deviate up to 30 days. This offset might be lessened by the use of remote sensing by studying the phenology of the crop. Vegetation index information, which can be obtained from low resolution imageries with high temporal visits such as AVHRR and MODIS, can provide the vegetation index temporal curve. These temporal curves can help us to detect the onset of greenness that can be related to crop emergence. The detection of emergence can in turn help us to detect the planting date by using heat units. This process can be performed spatially using temporal map algebra which has the capability of spatial-temporal analysis, the results of which may be employed to reset the planting data to a more accurate value than the initializing estimate.

Mali *et al.* (2005) used the temporal map algebra to obtain NDVI composites from AVHRR and MODIS. However, the temporal map algebra can be used in any spatio-temporal analysis in which a temporal dataset can be treated as a three-dimensional dataset and various spatial operations can be performed (Mennis and Viger, 2004).

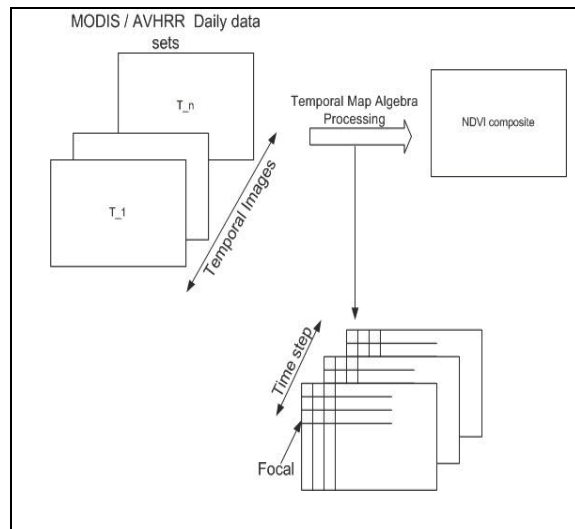


Figure 1. The NDVI composite is created by using temporal range of NDVI from temporal map algebra

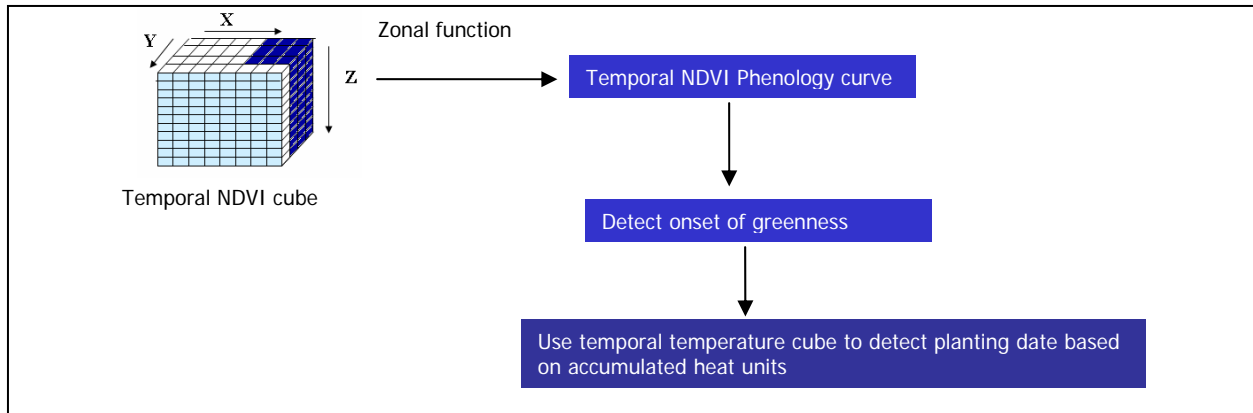


Figure 2. Planting date estimation using Vegetation Index and Temporal map algebra

CONCLUSIONS

One of the challenges in using the remote sensing based input for yield prediction is the unavailability of meteorological as well as vegetation based grid datasets in a spatially compatible form. In spatial analysis, the output dataset is as good as the lowest resolution available. Therefore, in the process of operational yield modeling, the resolutions of different spatial datasets need to be as compatible with each other as possible. This theory is not true for only spatial resolution but also for radiometric variations as well as for temporal time frame in which the datasets will be used. Practically, we may not be able to achieve that goal but ample amount of consideration should be given so that the datasets are as much compatible as possible in radiometric, spatial and temporal aspects.

The Weather Research and Forecasting (WRF) model (Michalakes *et al.*, 2004), developed by the atmospheric community and supported by NASA, are being adopted as the model of choice by the GeoResources Institute (GRI). The WRF has been successfully coupled to the LIS via the ESMF. The WRF model has been demonstrated to improve precipitation estimates when coupled to the LIS (using ESMF) on the same grid. The atmospheric fields and products generated using WRF can be used to contribute vital environmental information for several cross-cutting applications, including mechanistic crop modeling. The coupled LIS-WRF analysis of the atmospheric conditions in the surface layer of the atmosphere and the ground surface will provide useful input parameters at high temporal and spatial resolutions to the Sinclair crop model, including air and ground temperature, humidity (dew point), solar and long wave radiation components, energy and moisture fluxes across the land-atmosphere interface, and soil moisture estimates at different levels.

In case of the LAI, the use of MODIS based LAI or EVI needs to be studied more. If validated, the process of operational yield forecasting will be much more efficient.

The current unavailability of daily datasets due to cloud cover makes the daily simulation of Sinclair model almost impossible. Current limitations in temporally composited products that remove cloud cover restrict the model to be performed weekly or biweekly instead of daily. An objective of ongoing research at GRI is to generate representative composite datasets free of cloud cover within temporal windows of a shorter time frame (rather than using the absolute time frame of weekly or biweekly periods) to determine if this improves the ability of the model to simulate crop growth effectively.

Figure 3. illustrates how a temporal image “hypercube” can be evaluated in different ways utilizing 1) a *sliding temporal window* to create a continuum of composites over time to better track the progress of vegetation growth and monitor phenological growth processes, and 2) a *closing temporal window* to create representative data sets over increasing or decreasing intervals of time, that may be evaluated to determine the temporal limitations and assist in finding the minimum temporal range of time for images that may be composited.

The use of temporal map algebra to obtain a phenological curve to help detect the planting date seems to be a good start in the approach to apply use of remote sensing and spatial analysis for effective results.

The discussions on this paper and the concepts provided as a possible solution are some of the work that the team at GRI is looking forward to achieve in the near future.

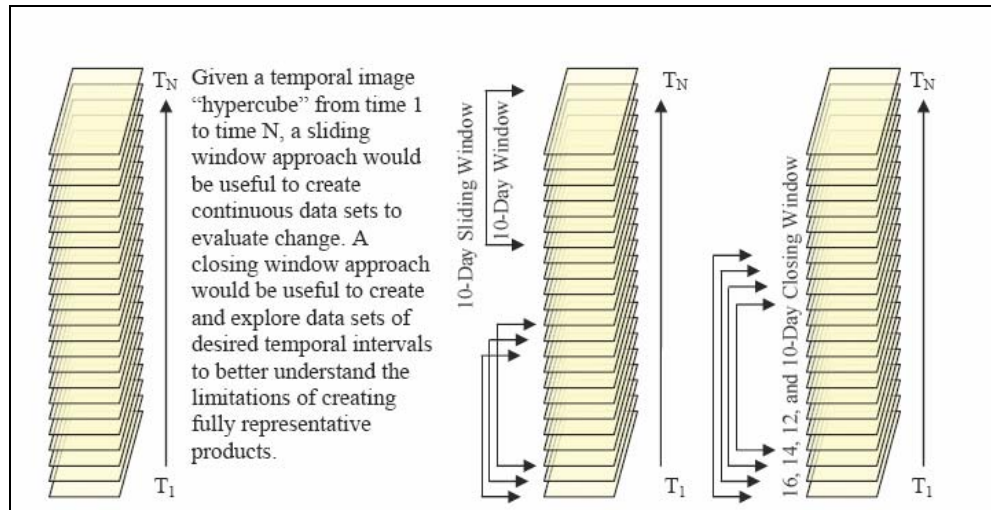


Figure 3. The representation of sliding window composite and closing window composites.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Robert Tetrault for his inputs which helped us form the conceptual framework of this paper. We are also grateful to Mr. Thomas Sinclair for his cooperation on the model.

REFERENCES

- Bastiaanssen, W. G. M. and S. Ali (2003). A new crop yield forecasting model based on satellite measurements applied across the Indus Basin Pakistan. *Agriculture, Ecosystems and Environment*, 94:321–340.
- Carbone, G.J., S. Narumalani, and M. King (1996). Application of Remote Sensing and GIS Technologies with Physiological Crop Models. *Photogrammetric Engineering & Remote Sensing*, 62(2): 171-179.
- Dabrawoska, K. Z., F. Kogan, A. Ciolkosz, M. Gruszczynska, and W. Kowalik (2002). Modeling of crop growth conditions and crop yield in Poland using AVHRR-based indices. *International Journal of Remote Sensing*, 2(6):1109–1123.
- Doraiswamy, P.C., T.R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, J. Prueger (2005). Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, 97: 192-202.
- Huete, A.R., C.O. Justice, and W.J.D Van Leeuwen (1999). MODIS Vegetation Index (MOD13) Version 3.0. URL <http://modis.gsfc.nasa.gov/data/atbd/land-atbd.html>.
- Huete, A. R., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83: 195–213.
- Kumar, S. V., C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. L. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood and J. Sheffield (2004). Land Information System - An Interoperable Framework for High Resolution Land Surface Modeling. *Submitted to Environmental Modelling & Software*.
- Liang, S., H. Fang, G. Hoogenboom, J. Teasdale, and M. Cavigelli (2004). Estimation of crop yield at the regional scale from MODIS observations. *Geoscience and Remote Sensing Symposium, IGARSS'04, Proceedings*. 2004 IEEE International, 20-24 Sept, vol. 3, 1625- 1628
- Lobell, D. B., G. P. Asker, J. I. OrtizMonasterio, and T. L. Benning (2003). Remote sensing of regional crop production in the Yaqui Valley Mexico: estimates and uncertainties. *Agriculture, Ecosystems and Environment*, 94:205–220.
- Mali, P., C. G. O’Hara, B. Shrestha, and V. Vijayaraj (2005). Use and Analysis of Temporal Map Algebra for Vegetation Index Compositing. *Proceedings of Third International Workshop on the Analysis of Multi-temporal Remote Sensing Imagery*, May 2005, Gulf Coast, Mississippi, 90 – 94.

- Mennis, J., and R. Viger (2004). Analyzing time series of satellite imagery using temporal map algebra. *Proceedings of the ASPRS 2004 Annual Convention*, 23-28 May, Denver, Colorado (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), unpaginated CD-ROM.
- Michalakes, J., J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang (2004). The Weather Research and Forecast Model: Software Architecture and Performance. *11th ECMWF Workshop on the Use of High Performance Computing In Meteorology*, 25-29 October 2004, Reading U.K. Ed. George Mozdzynski.
- Moulin, S., A. Biondeau and R. Delecalle (1998). Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing*, 19(6):1021–1036.
- NASA (National Aeronautics and Space Administration). MOD15-Leaf Area Index (LAI) and Fractional Photosynthetically Active radiation, URL: http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD_NUMBER=15.
- Peters-Lidard, C. D., S. Kumar, Y. Tian, J. L. Eastman, and P. Houser (2004). Global Urban-Scale Land-Atmosphere Modeling with the Land Information System. *Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone*, 84th AMS Annual Meeting 11-15 January 2004 Seattle, WA, USA.
- Rasmussen, M. S. (1998). Developing simple, operational, consistent NDVI-vegetation models by applying environmental and climatic information. Part II: Crop yield assessment. *International Journal of Remote Sensing*, 19(1):119–139.
- Reynolds, C. A. (2001). Summary of Global Data Sources and Processing Models. URL: <http://www.pecad.fas.usda.gov/cropexplorer/datasources.cfm>.
- Reynolds, C. A., M. Yitayew, D. C. Slack, C. F. Hutchinson, A. Huete, and M. S. Peterson (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. *International Journal of Remote Sensing*, 21(18):3487–3508.
- Sinclair, T.R. (1986). Water and Nitrogen limitations in soybean grain production. *Field crops Research*, 15:125-141.
- Speath, S.C., T.R. Sinclair, T. Ohnuma and S. Konno (1987). Temperature, Radiation and Duration dependence of high soybean yields: measurement and simulation. *Fields Crop Research*, 16:197-307.
- Tian, Y., C. D. Peters-Lidard, S. Kumar, J. Geiger, P. R. Houser, J. L. Eastman, P. Dirmeyer, B. Doty and J. Adams (2004) High Performance Land Surface Modeling with a Beowulf Cluster. *Submitted to Computing in Science & Engineering*.
- Wiegand, C. L., A. J. Richardson, D. E. Escobar and A. H. Gerbermann (1991). Vegetation indices in crop assessments. *Remote Sensing of the Environment*, 35:105–119.