SOYBEAN CROP-TYPE CLASSIFICATION IN BRAZIL USING TEMPORAL MODIS DATASETS

Jon Fridgen, Research Scientist
Karly Hellrung, Image Analyst
Ken Copenhaver, Program Director
Institute for Technology Development
Savoy, IL 61874
jfridgen@iftd.org
khellrung@iftd.org
kcopenhaver@iftd.org

ABSTRACT

Identification of specific crop types is important to deriving reasonable estimates of world-wide food and fiber production. Use of remotely sensed satellite imagery has the potential to provide stakeholders with crop type information over large geographic areas. During the 2004-2005 growing season, a project was conducted in the Matto Grosso state of Brazil to evaluate the use of temporal remotely-sensed datasets for the identification and mapping of soybean. Specifically, temporal surface reflectance data and vegetation index products derived from the 250 meter bands of the Moderate Resolution Imaging Spectrometer (MODIS) were used to develop a land cover/crop type classification for a study area near the city of Campo Novo do Parecis. Additionally, Shuttle Radar Topography Mission (SRTM) elevation data were obtained over the study area and fused with the MODIS data to account for changes in landscape topography. Data were analyzed using a combination of repeated measures statistical analysis, analysis of covariance and supervised classification techniques. Overall classification accuracies when considering all land cover/crop types was 75%. However, when classes were merged to develop a soybean-only classification, the overall accuracy increased to 83%.

INTRODUCTION

Land cover refers to the vegetative and manmade features covering the surface, which can be measured by remotely sensed imagery (Cambell, 2002). Classifying different types of land cover, specifically vegetation and crops, is extremely important for monitoring world-wide food production. Remotely sensed imagery has been used for many years to classify different types of land cover for this purpose. In a study conducted by Fang (1998), Landsat imagery was utilized to classify rice areas in the Hubei province of China. These classifications were then used to accurately predict rice crop yields for 1992 at 84.5 percent. Another example where satellite imagery monitored and estimated crop yields was performed by Lobell (2003) in the Yaqui Valley. This study used land cover classifications to estimate wheat and soybean yields within three to five percent of the actual harvested acres for the 1994 growing season. In 1999, NASA launched the Terra spacecraft that carries the Moderate Resolution Imaging Spectrometer (MODIS) sensor. MODIS provides data at resolutions of 250 meters to 1 kilometer for regional and world-wide research in vegetation dynamics. The MODIS data also has different vegetation and land cover products that may be utilized to produce higher accuracies when classifying different types of land cover and/or crop types.

The ability to classify land cover world-wide is a primary concern for many geographers for many different reasons. One reason that land cover is important is for global climatology. Knowing the different land cover can help research explain different weather and biochemical phenomena (Townshend et al., 1991). Another purpose for land cover classification research is land cover change, this is extremely important for urbanization and deforestation. The reason this project was conducted is because of food security and disease detection for vegetative land cover. Land cover classifications can be produced accurately by running analysis of satellite imagery combined with ground data for each land cover type in the study area.

Satellite and airborne-based remote sensing has been used to monitor crop health for a number of agricultural applications including drought and herbicide stress (Adcock et al., 1990; Carter and Miller, 1994), and nutrient deficiencies (Mariotti et al., 1996). Remotely sensed data has also been used for the purpose of disease detection. Jackson (1986) discusses a number of examples on the use of ground-based radiometry for the detection of plant diseases. Data simulating the spectral range of Landsat MSS or TM data was particularly useful for distinguishing between infected and non-infected plants. A study conducted by Clark et al. (1981) showed aerial infrared photography could be used to distinguish between healthy cereal crops and those infected with the Barley Yellow Dwarf Virus. Additionally, the southern corn blight watch project in the U.S. demonstrated the efficacy of aerial infrared photography for crop disease surveillance (Bauer et al., 1971; Myers, 1983).

This paper discusses the use of temporal MODIS datasets for the identification of soybeans grown in Brazil. The MODIS 250 meter resolution vegetation index products were used to identify different types of land cover with an emphasis on the identification of soybean. The work presented here is part of a larger effort in which NASA Earth science data was used in an attempt to identify soybean rust. Analyses focused on: 1) identification of various land cover types to develop a soybean-only classification for the study area; and 2) statistical modeling of the temporal vegetation products.

METHODOLOGY

Study Site

The study site was located in Brazil near the city of Campo Novo do Parecis in the state of Mato Grosso (Figure 1). In the area surrounding Campo Novo de Parecis, approximately 370,000 ha of soybeans are grown annually, contributing to Brazil's role as one of the larger soybean-producing countries in the world. The study area is characterized by a warm, humid tropical climate making it conducive for the development of fungal crop pathogens, especially soybean rust. These characteristics and the known presence of soybean rust in Brazil, are the reasons Campo Novo de Parecis was chosen as the focus of this research.

Figure 1. Location of the study area in Brazil. Ground data was collected near the city of Campo Novo de Parecis.

Remotely Sensed Data

A number of MODIS products were used in this project. First and foremost, the vegetation index products (MOD13) were utilized because of their ability to provide crop condition assessments. An initial classification was developed using the MOD12 product for the 2004/2005 growing season, and was then refined using the ground data collected by project team. MODIS land cover product's main goal is to output biophysical information at the regional and global scale, which is why it was utilized for this research. Another reason the MOD 12 land cover classification was used for soybean classification analysis is it has been proven to be a significant improvement in classification quality and accuracy for global land cover compared to the advanced very high-resolution radiometer (AVHRR) global land cover data (Friedl, et al., 2002).

The MOD09 daily and 8-day composite surface reflectance products were also used for the project. The MOD09 daily surface reflectance product serves as the primary input to many of the higher level products derived from the sensor (i.e., LAI, vegetation indices). In this case, the MOD09 daily and 8-day composites were used to create vegetation index products that have increased temporal resolution over the standard MOD13 16-day vegetation index product.

The MODIS data products covering the study area were located in cell h12v10 on the MODIS Sinusoidal grid. The data products obtained, their spatial resolutions, and dates collected are shown in Table 1. All MODIS products obtained were from the Terra spacecraft.

Shuttle Radar Topography Mission (SRTM) data was also obtained for the study site. The SRTM digital elevation data is available world wide with a spatial resolution of 90-m. The data were acquired in February 2000 during the STS-99 mission of Space Shuttle Endeavor. The published accuracy of the SRTM data is 16-m at the 90% confidence level. The digital elevation data, when combined with other topographic attributes (i.e., slope, aspect) is useful for quantifying landscape position within a given region or watershed. Ground-based observations of land cover seemed to indicate that elevation and landscape position may play a role in land cover type. Thus, it was hoped that using the SRTM in conjunction with the MODIS data may lead to higher classification accuracies.

Table 1. Terra MODIS products, descriptions, resolutions, and acquisition dates used in the project.

Product Name	Description	Resolution	Acquisition Day Interval†	Acquisition Year
		— m —	Julian Day	
MOD09A1	Surface Reflectance – 8 day	500	273 - 097	2001/02; 2004/05
MOD09Q1	Surface Reflectance – 8 day	250	273 - 097	2001/02; 2004/05
MOD09GHK	Surface Reflectance – daily	500	273 - 097	2001/02; 2004/05
MOD09GQK	Surface Reflectance – daily	250	273 - 097	2001/02; 2004/05
MOD12Q1	Land Cover Type	1000	NA‡	2001
MOD13A1	Vegetation Indices – 16 day	500	273 - 097	2001/02; 2004/05
MOD13Q1	Vegetation Indices – 16 day	250	273 - 097	2001/02; 2004/05

† The Julian day range given represents the South American growing season. For example, the 2004/05 growing season would span the time from day 273, 2004 to day 97, 2005.
‡ Not applicable

Ground Data

Ground data was collected during the 2004-05 soybean growing season near the city of Campo Novo do Parecis. A cooperative agreement was established with EMBRAPA (The Brazilian Agricultural Research Corporation) to assist in obtaining ground data. The ground data was initially collected at the end of January 2005 by the research group, while throughout the remainder of the growing season (February, March and early April), EMBRAPA arranged for a scout to collect the ground data.

The ground data points were collected randomly throughout the study site using a hand-held computer equipped with a global positioning system (GPS). At each data acquisition point, data recorded included geographic location along with the land cover/crop type. If the crop type was soybean, additional data was collected that included growth stage (Table 2), presence of soybean rust, percent defoliation (from soybean rust), and if available, planting date and management practices. In all, a total of seven different land cover classes were identified (Table 3).

Table 3. Land cover classification categories used for the study area near the city of Campo Novo do Parecis.

Land Cover Category
Soybeans
Deciduous Forest
Sugarcane/Pasture
Savanna
Savanna Being Cleared
Cotton
Urban

Table 2. Soybean growth stages used in Brazil.†

Stage	Description
Vegetative Phase	
Vc	Cotyledon stage
V1	First trifoliate leaf stage
V2	Second trifoliate leaf stage
V3	Third trifoliate leaf stage
Vn	Nth trifoliate leaf stage
Reproductive Phase	
R1	~ 50% bloom
R2	Full bloom
R3	Beginning pod development; pods > 1.5 cm
R4	Pods 2-4 cm; no seed development
R5.1	< 10% seed development complete
R5.2	10% - 25% seed development complete
R5.3	25% - 50% seed development complete
R5.4	50% - 75% seed development complete
R5.5	75% - 100% seed development complete
R6	Complete seed development
R7.1	50% yellow leaves
R7.2	51% - 75% yellow leaves
R7.3	> 75% yellow leaves
R8.1	< 50% defoliation
R8.2	> 50% defoliation
R9	Full maturity

† Adapted for Brazil by J. T. Yorinori, 1996.

Data Processing and Analysis

The MOD09 eight-day composite surface reflectance product was utilized to calculate the normalized difference vegetation index (NDVI) and Enhanced Vegetation Index (EVI) datasets. After the vegetation indices were calculated, a compositing algorithm was used to create the vegetation index composites. The composites were created by selecting the maximum value for a given pixel over a specified number of days. After creating the vegetation composites with the MOD09 eight-day composite surface reflectance, it was determined that the eight-day composites contained numerous no-data pixels resulting from the extensive cloud cover over the study area. Thus, it was decided that the MOD13 16-day vegetation index composites would be used in the classification process.

Classification analysis was mainly conducted on the 2004/05 dataset. However, analysis were also performed using the MOD13 data collected in 2001/2002 as it coincided with the MOD12 land cover classification product. This was accomplished to determine how accurate the MODIS data could classify different types of land cover. A majority of the analysis focused on image processing techniques to separate soybean from all other land cover/crop types. Additional analyses were performed using statistical repeated measures analysis.

Statistical analysis was performed using the MIXED procedure in the SAS System Version 8 (SAS Institute, Cary, NC). The MIXED procedure was used because of its ability to model the covariance structure of repeated measures (i.e., the vegetation index measured across time) (Littell et al., 1996). The analysis modeled the vegetation indices as a polynomial function of time. This analysis was also performed on the 2001/2002 vegetation index data and the associated land cover type. For all statistical analyses, a significance level of 0.05 was used.

Image analysis focused on the use of supervised classification algorithms to first classify the different land cover types and then separate the soybean from all other land cover types. Prior to performing any analysis on the NDVI and EVI datasets, a classification method had to be selected. The primary algorithms evaluated were Maximum Likelihood Classification and the Spectral Angle Mapper (SAM). The Maximum Likelihood Classification algorithm is based on the probability that a pixel belongs to a particular class. This is based off the assumptions that the bands have a normal distribution and that the probabilities are equal for all of the classes (Jensen, 1996; Leica Geosystems GIS and Mapping, 2003). The SAM algorithm classifies data based on a set of reference spectra that define each class. The angle is then computed between each band vector and each reference spectra vector. The algorithm assigns a given pixel to the class that has the smallest angle between the band vector and a reference spectra vector.

Statistical analysis was performed using the MIXED procedure in the SAS System Version 8 (SAS Institute, Cary, NC). The MIXED procedure was used because of its ability to model the covariance structure of repeated measures (i.e., the vegetation index measured across time) (Littell et al., 1996). The analysis modeled the vegetation indices as a polynomial function of time. This analysis was also performed on the 2001/2002 vegetation index data and the associated land cover type. For all statistical analyses, a significance level of 0.05 was used.

When evaluating the different classification methods, the 2001/2002 MOD13 vegetation index data was used in conjunction with the MOD12 land cover classification product. Although the MOD12 product contains five land cover classification schemes, only the Type 5 land cover classification scheme was used in this analysis. The Type 5 classification scheme is referred to as the Plant Functional Types (PFT) classification and identifies twelve different land cover types. The Type 5 PFT scheme was selected because it identifies both broadleaf and cereal crops and has an overall classification accuracy of seventy-eight percent when using the maximum likelihood classification algorithm (LPDAAC, 2005). Also, since soybean is considered a broadleaf crop, the Type 5 scheme was considered to be a good starting point in developing a soybean crop type classification.

The maximum likelihood algorithm performed best with the above 2001/2002 dataset and was then used to develop the crop type/land cover classification for the 2004/2005 growing season. The MOD12 product functioned as a guide to initiate and determine the land cover classes to be used for the 2004/2005 dataset, but was not used in the accuracy assessment for the 2004/2005 land cover classification. To ensure classification error rates were determined in an unbiased manner, the ground data was randomly split into two groups: two-thirds of the data for training and one-third for evaluation/accuracy assessment. After performing all classifications, an accuracy assessment was computed using a confusion matrix for the remainder of the data that was excluded from the classification algorithm training process. Overall accuracies as well as the producer’s and user’s accuracy were computed.

RESULTS

Statistical Analysis

Statistical analysis of the MOD13 vegetation index data and the MOD12 land cover product was performed using the MIXED procedure in SAS. The goal of this initial analysis was to determine if one of the vegetation indices (NDVI or EVI) performed better than the other at distinguishing land cover types. Each of the vegetation indices was modeled as a polynomial function of time within each of the land cover classes. Least squares means were calculated and plotted versus Julian day (Figure 2) to observe the patterns of the vegetation indices over time.

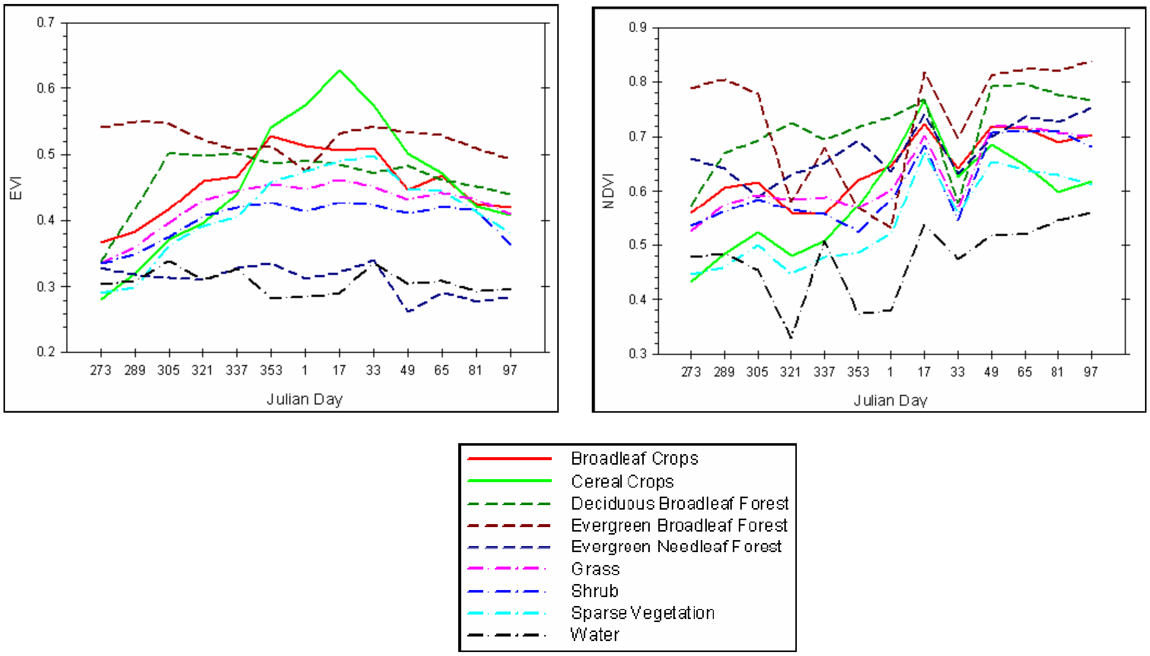

Figure 2. Plots of the least square means for the EVI (left) and NDVI (right) versus Julian day for each of the land cover classes.

Results of the analysis indicate a great deal of fluctuation in the NDVI across time within a given land cover type, whereas the EVI exhibits a much smoother trend with few spikes. This possibly could be explained by the ability of EVI to be more resistant to atmospheric effects than the NDVI (Huete et al., 1999). The EVI was modeled as a polynomial function of time for each of the growth stages observed in the ground data. This analysis only utilized the last six dates of the temporal dataset (Julian day 17 through Julian day 97) as this represented the time period in which a majority of the soybeans entered the reproductive phase and proceeded toward senescence and maturity.

Based on the initial ground data collected in late January, eight soybean growth stages were identified and utilized in this analysis. The analysis was somewhat confounded by the fact that as the soybean crop matured and was harvested, a second crop such as cotton was planted. Thus, the EVI decreased as the soybean crop progressed through senescence to maturity and would then begin to increase as the follow-on crop began to emerge and grow.

Of the eight reproductive growth stages of soybeans examined, the soybeans that were at growth stage R3 in late January appeared to show the greatest difference in vegetative status (as determined by the EVI). For soybeans that were in the more advanced reproductive phases (i.e., R4 through R6), the analysis tended to be confounded by the fact that the crop was reaching maturity and another crop was then being planted (e.g., cotton). This is evidenced by the increase in the after it reaches a minimum. A more thorough analysis could have been performed if a crop growth model could have been utilized to augment ground observations on growth stage.

Image Analysis

Results of the statistical analysis indicated a large amount of noise in the NDVI data, most likely due to atmospheric influences. This suspicion was further confirmed when land cover classifications were attempted using the Maximum likelihood algorithm. When classification accuracies were compared, the EVI classification accuracy exceeded that of the NDVI. Again, this was likely due to the fact that the EVI is not as sensitive to the atmospheric conditions as the NDVI (Huete et al., 1999).

Since the initial evaluation of classification techniques indicated the Maximum Likelihood algorithm generally provided higher classification accuracies, the next step was to classify the 2004/05 data using the Maximum Likelihood algorithm (Figure 3). Using the remaining one-third of the data that was excluded from the classification, an accuracy assessment was performed by computing a contingency table (Table 4). The accuracy assessment results revealed an overall accuracy of 75%. user accuracies (89% and 93%) were obtained for the soybean and deciduous forest classes, respectively.

Table 4. Land cover classification accuracy assessment results using the 2004/2005 temporal EVI data and Maximum Likelihood classification.

Land Cover Classification	Represented Color	User Accuracy (%)	Producer Accuracy (%)
Soybeans	Green	89	58
Deciduous Forest	Yellow	93	82
Sugarcane/Pasture	Light Purple	0	0
Savanna	Dark Brown	0	0
Savanna Being Cleared	Light Brown	30	6
Cotton	White	57	75
Urban	Purple	78	26
Overall Accuracy	75%		

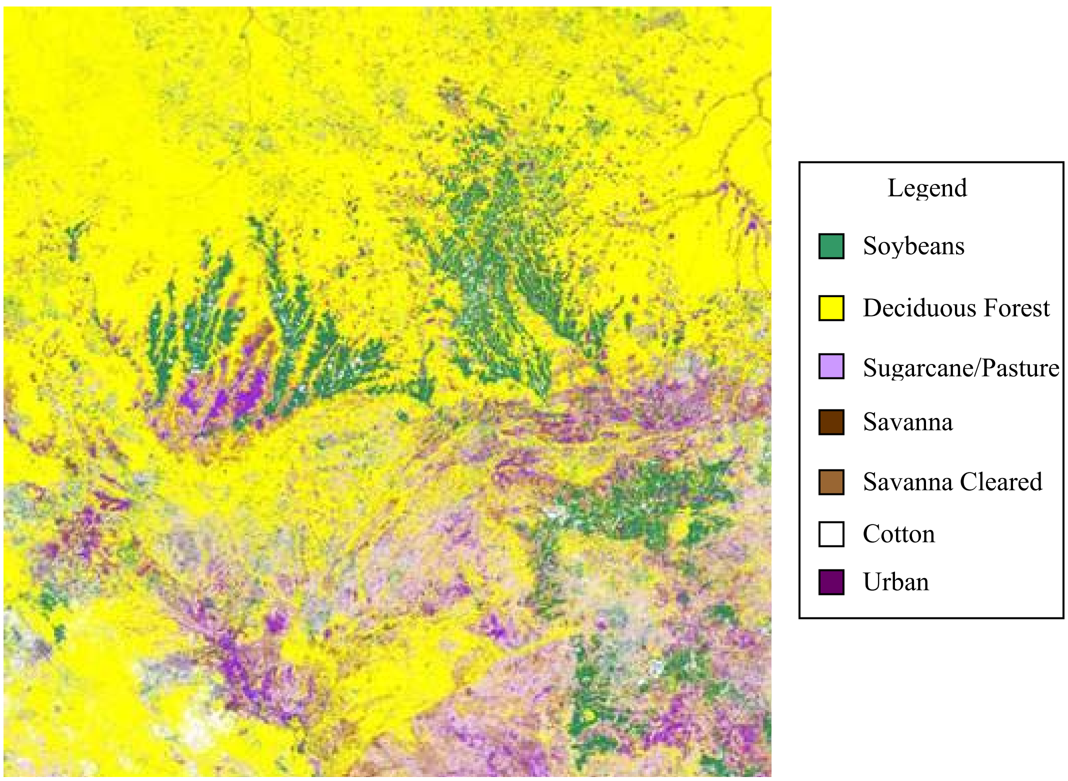

Figure 3. Land cover classification derived using the 2004/2005 temporal EVI dataset and Maximum Likelihood classification.

Using the above classification, all pixels that represented any land cover/crop type class other than soybean were masked out. The maximum likelihood classification was preformed on the soybeans and the masked out other class, which consisted of all of the other land cover class signatures grouped together. Below is the output classification (Figure 4). Masking of the other pixels resulted in a slightly higher classification producer accuracy of eighty-five percent for the soybean category and an overall accuracy of eighty-three percent (Table 5).

Table 5. Accuracy assessment results obtained with soybean versus all other land cover/crop types.

Land Cover Classification	Represented Color	User Accuracy (%)	Producer Accuracy (%)
Soybeans	Green	60	85
Other	Black	95	82
Overall Accuracy	83%		

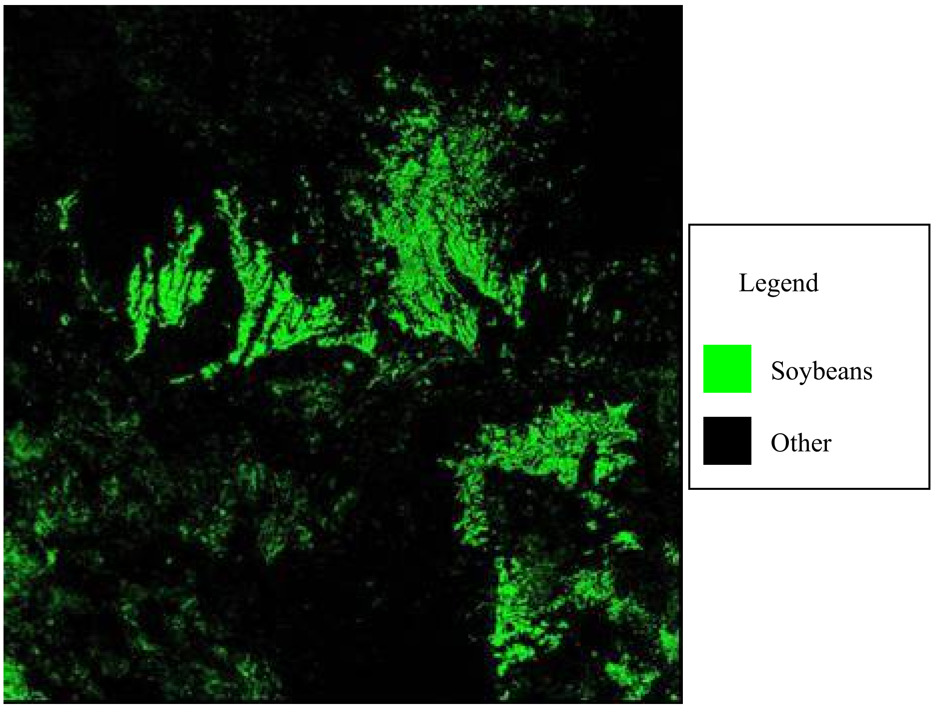

Figure 4. Maximum Likelihood classification result of classifying soybeans against all other land cover.

CONCLUSIONS

Land cover is an important topic of research for many reasons, food security and disease detection being top priority. The project's main goal was to use MODIS satellite imagery to accurately classify land cover, with a focus on the soybean class. This has been obtained at accuracy levels above eighty percent. These classifications can then be used to aid in detecting fungus, such as soybean rust, and hopefully stop it from destroying yield. Research also indicated that for this project the MODIS EVI product classifies land cover more effectively potentially due to being less susceptible to atmospheric conditions. The NDVI product did not prove to be useful potentially because the atmospheric conditions cause the data to not be attainable.

Future studies could be conducted with MODIS data for land cover classifications. A focus on different land cover types would be of interest for areas of deforestation and land cover change over time. Other potential research topics for land cover would be using other vegetation products to classify land cover, such as LAI or fPAR. Other products that could aid in higher classification accuracies are the Tropical Rainfall Measuring Mission (TRMM), solar radiation, water balance, and net photosynthesis. MODIS products have proven to accurately classify land cover, other MODIS products should be tested to determine if they can create higher classification accuracies.

REFERENCES

Adcock, T.E., F. W. Nutter Jr., and P. A. Banks (1990). Measuring herbicide injury to soybeans (Glycine max) using a radiometer. *Weed Science*, 38: 625-627.

Bauer, M.E., R. P. Mroczynski, R. B. MacDonald, and R. M. Hoffer (1971) Detection of southern corn blight using color infrared aerial photography. *Proc. Biennial Workshop on Color Aerial Photgraphy in Plant Science,*. 225-226.

Campbell, J.B. (2002). *Introduction to Remote Sensing, 3rd Edition*. The Guilford Press, New York.

Carter, G.A. and R. L. Miller (1994). Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment*, 50: 295-302.

Clark, R.V., D.A. Galway, and Y.C. Paliwal (1981). Aerial infrared photography for disease detection in field plots of barley, oats, and wheat. *Phytopathology*, 71(8): 867.

Fang H., B. Wu, H. Liu, and X. Huang (1998) Using NOAA AVHRR and Landsat TM to estimate rice area year-by-year. *International Journal of Remote Sensing*, 19(3):521-525.

Friedl, M.A., D.K. McLver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C Schaaf (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83: 287-302.

Huete, A., C. Justice, and W. van Leeuwen. (1999). MODIS vegetation index (MOD13) algorithm theoretical basis document. NASA-Goddard Space Flight Center, Greenbelt, MD. http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf

Jackson, R.D. (1986). Remote sensing of biotic and abiotic plant stress. *Ann. Rev. Phytopathology*, 24: 265-287.

Jensen, J.R. (1996). *Introductory digital image processing: A remote sensing perspective, 2nd edition*. Prentice-Hall, New Jersey.

Leica Geosystems GIS and Mapping, LLC, ERDAS Field Guide 7th ed., Atlanta, GA: Leica Geosystems GIS and Mapping, LLC, 2003.

Littell, R.C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger (1996). SAS system for mixed models. Cary, NC: SAS Institute, 87-134.

Lobell, D.B., G. P. Asner, J.I. Ortiz-Monasterio, 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.

LPDAAC. (2005). MODIS/Terra Land Cover Type 96-day L3 Global 1km SIN Grid. LPDAAC User Services, Sioux Falls, SD. http://lpdaac.usgs.gov/modis/mod12q1v4.asp

Mariotti, M., L. Ercoli, and A. Masoni (1996) Spectral properties of iron-deficient corn and sunflower leaves. *Remote Sensing of Environment*, 58: 282-288.

Myers, V.I. (1983). Remote sensing applications in agriculture. *Manual of Remote Sensing*, 2111-2128.

Townshend, J., C. Justice, W. Li, and J. McManus (1991). Global Land Cover Classification by Remote Sensing: Present capabilities and Future Possibilities. *Remote Sensing of Environment*, 35: 243- 255.

Yorinori, J.T. (2005). Personal communication.