5-1-2010

Using remote sensing and grid-based meteorological datasets for regional soybean crop yield prediction and crop monitoring

Preeti Mali

USING REMOTE SENSING AND GRID-BASED METEOROLOGICAL DATASETS FOR REGIONAL SOYBEAN CROP YIELD PREDICTION AND CROP MONITORING

By

Preeti Mali

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Agronomy
in the Department of Plant and Soil Sciences

Mississippi State, Mississippi

May 2010

USING REMOTE SENSING AND GRID-BASED METEOROLOGICAL DATASETS FOR REGIONAL SOYBEAN CROP YIELD PREDICTION AND CROP MONITORING

By

Preeti Mali

Approved:

David R. Shaw Vice President Research and Economic Development William L. Giles Distinguished Professor (Major Professor)	Charles G. O'Hara Associate Research Professor Geosystems Research Institute (Co-Major Professor)
K. Raja Reddy Professor of Plant and Soil Sciences (Committee Member)	Daniel Reynolds Professor of Plant and Soil Sciences (Committee Member)
Roger L. King Director Center for Advanced Vehicular Systems William L. Giles Distinguished Professor (Committee Member)	William Kingery Graduate Coordinator Department of Plant and Soil Sciences

Melissa J. Mixon
Interim Dean of the College of Agriculture
and Life Sciences

Name: Preeti Mali

Date of Degree: May 1, 2010

Institution: Mississippi State University

Major Field: Agronomy

Major Professor: Dr. David R. Shaw

Co-major Professor: Dr. Charles G. O'Hara

Title of Study: USING REMOTE SENSING AND GRID-BASED METEOROLOGICAL DATASETS FOR REGIONAL SOYBEAN CROP YIELD PREDICTION AND CROP MONITORING

Pages in Study: 126

Candidate for Degree of Doctor of Philosophy

Regional crop yield estimations using crop models is a national priority due to its contributions to crop security assessment and food pricing policies. Many of these crop yield assessments are performed using time-consuming, intensive field surveys. This research was initiated to test the applicability of remote sensing and grid-based meteorological model data for providing improved and efficient predictive capabilities for crop bio-productivity.

The soybean prediction model (Sinclair model) used in this research, requires daily data inputs to simulate yield which are temperature, precipitation, solar radiation, day length initialization of certain soil moisture parameters for each model run. The traditional meteorological datasets were compared with simulated South American Land Data Assimilation System (SALDAS) meteorological datasets for Sinclair model runs and for initializing soil moisture inputs. Considering the fact that grid-based

meteorological data has the resolution of $1/8^{th}$ of a degree, the estimations demonstrated a reasonable accuracy level and showed promise for increase in efficiency for regional level yield predictions.

The research tested daily composited Normalized Difference Vegetation Index (NDVI) from Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (both AQUA and TERRA platform) and simulated Visible/Infrared Imager Radiometer Suite (VIIRS) sensor product (a new sensor planned to be launched in the near future) for crop growth and development based on phenological events. The AQUA and TERRA fusion based daily MODIS NDVI was utilized to develop a planting date estimation method. The results have shown that daily MODIS composited NDVI values have the capability for enhanced monitoring of soybean crop growth and development. The method was able to predict planting date within ±3.4 days. A geoprocessing framework for extracting data from the grid data sources was developed. Overall, this study was able to demonstrate the utility of MODIS and VIIRS NDVI datasets and SALDAS meteorological data for providing effective inputs to crop yield models and the ability to provide an effective remote sensing-based regional crop monitoring. The utilization of these datasets helps in eliminating the ground-based data collection, which improves cost and time efficiency and also provides capability for regional crop monitoring.

DEDICATION

To Bijay

To My Parents

To My Friends

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to Dr. Charles G. O'Hara and Dr. David R. Shaw. Dr. O'Hara has been my supervisor and a source of inspiration throughout this period. Dr. Shaw has always been there to help me with his superb guidance. I would like to express my sincere gratitude to my committee members, Dr. Kambham R. Reddy, Dr. Daniel Reynolds and Dr. Roger L. King for their invaluable comments and suggestions.

I would like to thank the institution of Geosystems Research Institute (GRI) and High Performance Computing Collaboratory (HPCC), which has been my second home since I came here in the August of 2003. I would like to thank all my friends, colleagues and staff members at GRI and HPCC, who have helped me tremendously with both personal as well as with their professional support.

A special appreciation of gratitude to all my friends, whose encouragement and companionship has helped me tremendously throughout this journey.

I would like to sincerely thank Dr. Thomas R. Sinclair and Dr. Luis Salado-Navarro for their help with data collection in Argentina and for assisting me with their expertise in Sinclair model. I would like to express my thankfulness to Dr. Luis Gustavo DeGoncalves for the SALDAS datasets and collaboration with the project. I would like to thank Kent Hilbert, Dr. David Lewis from ITD with their help with MODIS data collection. I would also like to thank Jefferey A. Russel and Dr. Kenton W. Ross from SSAI for their collaboration and help with ART toolbox.

A special mention of gratitude goes to my Parents (Mother and Father) and my two brothers whose words of encouragement kept me going through the years, though they were so far away. I would also like to thank my Father-in law (1936-2009) and my Mother-in-law for their support. Last but not the least, I cannot express enough gratitude to thank my husband, friend and colleague Bijay Shrestha for being there for me at all times.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES x

CHAPTER

I. INTRODUCTION 1

1.1 Research Introduction 1
1.2 Literature Review 4
1.2.1 Crop Yield Modeling 4
1.2.2 NDVI and Crop Productivity Monitoring 5
1.2.3 Crop Yield Models and Remote Sensing 7
1.2.3.1 Regression Based Emperical Method 7
1.2.3.2 Semi-Empirical Method 9
1.2.3.3 Mechanistic or Agro-meteorological Method 10
1.2.3.4 Methods useful for regional predictions 12
1.2.4 Sinclair Crop Model 13
1.3 Statement of Problem 15
1.4 Research Objectives 17
1.5 References 19

II. ANALYSIS OF SALDAS METEOROLOGICAL FORCINGS AND SALDAS SIMULATED SOIL MOISTURE FOR SOYBEAN YIELD ESTIMATION MODELING 21

2.1 Abstract 21
2.2 Introduction 22
2.2.1 Meteorological datasets in crop yield modeling 24
2.2.1.1 SALDAS forcing’s as input to Sinclair model 28
2.2.1.2 SALDAS for intializing soil moisture parameters 29

2.3 Study Area 31
2.4 Methodology 32
2.4.1 Data Collection 32
2.4.1.1 The South American Land Data Assimilation System (SALDAS)...32
2.4.1.2 Selection of Soybean Fields..33
2.4.1.3 Ground Meteorological Data 34
2.4.1.4 Soil Moisture Field Sampling 35
2.4.1.5 Yield, Cultivars and Planting Date 35
2.4.2 Data Preprocessing...35
2.4.3 Use of SALDAS forcings meteorological input 36
2.4.4 Use of SALDAS soil moisture as initializing inputs 37
2.5 Results and Discussion 38
2.5.1 Yield results from ground meteorological inputs 38
2.5.2 Yield results from all SALDAS forcings.......................................39
2.5.3 Yield results from SALDAS precipitation.....................................40
2.5.4 Yield results from SALDAS minimum temperature 41
2.5.5 Yield results from SALDAS maximum temperature.....................42
2.5.6 Yield results from SALDAS solar radiation 43
2.5.7 SALDAS soil moisture as initializing inputs.................................44
2.5.7.1 Use of SALDAS moisture values for initializing Sinclair model 45
2.6 Conclusions...51
2.7 References...53

III. THE UTILITY OF MODIS AND SIMULATED VIIRS IMAGERIES FOR MONITORING CROP PRODUCTIVITY 55

3.1 Abstract...55
3.2 Background and Introduction 56
3.3 Methodology 61
3.3.1 Study Area 61
3.3.2 Data Sources 62
3.3.2.1 Moderate Resolution Imaging Spectroradiometer (MODIS) 62
3.3.2.2 Visible Infrared Imaging Radiometer Suite (VIIRS)......63
3.3.3 Preprocessing 64
3.3.3.1 Field Boundary Delineation and Verification.................64
3.3.3.2 Data Subsetting, Reprojection and Format Change........65
3.3.3.3 VIIRS Simulation..65
3.3.3.4 NDVI Calculation 66
3.3.3.5 Zonal Value Extraction..66
3.3.3.6 Large Data Size and Volume Handling with Batch Processing 67
3.4 Analysis and Results 71
3.4.1 Comparison between MODIS and simulated VIIRS 71

3.4.1.1 Comparison using scatter plot..71
3.4.1.2 Validation of VIIRS simulation.....................................76
3.4.1.3 Validation of time-series curves based on daily NDVI with respect to Soybean Phenology for MODIS and simulated VIIRS 78
3.4.1.4 NDVI time-series depicting soybean growth..................79
3.4.1.5 Cross Validation Approach..83
3.5 Results and Discussions..93
3.6 Conclusions...95
3.7 References...97

IV. IMPROVEMENT IN PLANTING DATE ESTIMATION THROUGH THE USE OF NDVI DERIVED FROM SATELLITE IMAGERY...99

4.1 Abstract..99
4.2 Background and Introduction 100
4.3 Methodology..105
4.3.1 Data Collection and Preprocessing...105
4.3.1.1 Field Boundary Verification 106
4.3.1.2 Field Based Datasets Collected....................................107
4.3.2 Analysis and Results...107
4.3.2.1 NDVI Calculation and Fusion Based Compositing......107
4.3.2.2 Difference between AQUA and TERRA MODIS mean NDVI values..110
4.3.2.3 Cross Platform Fusion and Composite Creation...........110
4.3.2.4 Zonal Processing...111
4.3.2.5 Planting Date Estimation Process 112
4.4 Results and Discussion 113
4.5 Conclusions...118
4.6 References...120

V. FINAL SUMMARY AND RECOMMENDATIONS.....................................123

LIST OF TABLES

2.1 Satellite based meteorological data sources. 26

2.2 Integrated Data Sources. 27

2.3 Soybean fields used for the study for the year 2006/2007 in Argentina. 34

2.4 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield using SALDAS meteorological data. 40

2.5 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS precipitation. 41

2.6 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS minimum temperature. 42

2.7 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS maximum temperature. 43

2.8 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS solar radiation. 44

2.9 Comparison between SALDAS soil moisture values and field observed soil moisture. 45

2.10 Yield from six sets of simulations using different ESW and DEEP values results obtained. 49

2.11 Percentage differences of the simulated yields with ESW and DEEP from SALDAS with base yields (simulated yield results using all experts provided initializing variables and ground meteorological datasets as inputs). 50

2.12 Root mean square error (RMSE) calculated between yield from six sets of simulations using different ESW and DEEP values and base yield....................50

3.1 Comparison between spatial, temporal and spectral resolution of MODIS, VIIRS and AVHRR. 59

3.2 Soybean fields for the year 2006/2007 in Argentina. 69

3.3 Statistical differences between VIIRS images simulated from MODIS and AWiFS....................77

3.4 Correlation coefficient value between image to image comparison between MOD02 (atmospherically uncorrected level 1 image) vs. MOD09 (atmospherically corrected level 2 image)....................78

3.5 "Temporal Metrics" (Reed et al. 1994) and the corresponding phenological stages. 83

3.6 Values of various phenological stages for the soybean fields and the corresponding representative NDVI value from Terra MODIS and LAI values simulated from Sinclair Model....................86

3.7 Values of various phenological stages for the soybean fields and the corresponding representative NDVI value from Terra MODIS and LAI values simulated from Sinclair Model....................87

4.1 Estimated planting dates of early planted (October to early November) soybeans. 117

LIST OF FIGURES

1.1 Abridged schematic flowchart depicting inputs and program flow of major modules for Sinclair soybean model. 14

2.1 AWIFS imagery showing study sites selected for soybean farms and also for installing seven automatic weather stations (except Monte Buey)...31

2.2 Gridded SALDAS meteorological forcings and soil moisture data......................33

2.3 An example of fields selected for the study: Marcos Juarez fields over AWIFS imagery. 33

2.4 Geo-processing methods utilized to extract and process SALDAS grid datasets for providing “Sinclair model ready” inputs. 36

3.1 The AOI of our interest is shown by the red box inside the larger magenta colored box. The magenta colored box is the MODIS tile H12V12 used for this experiment...63

3.2 Simulated VIIRS NDVI (400 m) from MODIS data (250 m) using Application Research Toolbox (ART)..66

3.3 Flowchart of the geo-processing methodology used for the analysis. 68

3.4 Flowchart of the pre-processing steps performed before analysis. 70

3.5 TERRA MODIS NDVI and simulated VIIRS NDVI scatter plots for 14 fields in the area for 2006-2007 soybean growing season. 72

3.6 AQUA MODIS NDVI and simulated VIIRS NDVI scatter plots for 14 fields in the area for 2006-2007 soybean growing season. 74

3.7 NDVI curves depicting soybean growth characteristics for the fourteen soybean test fields in Argentina from AQUA MODIS (a), VIIRS simulated from AQUA MODIS (b), TERRA MODIS (c), VIIRS simulated from TERRA MODIS (d)..80

3.8 Charts showing the NDVI curves for various selected soybean fields in Argentina from TERRA MODIS and VIIRS simulated from TERRA MODIS and specific phenological stages for each field.............88

4.1 Time chart of various phenological stages for the soybean fields as calculated by the Sinclair crop model..103

4.2 Comparison between various zenith angle constraints for compositing process. The zenith angle of 48º provided a good cutoff point...............109

4.3 Flowchart of the geo-processing methodology used for the analysis.................112

4.4 Planting date estimation using Vegetation Index and Temporal map algebra (Mali et al., 2006)..113

4.5 Graph showing soybean phenology and growing degree days for soybean planted over soybean field in Rio Segundo. 114

4.6 Graph showing soybean phenology and growing degree days for soybean planted over wheat field in MonteBuey..115

CHAPTER I

INTRODUCTION

1.1 Research 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 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 pricing and trade policies. Thus, reliable yield prediction methods are highly important for national and global food security. The Production Estimation and Crop Assessment Division (PECAD) of United States Department of Agriculture (USDA)/ Foreign Agricultural Service (FAS) provide global crop yield forecasts for major food grain and oil seed crops. These estimates require a tremendous amount of ground data collection network. The ability of remote sensing and meteorological grid datasets to provide information on crop growth and environmental conditions that affect crop growth is a huge benefit for agencies such as USDA/FAS PECAD for regional yield predictions. With the benefits and limitations of remote sensing considered, this research has been developed with the hypothesis that utilization of remote sensing and spatial technologies can greatly benefit in regional scale crop yield estimation modeling. A detailed literature review has showed that most remote sensing based regional yield prediction models use coarse resolution imageries such as the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectro-radiometer (MODIS). The USDA/FAS PECAD currently utilizes (MODIS) Normalized Difference Vegetation Index (NDVI) for assessing the crop growth conditions. However, a future sensor Visible/Infrared Radiometer Suite (VIIRS) is expected to replace MODIS in the near future. This research compares MODIS and VIIRS to assess whether VIIRS sensor product is applicable to replace MODIS for regional yield prediction and crop productivity monitoring. The types of models used for crop predictions are highly varied. Different methods of crop prediction using remote sensing and spatial datasets for crop prediction have been researched. However, most of these methods either use highly empirical methods, or methods that use parameters that cannot be utilized for regional level predictions (Rasmussen, 1998; Dabrawoska et al., 2002; Bastiaanssen and Ali, 2003; Lobell et al., 2003). Doraiswamy et al. (2005) applied remote sensing to a semi-mechanistic crop model for regional yield assessment and found that with the use of the correct crop model, the information from remote sensing observations can be effectively integrated into crop modeling methodologies.

This research uses Sinclair model for soybean yield prediction for Argentina. Sinclair model has been used operationally by USDA/FAS PECAD to provide estimation on soybean 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, p. 298);

therefore, suitable for adapting to a geoprocessing environment. The model uses daily inputs of temperature (daily minimum and maximum), precipitation, solar radiation, day length and planting date. The input variables are obtained from meteorological stations. Besides these variables, the model also requires initialization of certain parameters during each model run, of which planting date and soil moisture content are very sensitive to the yield estimates.

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 MODIS and 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 field, 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 model is being used for regional level estimates, and on the other hand, the input parameters are estimated from remote sensing. Even then, it can be quite beneficial 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). Therefore, the research evaluates the use of remote sensing (current and future sensors) and grid-based meteorological datasets to reduce the need for detailed time consuming field data and for providing improvement in efficiency to monitor and model crop bio-productivity.

1.2 Literature Review

1.2.1 Crop Yield Modeling

Crop growth modeling was initiated and developed by C.T. de Wit. The origin of crop modeling can be traced into the publication in 1965 by C.T. de Wit on modeling photosynthesis as a function of leaf canopies (Boumat et al., 1996). The crop models that follow the modeling philosophy of C. T. de Wit are considered to belong to the "School of de Wit". De Wit and Penning (1982) proposed a basic classification of modeling system that consisted of four production situations: a) potential production b) water limited production c) nitrogen limited production and d) nutrient limited production

(Bouman et al., 1996). De Wit (1965) demonstrated that canopy photosynthesis is the sum of photosynthesis of all the individual leaves. De Wit (1982) introduced growth rate calculation as a function of time, a dynamic system of crop modeling was introduced.

1.2.2 NDVI and Crop Productivity Monitoring

The utilization of remote sensing in crop yield estimation and crop growth monitoring can be traced back to the development of vegetation indices that are based upon the plant spectral characteristics. Vegetation indices utilize the properties of the chlorophyll reflectance in the red and near-infrared region of the electromagnetic spectrum (Myneni et al., 2005). Among the many vegetation indices, NDVI is the most researched and widely used index. The development of the NDVI is shared by Tucker (1979) and Deering (1978). Tucker (1977) found that leaf water content was best estimated in the region of 0.4-0.5, 0.63-0.69 and 0.74 to 0.8 μm in the electromagnetic spectrum. He concluded that these resulted due to the strong chlorophyll absorption on the 0.4-0.5 μm, 0.63-0.69 μm and high reflectance of vegetation in the 0.74 to 0.8 μm regions. Tucker (1979) studied *in-situ* spectral reflectance of grass for 80% green biomass, 50% green biomass and dead biomass and compared various vegetation indices which included NIR/R, Visible/IR ratios. He concluded that for vegetation studies, the IR/Red ratio was most useful. Tucker (1979) also pointed out that some means of normalization for different irradiation conditions would be useful for studying the green leaf biomass of crops. Tucker (1979) found that the normalized difference transformation was effective in compensating for the variation in irradiation conditions. Tucker (1979) also found that the percentage of crop cover was closely related to vegetation indices. As crop cover increased or decreased, the vegetation index values measured had a corresponding change. Thus, he concluded that due to the observed relationship between the vegetation indices and crop development, crop conditions could be monitored through spectral measurements. The early set of researches on relationship between the vegetation growth and the spectral bands of NDVI as well as the IR/Red ratios has led to a great deal of research such that NDVI has been used as proven index for monitoring vegetation condition. Wiegand and Richardson (1984) found that plants express their development, stress response, and yield capability through spectral observable canopy relating vegetation index to leaf area index, fractional observed photosynthetically active radiation (PAR) and economic yield (Richardson, 1990).

Wiegand and Richardson (1990a) proposed a rationale in which the spectral observations i.e. vegetation indices were related to plant processes specifically leaf area, evapo-transpiration and yield. Wiegand and Richardson (1990b) tested the rationale proposed on relating vegetation indices to plant processes to cotton, wheat and corn. The vegetation indices used include NDVI, Perpendicular Vegetation Index (PVI), and red index and found that although limited, the vegetation indices do have relationships to crop growth and development and can be used to infer leaf area, evapo-transpiration and yield. Various researchers have since utilized NDVI to predict crop yield.

1.2.3 Crop Yield Models and Remote Sensing

The review of different methods of crop yield predictions has shown that crop yield predictions using remote sensing and spatial technologies can be basically categorized into the following (Moulin et al., 1998):

- Regression based empirical method
- Semi-empirical based method (Monteith based model)
- Mechanistic or agro-meteorological based method

1.2.3.1 Regression Based Emperical Method

Regression based crop yield models are developed on the basis of the relationship between crop yield to a variety of biophysical factors such as crop vigor, rainfall, temperature, and soil. Boken et al. (2002) used NOAA-AVHRR based composited NDVI for spring wheat model in Canadian Prairies, using a monthly model based on a cumulative moisture index. The main purpose of this research was to improve an operational wheat model using remote sensing information based on monthly weather data. The model uses monthly temperature and precipitation data, estimated daily crop water requirement to obtain the Cumulative Moisture Index (CMI), which provides the daily moisture data; these data are cumulated for the whole of the growing season (from sowing to harvest). The use of CMI is based on the theory that if soil moisture requirement has been met, optimum growth will be attained. The research found that the use of NDVI based variable in a regression model with CMI improves the prediction power of the model significantly. The coefficients of determination in a NDVI based model were 0.79, 0.96, 0.83, 0.95, and 0.39 in five districts as opposed to 0.13, 0.70, 0.70, 0.75, 0.50, and 0.00 in a regular monthly model. Boken et al. (2002) also compared different variables obtained from NDVI and crop growth with wheat yield, and found that average NDVI during the heading period correlated highly with the wheat yield.

In other research, Rasmussen et al. (1998) used NOAA-AVHRR NDVI based model for predicting crop yield in Senegal, West Africa. The nine-day maximum value composited NDVI imagery of the years 1990 and 1991 were used. Besides NDVI, percent tree cover data were collected through low altitude systematic reconnaissance flights. Similarly, Tropical Livestock Unit (TLU) densities and percentage cultivated land data and population density data for grownup males were collected. These data were collected as point data and were interpolated using inverse distance weighted method with grid cell size of 500 m. On regression analysis of the various data parameters, it was found that grain yield and time weighted NDVI values (iNDVI) were highly correlated. The application of the model to stratified data with greater than 22% cultivated land improved the yield from $r^2 = 0.62$ to 0.73. The TLU density showed a significant relation with yield. The regression model, for cultivated area of percentage value greater than 22%, based on the iNDVI and TLU density, improved the yield prediction to $r^2 = 0.88$.

Dabrawoska (2002) also used AVHRR NDVI based regression model for cereal yield estimation. In this research NOAA-AVHRR GAC (global area coverage) data with 4 km resolution was used to calculate NDVI and brightness temperature (BT). The NDVI and BT were further used to obtain VCI (vegetation condition index) and TCI (temperature condition index), respectively. Landsat data were also used to obtain agricultural distribution map to obtain pixels with agricultural land less than 50, 50- 70 and 70-100%. High correlations of yield with VCI were noted in the weeks of 16, 22 of crop growing period and TCI in the week 25 of the crop growing period. Therefore, Dabrawoska et al. (2002) developed a regression-based model using TCI at weeks 16 and 22 and VCI at week 25 of crop growing season. The prediction result showed only a mean average error of 4%.

1.2.3.2 Semi-Empirical Method

Montieth-based models can be considered semi-empirical in nature (Moulin et al., 1998). Bastiaanssen and Ali (2003) used a Monteith-based model that uses accumulated above ground biomass to predict yield. The biomass is derived from APAR (absorbed photosynthetically active radiation) values, which are derived from NDVI-derived PAR and APAR/PAR fraction. PAR values are obtained from incoming solar radiation values measured from Gumble Stokes recorders at ground meteorological stations. The method used also required estimation of light use efficiency values. The estimation of light-use efficiency requires values of ‘impact soil moisture’, ‘heat effect factors’, and ‘evaporative fraction’ that require complex derivations and assumptions. The authors found that their method was successful in predicting wheat, rice, and sugarcane yields, with 22, 29 and 23% relative deviation from the observed yield values. However, this method was not successful in predicting cotton yield and the reason has been given as the inability of AVHRR in distinguishing cotton fields. Lobell et al. (2003) also used a Montieth-based model to predict crop yield using Landsat TM. But the derivation of variables of the model is different from the method used by Bastiaanssen and Ali (2003). The APAR/PAR fraction and light-use efficiency are calculated using a different procedure. PAR values are calculated in the field using a pyranometer. APAR values are calculated using simple ratio (SR) and maximum and minimum possible APAR. The light-use efficiency is calculated from plot-based harvested biomass and APAR values.

1.2.3.3 Mechanistic or Agro-meteorological Method

Mechanistic models usually contain a defined process using crop state variables and energy, carbon, nutrient fluxes at crop/soil/atmosphere interfaces (Moulin et al., 1998). One such mechanistic agro-meteorological model is FAO-based crop specific water balance model (CSWB) used by Reynolds et al. (2000). In this method, near real time satellite products such as NDVI, RFE (rainfall estimate) images are used. The NDVI is derived from NOAA-AVHRR and RFE images are obtained from stationary Meteosat-5 satellite. Ground-based PET (potential evapo-transpiration data) from meteorological station was also used. This method incorporates remote sensing data with a ground-based model. The data are integrated in a GIS-based model called WINDISP3. This method also requires locally derived information such as yield reduction factor, maximum yield that differ spatially. The agrometeorological-based method has even been used by PECAD FAS to provide estimation on global agricultural production (Reynolds, 2001; NASA, 2003). PECAD's method is based on an automated decision support system called Crop Condition Data Retrieval and Evaluation (CADRE). CADRE is an operational outgrowth of the Large Area Crop Inventory Experiment (LACIE) and Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS). CADRE integrates remote sensing data, crop and soil models with weather information. It serves as an interface to different models and outputs data through GIS software, time-series plot and web interface displays. The agro-meteorological data to CADRE is provided by Agricultural Meteorological Model (AGRMET) and World Meteorological Organization (WMO) network of weather stations. AGRMET provides precipitation, minimum and maximum temperature, snow depth, solar and long wave radiation and potential and actual evapo-transpiration (ET). CADRE computes its own ET from temperature inputs using the Penman-Monteith equation. The satellite-based data include AVHRR and SPOT vegetation data. CADRE also requires baseline data which are digital elevation model which comprises of, FAO generated Digital Soil Map of the World (DSMW) information, historical crop production database from FAS, average temperature, rainfall spatial data, and administrative boundaries. The CADRE crop model requires crop calendar models, crop stress models (CERES, AgRISTARS, Maas, URCROP, Sinclair) and a two-layer soil moisture model. The two-layer soil moisture model runs the crop calendar and crop stress models. The soil moisture model accounts for the total water gained or lost in the soil profile by recording the amount withdrawn by evapo-transpiration and replenished by precipitation. The crop calendar model is based on the growing degree-days algorithm that uses minimum, maximum and threshold temperatures defined by a particular crop report. The crop stress model developed by AgRIStars informs analyst on abnormal temperature or moisture stress that may affect yields. Thus, PECAD uses a highly operational crop yield prediction system that requires an extensive input of time series data, baseline data and crop information and models from various sources.

1.2.3.4 Methods useful for regional predictions

Most of the methods reviewed have used NOAA-AVHRR-based NDVI images to estimate crop yield. The regression-based models may be used for regional yield estimations; however, it is highly empirical in nature. The three methods based on regression reviewed here (Boken, 2002; Rasmussen, 1998; Dabrawoska, 2002) are all empirical in nature and are locally based such that the models derived cannot be used for regional or global application. However, if a globally derived model can be obtained, regression-based yield models are the simplest of all models. According to Moulin et al. (Moulin et al., 1998), regression-based models based on vegetation index and yield are empirically derived; hence, not based on a theoretical and experimentally proved relation. Therefore, Moulin et al. (1998, p.1023) stated that, "more mechanistic and physiologically sound models are necessary to assimilate remote sensing data and to predict production of major crops". However, even though Montieth-based models are physiologically sound and experimentally proven, their uses in regional based application remain questionable. The Montieth-based models reviewed required values such as light-use efficiency values which were calculated using highly complex relations in Bastiaansen and Ali (2003) and using field based values as in Lobell et al. (2003). Thus, the applicability of such models in regional yield predictions whose variables cannot be computed regionally or globally need to be further studied. Contrary to the empirical methods and Montieth-based method, the agro-meteorological-based crop yield prediction method seems to have a good scope in regional yield predictions using remote sensing. The variables in these methods are mostly obtained from either meteorological stations or remote sensing satellites; thus, they have global or regional applicability. One such model used by PECAD is the ‘Sinclair’ model. Sinclair model has been used operationally by PECAD to estimate regional soybean yield.

1.2.4 Sinclair Crop Model

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, p.298). According to Speath et al. (1987, pp. 299-300) 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; and g) calculation of development rate (Sinclair, 1986). The meteorological data required for the model for its daily runs are obtained from ground stations. The model uses daily inputs of temperature (daily minimum and maximum), precipitation, solar radiation, day length and planting date (Figure 1.1). The temperature, precipitation, solar radiation data are usually obtained from meteorological stations. The planting date is estimated from reports and local knowledge. Besides these daily input variables, the model also requires initializing parameters for planting date and initial soil moisture conditions. The day length is calculated based on latitude. Due to its mechanistic nature, the model has potential to be effectively used in a geo-processing environment using remote sensing inputs.

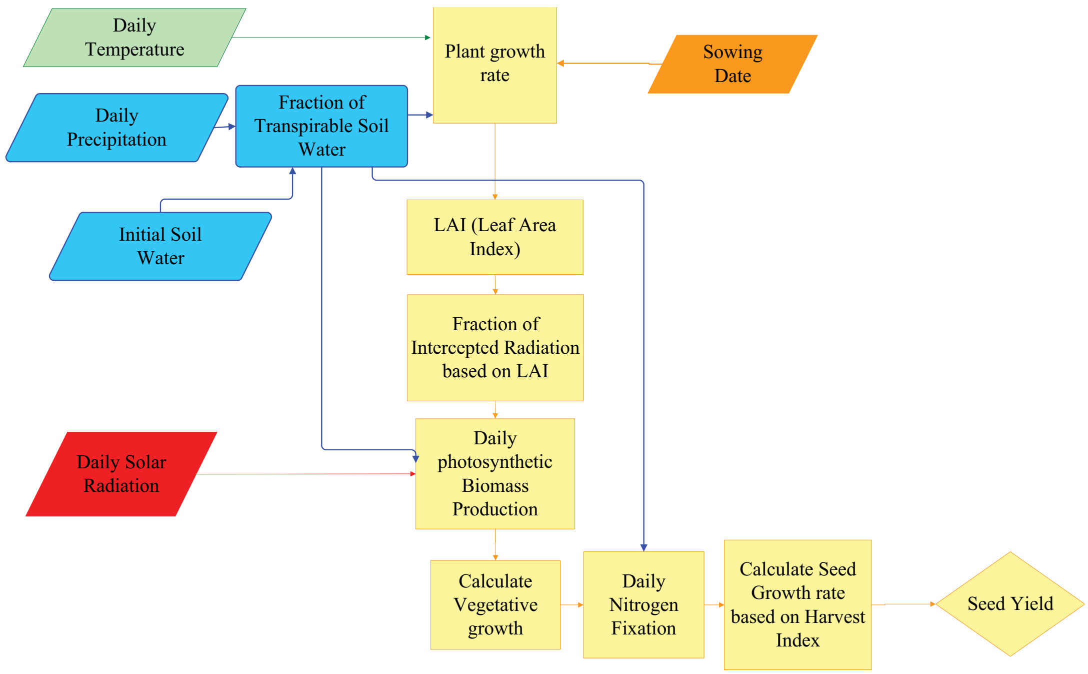


Figure 1.1 Abridged schematic flowchart depicting inputs and program flow of major modules for Sinclair soybean model.

The model requires initializing conditions for soil water conditions during sowing, sowing date and soil evaporation coefficient for no-till residue on soil surface. Sinclair et al. (2007) found that the model is highly sensitive to initial soil water content and the predicted yields are highly sensitive to the initial soil water conditions at sowing date. In the absence of a process for estimating initial soil water conditions, the simulations were initiated "at the harvest of the previous crop and assuming zero water content at that time (or even simulating the water use by the preceding crop if it might leave significant amounts of water in the soil)". The conclusions drawn from the above study by Sinclair et al. (2007) emphasized the need for methods for measuring or predicting accurate estimation method for initial soil water conditions.

Similarly, the daily meteorological input requirements for temperature, precipitation, and solar radiation drive the various modules such as daily calculation of leaf area growth, phenological stages, seed growth and daily water balance, which finally leads to the estimation of yield (Sinclair et al., 1986; Speath et al., 1987). These daily input variables are traditionally derived from ground meteorological stations. The output values are therefore only correct for a given area in which the meteorological values have influence. Utilization of gridded meteorological data sources can provide better capability to reduce the error in the modeling process due to spatial differences in location of the source of input variables and the location of the field to be modeled. Ultimately, the utilization of gridded datasets with spatial reference provides the opportunity to utilize the model in a geo-processing framework and also helps in improving the capability of the model to be used for regional level predictions.

1.3 Statement of Problem

Pre-harvest crop yield estimations and growth monitoring of crops can provide early warnings on status of crops that allows for making timely policy decisions for crop pricing as well as import and export quantities for major crops. Use of remote sensing has been utilized by many agencies and has been proved in different studies as an efficient way in comparison to the ground surveys that are required to collect crop-based information. Remote sensing is highly useful in providing a wide spatial view for regional-level crop estimation and monitoring. Review of various previous studies has shown that there are two basic methods through which such crop yield predictions and growth monitoring through the use of remote sensing is being performed: a) direct use of NDVI from satellite imagery to relate with biomass and yield; and b) use of remote sensing-based variables in a crop yield model as inputs to drive the processes.

Both of these methods are relatively new and still in the development phase as far as providing proven technological methods are concerned. In the methods where NDVI have been used, researchers have focused mostly on NOAA-AVHRR NDVI products. NOAA-AVHRR is in the process of being phased out and MODIS is considered as the immediate successor. However VIIRS, a new sensor is already being planned as a future successor to MODIS sensors. NOAA-AVHRR NDVI has been used by agencies worldwide to obtain wide-area information on crop status. In the US, USDA/FAS PECAD as well as USDA National Agricultural Statistics Service utilizes AVHRR NDVI datasets for their national as well as worldwide crop prediction and monitoring activities. Ultimately, in the future, these activities will need to be replaced with MODIS and then with VIIRS. However, a lack of research specifically on the applicability of MODIS for such crop monitoring and yield predictions were found.

With respect to the use of crop models for crop yield estimation and growth monitoring activities, the requirement of field level information to run crop models act as hindrance in providing accurate yield estimations in regional level. In most of the cases, highly detailed crop models that utilized a lot of field-based variables were not very effective for regional applications. However, use of crop models can greatly benefit accurate yield predictions and research that can make such applications possible is needed. Therefore for crop yield estimations, crop models that provide accuracy as well as require less information which can be easily obtained should be used. The semi-mechanistic nature of Sinclair model was found to be suitable for regional level adaptation. This model was originally developed for field-based analysis. However, the requirement of fewer field level inputs can provide an opportunity to utilize the model for regional level analysis and for spatially-based inputs that can provide crop growth and yield monitoring capabilities with fewer ground level data collection.

Therefore, based on the problem statement, this research aims are to look at two specific areas within the scope of regional crop yield and growth monitoring:

a. Efficient means of input variables to crop models and to research ways to integrate new sources of variables that are cost effective and time saving.
b. Applicability of MODIS data for crop yield prediction and monitoring as well as for the future sensor VIIRS that is expected to follow the heritage of AVHRR and MODIS.

1.4 Research Objectives

Based on the problem statement, the main research objective was to evaluate the use of remote sensing (current and future sensors) and grid-based meteorological datasets to reduce the need for field data and for providing improved predictive capabilities to monitor and model regional agricultural bio-productivity.

To fulfill the main objective, the study has three specific research objectives:

1. To test the ability of integrated grid-based meteorological datasets to provide inputs to crop model. In this study, the focus is directed on testing the South American Land Data Assimilation System (SALDAS) forcings data and soil

moisture to provide input to the Sinclair soybean yield model as a drop-in replacement to ground-based meteorological station datasets and to provide initializing soil moisture conditions.

2. To test the regional crop monitoring capability of MODIS, an existing 250 m resolution sensor, and to test the capability of VIIRS, a future 400 m sensor planned to replace MODIS, in providing continuity to crop monitoring application of MODIS. The study focuses on MODIS NDVI, an existing sensor product and VIIRS NDVI, a future sensor product. The baseline MODIS evaluation is used to compare against simulated VIIRS to verify capabilities of the future sensor product to provide future continuity to regional crop productivity monitoring capability that the existing MODIS sensor provides.
3. To test the ability of MODIS NDVI time-series data in estimating planting date for improving soybean yield predictions using Sinclair model. Planting date is a sensitive initialization parameter used in the model and is usually estimated due to difficulty in obtaining actual field data.

1.5 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.

Boken, V. K., and C. F. Shaykewich (2002), Improving an operational wheat yield model using phenological phase based Normalized Difference Vegetation Index, *International Journal of Remote Sensing*, 23: 4155-4168.

Bouman, B. A. M., H. van Keulen, H. H. van Laar & R. Rabinge (1996), The 'School of de Wit' Crop Simulation Models: A Pedigree and Historical Overview, *Agricultural Systems*, 52:171-198.

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*, 2:1109–1123.

De Wit, C. T. (1965), Photosynthesis of Leaf Canopies, *Agricultural Research Report 663*, PUDOC, Wagenigen, The Netherlands, 57pp.

Deering, D. W. (1978), Rangeland reflectance characteristics measured by aircraft and spacecraft sensors, Ph.D. Diss. Texas A&M Univ., College Station, 338pp.

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

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.

Moulin, S., A. Biondeau, and R. Delecolle (1998), Combining agricultural crop models and satellite observations: from field to regional scales, *International Journal of Remote Sensing*, 19:1021–1036.

Myneni, R. B., F. G. Hall, P. J. Sellers, and A. L. Marshak (1995), The interpretation of spectral vegetation indices. IEEE Transactions on Geosciences and Remote Sensing, 33: 481-486.

NASA (National Aeronautics and Space Administration) (2003), Decision Support Tools Evaluation Report for FAS/PECAD, Version 1.0'. [Unpublished].

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, 19:119–139.

Reynolds, C. A. (2001), Summary of Global Data Sources and Processing Models, Production Estimates and Crop Assessment Division (PECAD), Available online:http://www.pecad.fas.usda.gov/cropexplorer/datasources.cfm [Downloaded on: August 3, 2004].

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*, 21:3487–3508.

Sinclair, T. R. (1986), Water and nitrogen limitations in soybean grain production, *Fields Crops Research*, 15: 125-141.

Sinclair, T. R., L. R. Salado-Navarro, Graciela Salas, and L. C. Purcell, Soybean yields and soil water status in Argentina: Simulation analysis, Agricultural Systems, 94:471-477.

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.

Tucker, C. J. (1979), Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sensing of Environment*, 8:127-150.

Wiegand, C. L. and A. J. Richardson (1990a), Use of spectra indices to infer leaf area, evapotranspiration and yield: I. Rationale, *Agronomy Journal*, 82: 623-629.

Wiegand, C. L. and A. J. Richardson (1990b), Use of spectral indices to infer leaf area, evapotranspiration and yield: II. Results, *Agronomy Journal*, 82: 630-636.

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.

CHAPTER II

ANALYSIS OF SALDAS METEOROLOGICAL FORCINGS AND SALDAS SIMULATED SOIL MOISTURE FOR SOYBEAN YIELD ESTIMATION MODELING

2.1 Abstract

South American Land Data Assimilation System (SALDAS) is a part of the NASA/GSFC Global Land Data Assimilation System (GLDAS) project. In this research, a new meteorological grid-based data source provided by SALDAS was tested for its capability to provide inputs to Sinclair soybean yield prediction model. The Sinclair model requires daily inputs of meteorological variables for precipitation, solar radiation, and daily minimum and daily maximum temperatures. This model also requires initializing parameters for the soil moisture.

The hypothesis of this research is that the assimilation of SALDAS or similar data sources can benefit the yield modeling process significantly through simplification of the data collection effort for model runs. For the research analysis, SALDAS forcings datasets which have a similar geographic coverage and daily temporal resolution and 1/8 of a degree of spatial resolution for the precipitation, solar radiation, minimum temperature, and maximum temperature were utilized as drop-in replacements to ground meteorological data. The yield values, obtained from the Sinclair model runs using

traditional ground-based meteorological inputs, were used as the baseline result to compare with yield results from model runs with inputs from SALDAS forcings datasets. When the SALDAS data were tested with one on one replacement for Sinclair model runs, the yields values were comparable with yield values obtained from the use of ground meteorological datasets as inputs. However, when all the ground meteorological datasets were replaced with SALDAS inputs, the resulting simulated yields had a higher amount of deviation from the baseline yield values seemingly as a result of cumulative effect, as each variable input contributed to the deviation resulting in a larger observed total deviation of yield values. In the case of utilizing SALDAS soil moisture as initializing variables for water budgeting within the Sinclair model, the SALDAS soil moisture values showed good potential.

2.2 Introduction

The world today faces an increase in demand for agricultural production to supply the needs of an increasing population. For a geographic region, a reduction in any major crop's yield means potential food shortages. Information of crop yield forecasts before harvest can provide the administration the ability to make decisions for the prevention of such shortages. Crop yield prediction also has an additional economic importance; since excessive or deficient food imports are damaging to national interests. Imports that exceed actual demand can lead to economic damage whereas insufficient imports may lead to severe food shortages. Such yield predictions have become even more important with the advent of unstable rainfall patterns and other climatic variability due to global climatic changes (Reynolds et al., 2000; Bastiaanssen and Ali, 2003). According to

Lobell et al. (2003), timely regional crop yield predictions are also important to manage regional agricultural lands and to regulate regional food prices and trade approaches. Thus, reliable yield prediction methods are highly important for regional as well as national and global food security.

Yield predictions have traditionally been carried out using rigorous field-based assessments. The traditional method of obtaining crop yield information is considered to be highly time consuming, labor intensive and costly. These methods also have large data gaps resulting in inaccurate yield predictions even after extensive investment in labor and expenditure of large amounts of funds (Lobell et al., 2003 Reynolds et al., 2000; Bastiaanssen and Ali, 2003). Therefore, efficient ways of providing crop information before harvest are highly desirable. Although remote sensing-based methods have been used in the past (Rasmussen, 1998; Dabrawoska et al., 2002; Bastiaanssen and Ali, 2003; Lobell et al., 2003; Doraiswamy et al., 2005) for crop yield estimations, there is need for research to increase the efficiency of such methods allowing for a reduction of field data collection, reduction in inconsistencies in providing crop initializing parameters and for continued increase in the estimation and monitoring of crop condition and yield. In order to utilize remote sensing-based data inputs and methods, geo-processing friendly data sources and crop yield models need to be integrated together.

Mechanistic crop yield models have been found to be more efficient in providing regional level crop yield estimates. These crop models, however, require daily input of meteorological variables such as precipitation and temperature. The crop model utilizes the meteorological input variables, and outputs daily balances on soil water and crop growth parameters, which are then utilized for harvest index and ultimately crop yield.

Traditionally ground meteorological station datasets have been used to provide daily meteorological data inputs to the models. However, the distribution of ground meteorological stations is sparse and absent in many areas. National and regional agencies that require timely crop forecasts need estimates on international crop production levels as well. The lack of sufficient meteorological stations and their sparse locations create hindrances for efficient regional or global yield estimations.

2.2.1 Meteorological datasets in crop yield modeling

Many researchers have tested various meteorological inputs for regional yield predictions. 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 one-eighth of a degree for coupling with the crop model.

In cases of both local and regional predictions, the most popular source of meteorological datasets has been data from meteorological stations. One source of meteorological data is the National Climatic Data Center (NCDC) which provides daily meteorological datasets that includes precipitation, daily minimum and maximum 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.

A qualitative comparison of various available meteorological data sources is given in Tables 2.1 and 2.2. Comparison between some example sources of meteorological data sources from satellite images and ground stations with respect to their spatial and temporal distribution is given in Table 2.1. The comparative analysis shows that using a single source of data from a satellite can be problematic because each satellite source tends to focus on a single meteorological variable. For example, TRMM provides precipitation data for tropical regions; Meteosat provides precipitation and thermal datasets. These satellite meteorological sources also have varying spatial and temporal aspects. The local ground stations can provide most of the needs of meteorological data for agricultural modeling; however, the densities of meteorological stations are not distributed evenly in all regions. Another problem is the data collection effort required to collect the local ground station data for regional analysis. The lack of uniform distribution and sufficient density of meteorological station is one of the drawbacks of utilizing local ground stations for regional analysis.

Mechanistic crop yield models usually require daily inputs of precipitation, temperature, and solar radiation. Combining various satellite sources and collecting ground station data is a challenge for regional applicability. Therefore, integrated data sources provide a more viable source of meteorological data for inputs to crop yield models.

Table 2.1 Satellite based meteorological data sources.

INPUT DATA SOURCES	LOCAL GROUND STATIONS	GOES Satellite Systems	METEOSAT, TRMM	AVHRR, MODIS
Data	Temperature, Precipitation, Solar Radiation,	Precipitation	METEOSAT: Precipitation, Thermal TRMM:Precipitation	Land Surface Temperature
Resolution	Needs interpolation	4 km	2.5-5km	MODIS : 1km AVHRR LAC: 1km
Temporal cycle	Hourly, Daily, Weekly	Daily	Daily	Daily
Coverage	Depends upon countries	North and South America	METEOSAT: Europe/Africa/Indian Ocean TRMM: Tropics	Global

An example of utilization of integrated data source is PECAD's operational global crop assessment method. The system used by PECAD is based on an automated decision support system called Crop Condition Data Retrieval and Evaluation (CADRE). The agro-meteorological data input 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 have a resolution of 40 km. The vegetation datasets can be obtained in 250 m for MODIS and 1 km for AVHRR, but it is almost impossible to obtain the same resolution data for meteorological datasets. The use of 1 km 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 the 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).

Table 2.2 Integrated Data Sources.

INPUT DATA SOURCES	**NCDC (National Climatic Data Center)**	**USAF-AGRMET (Agriculture Meteorology model)**	**NASA-LIS (Land Information System)**
Source	Ground Met Stations	Integrated, Interpolated and Assimilated dataset	High-performance land surface modeling and data assimilation system
Data	Temperature, Precipitation, Solar Radiation,	Precipitation, Temperature, Soil Temperature, Soil Moisture, Evapo-transpiration etc	Precipitation, Temperature, Soil Moisture etc
Resolution	Needs interpolation	½ degree (~ 40 km)	1/8 degree
Temporal cycle	Hourly, Daily, Weekly	3 hourly,Daily	Daily
Coverage	United States	Global	Global

In this research, a new meteorological data source, Land Data Assimilation System (LDAS) is tested for its ability to provide daily meteorological variables as inputs to crop models as well as to provide initial soil moisture estimates. The research is based in Pampas in Argentina, one of the largest soybean producing regions of the world. Argentina's soybean production is one of the highest in the world and the prevalence of no-till agriculture as well as mostly rain-fed agriculture makes Argentina a good choice for soybean yield estimation studies. Large soybean fields in Argentina also provide a good opportunity to utilize coarser resolution imageries such as MODIS for field level validation. A simple mechanistic model, Sinclair soybean model, is used for this study. For South American region, South American Land Data Assimilation System (SALDAS) gridded meteorological datasets are tested. The model of SALDAS is similar to North American Land Data Assimilation System (NLDAS).

2.2.1.1 SALDAS forcing's as input to Sinclair model

Precipitation is one of the most important input parameters for running the Sinclair model. The model is highly sensitive to soil water as soil moisture affects several physiological processes leading to changes in 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 another input that is required daily for the model.

Through SALDAS, daily gridded inputs of SALDAS forcings for Sinclair inputs are available, which include daily minimum and maximum temperature, daily precipitation, and daily solar radiation. The SALDAS forcings data are from South American Regional Reanalysis (SARR) data from CPTEC/INPE (Centro de Previsão do Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais). SARR data is based upon the ETA Model, Regional Physical-space Statistical System (RPSS). These data are also utilized by SALDAS as forcings to provide soil moisture datasets (Larozza et al., 2007; Personal communications, Goncalves, 2007).

2.2.1.2 SALDAS for intializing soil moisture parameters

The Sinclair model requires certain initializing parameters during sowing time which are related to the water budget module. The model requires soil moisture at the planting date for initial soil moisture conditions at 15 and 90 cm depths. These initializing parameters are CSEVP, DYSE, DEEP, ESW and WH (water holding capacity). ESW is the water contained in the top 15 cm of the soil layer. CSEVP is the water evaporation coefficient based on the previous crop mulch. DEEP is the initial water in 90 cm depth. DYSE denotes days since last rainfall of more than 4 mm. WH is the water holding capacity. Detailed descriptions of the initializing parameters are given in the forthcoming pages. The main function inside the water budget model calculated by the Sinclair model is a daily calculated parameter called FTSW (fraction of transpirable soil water). The FTSW factor is coined by Sinclair and Ludlow (1986). FTSW factor is important for seed growth module, nitrogen budgeting, and leaf growth module. According to Sinclair (1986), the leaf area growth is restricted when FTSW is less than 0.2, and the leaf area growth stopped at the FTSW value of 0.05. The other physiological processes within the model are sensitive to the values of FTSW.

Sinclair et al. (2007) found that an accurate estimate of initial water content can help in predicting yield accurately. The study also pointed out the need for methods to either predict or measure initial water content. The study used predicted soil water content at sowing obtained by initiating simulations at the harvest of the previous crop and also by assuming zero water content at that time. However, for initiating simulations based on previous crops, information regarding the farming practices of a particular farm in study is necessary. At the same time, expert experience in handling simultaneous runs of crop models is also required. Therefore, this method may not provide efficient means for model runs especially for regional yield predictions.

Another study showed that soil moisture conditions may be estimated by running a 10 year model with information from farmers about the previous cropping and mulch type information (Salado and Sinclair, 2008). This method requires detailed information on crop rotations for previous years and an initial water budget condition. However, the availability of such detailed yearly cropping information for larger areas is not usually available. In practice, for local studies, initial soil water condition may be determined by local soil sampling and performing lab work needed to determine soil moisture for the field before planting. For large areas or regional implementation, direct sampling or long-term modeling to estimate soil moisture condition for setting initializing conditions in a crop model is highly inefficient. Again, in most cases of yield prediction modeling, the modeling is performed for off-site locations where field-based information is not accessible. In these cases, the unavailability of initial estimates for the initializing values of soil moisture conditions may be a highly limiting factor in providing accurate yield predictions. A more suitable method for measuring and/or estimating soil moisture conditions across a region is a practical necessity for implementing crop modeling larger areas or for regional level analysis.Therefore, in this study soil moisture data available from SALDAS is tested. If effective, soil moisture data sources such as SALDAS can be highly useful as an efficient source of providing initial soil moisture values in comparison to the current practice of utilizing a multi-year simultaneous crop modeling for water budget calculations.

2.3 Study Area

The study area is on the main soybean growing region of Argentina (The Pampas) and within the following coordinates (Upper left corner -65.2237, -31.089747; Lower right corner -60.091664, -35.0000). This includes the Cordoba region and other nearby areas (See Figure 2.1). From the eight different locations identified, various fields were selected with basic three crop rotation types: full season soybean after full season soybean, full season soybean after maize and double cropped soybeans after wheat. A near monthly field trip was taken to collect information including planting date, soil moisture, crop emergence and growth, crop residue information in addition to other relevant information important for the experiment.

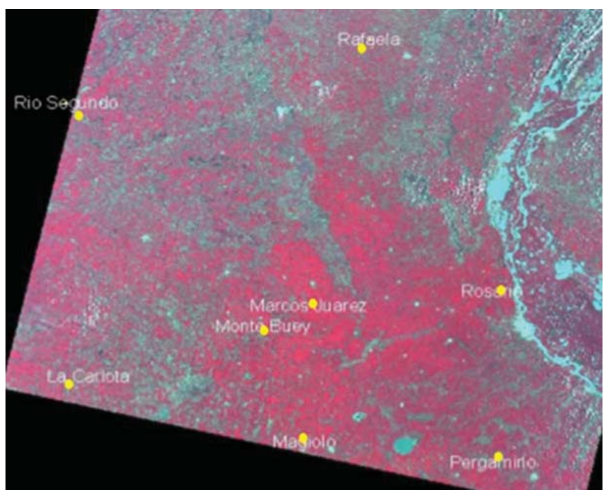


Figure 2.1 AWIFS imagery showing study sites selected for soybean farms and also for installing seven automatic weather stations (except Monte Buey).

2.4 Methodology

2.4.1 Data Collection

2.4.1.1 The South American Land Data Assimilation System (SALDAS)

SALDAS (de Goncalves et al., 2006) uses NASA's Land Information System (LIS - Kumar et al., 2006) as framework to run a suite of land surface models over South America. SALDAS is part of the NASA/GSFC Global Land Data Assimilation System (GLDAS) project. The SALDAS data were provided in one-eighth of a degree resolution. SALDAS utilizes meteorological forcings provided from CPTEC inputs to generate soil moisture, evapo-transpiration and other outputs. In this study, only soil moisture data from SALDAS is tested. The soil moisture values provided by SALDAS had data units in kg/m^2 for a given layer of soil moisture.

The soil moisture values are divided by range of soil layer. The total layer of soil moisture provided was 200 m in depth. At the same time, SALDAS forcings for daily minimum and maximum temperatures, total daily precipitation, and average daily solar radiation were tested as inputs to the Sinclair model meteorological data requirement (Figure 2.2). These data are also utilized by SALDAS as forcings to provide soil moisture datasets.

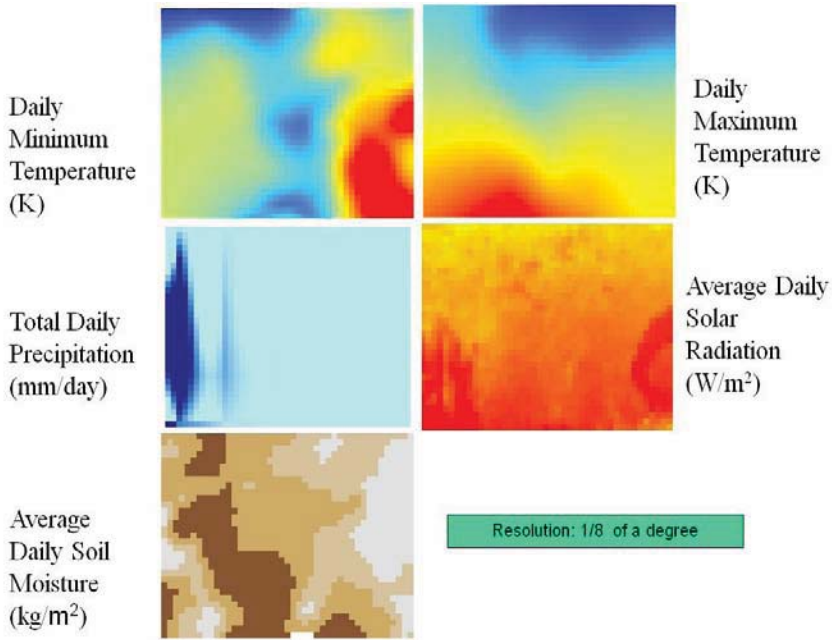


Figure 2.2 Gridded SALDAS meteorological forcings and soil moisture data.

2.4.1.2 Selection of Soybean Fields

Various soybean fields with large areas were selected in this study area of interest in Argentina. As applicable as possible (See Figure 2.3, Table 2.3), large soybean fields with at least 50 ha in area were selected. GPS points were obtained and field boundaries were delineated from those points using AWiFS, Google Earth and ArcGIS.

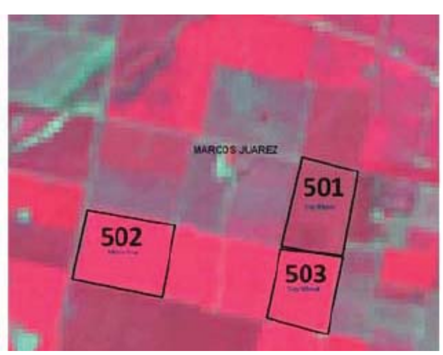


Figure 2.3 An example of fields selected for the study: Marcos Juarez fields over AWIFS imagery.

Table 2.3 Soybean fields used for the study for the year 2006/2007 in Argentina.

Field Location	Crop Rotation	Zonal Id	Area in Hectares
Rio Segundo	soy/soy	101	143
Rio Segundo	soy/wheat	102	110
Rio Segundo	soy/maize	104	77
Monte Buey	soy/wheat	401	42
Marcos Juarez	soy/maize	501	65
Marcos Juarez	soy/wheat	503	57
Magiolo	soy/wheat	601	109
Magiolo	soy/maize	602	117
Magiolo	soy/soy	604	161
Pergamino	soy/soy	703	54
Pergamino	soy/maize	704	81
Pergamino	soy/wheat	705	51
Rosario	soy/soy	801	60

2.4.1.3 Ground Meteorological Data

Ground meteorological stations were installed near the site of the field to collect daily rainfall, minimum and maximum temperatures, and solar radiation. The seven automatic weather stations, which included all of our sites except for MonteBuey, were installed in early November 2006, from which data are recorded every 15 minutes on air temperature, solar radiation and rainfall. These data are then aggregated to obtain daily datasets for each of the measurements. In case of areas where data are missing, the daily rainfall measured by the farmer, from the available locations, were collected. In most cases, the rainy days measured by the weather stations coincide with the farmer's observations, although the monthly totals showed some differences.

2.4.1.4 Soil Moisture Field Sampling

Soil moisture samples were taken from different soybean and maize fields. Samples were taken on each replication (three to four reps per field), up to a depth of 1.5 m, at 30 cm intervals. The available soil water was determined gravimetrically on each location by the Soils Lab of INTA Manfredi Experimental Station in Argentina.

2.4.1.5 Yield, Cultivars and Planting Date

Planting dates, cultivars, and crop yield for each field and farm samples were collected from the farmers.

2.4.2 Data Preprocessing

Daily SALDAS meteorological forcing datasets (maximum and minimum temperatures, solar radiation, precipitation, and soil moisture) for Argentina, South America were received in netCDF format. The data in netCDF format were converted to ASCIIGrid format using an IDL/ENVI script. These SALDAS forcings grid datasets were used for further geo-processing using a zonal analysis Arc-aml script with batch processing capability (O'Hara, 2008). For the selected zonal analysis script, the selected soybean farms were considered as zones, and were utilized to extract daily values of precipitation, minimum and maximum temperatures, and solar radiation. In the table below are the field zones used for the analysis and the code names given for zonal extraction. Each zone or field site was given a specific code which was used for zonal function within ArcGIS. The zonal function created a separate text output files for each zonal NDVI value for each day. These were rearranged into a spreadsheet for analysis using script for database manipulation written in Matlab by Shrestha (2008). The script extracted the mean zonal values from separate files created from Arc-aml zonal analysis function into an integrated text file in which the variable values were arranged for each day and was used as inputs to the Sinclair model (Figure 2.4).

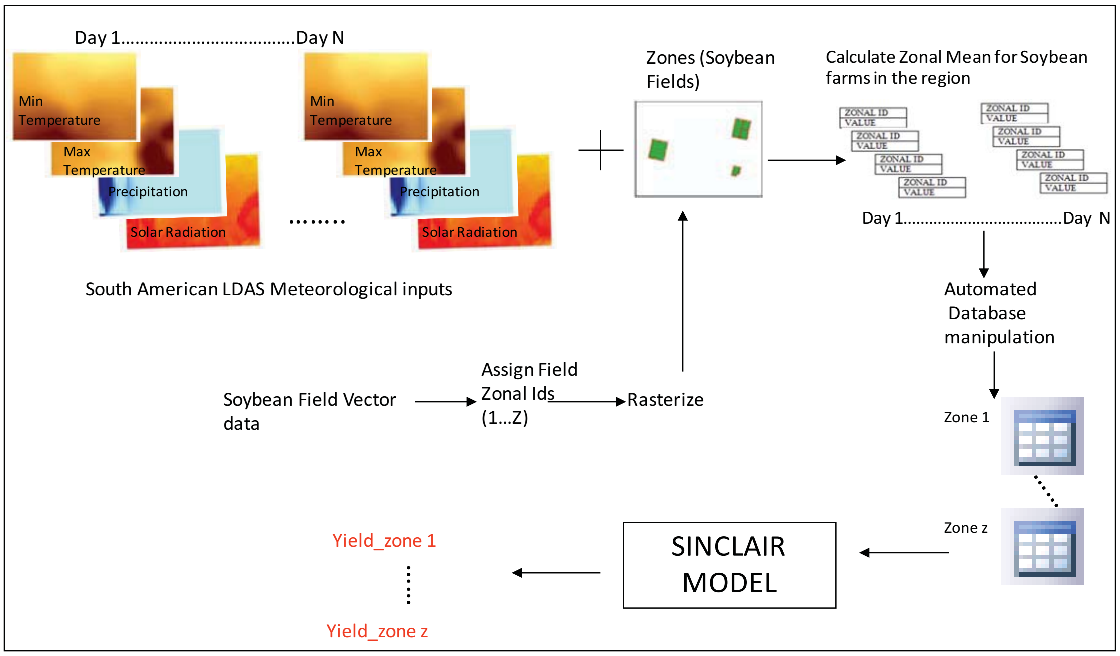


Figure 2.4 Geo-processing methods utilized to extract and process SALDAS grid datasets for providing "Sinclair model ready" inputs.

2.4.3 Use of SALDAS forcings meteorological input

This study incorporated a field data collection campaign which included compilation of crop plantings, varieties, yields, and a ground meteorological baseline database. A baseline model testing procedure for evaluating regional yield prediction was developed using the ground meteorological database for inputs to the Sinclair model and results were compared to observed yields. Results from the baseline model testing were compared to model results derived from the use of gridded SALDAS-based meteorological datasets as inputs to the Sinclair model.

To effectively use SALDAS and other spatial data, a geo-processing based framework was developed to manipulate and extract model-ready data from various spatially referenced grid datasets (See Figure 2.4). The SALDAS daily meteorological data were compared with ground-based meteorological data. SALDAS provided daily precipitation, minimum and maximum temperatures and solar radiation. These SALDAS variables were tested against the ground meteorological data by replacing one variable at a time and comparing the results against the base simulation results that contained all the input variables from the ground meteorological stations.

2.4.4 Use of SALDAS soil moisture as initializing inputs

The use of SALDAS soil moisture values to initialize inputs to the Sinclair model required conversion of units. The SALDAS soil moisture values were available in kg/m^2 for a given amount of depth. This value can be converted to mm per given amount of depth [$1\ kg/m^2 = 1$ mm]. For the purpose of this analysis, the SALDAS layers were combined for a column depth of 200 cm. In the field, soil moisture values were measured for selected study areas. The soil moisture values were determined gravimetrically on each location and converted to volumetric values. The soil moisture measurements were performed at the Soils Lab of INTA Manfredi Experimental Station.

Comparative analysis was performed between the field measurements of soil moisture and SALDAS simulated soil moisture for the month of November, in which soybeans were planted in most of the fields. Field conditions of very dry soils resulted in lack of soil samples for months before November. For the analysis, the field calculated measurement of volumetric soil water for 150 cm in mm of units were used. The volumetric fractional values were multiplied by the each depth of measurement (300 mm) and the summation of the values over the five layers of each 300 mm depth provided the volumetric content of soil water in mm for 150 cm depth. The SALDAS soil moisture values were in kg/m^2 for 200cm depth. For the data obtained for the year 2006, there were four layers of the following depths 100, 300, 600, and 1000 mm with the total depth of 2000 mm (200 cm). The SALDAS values in kg/m^2 could be directly converted to mm of water. The values were multiplied by 0.75 to obtain the representative fractional water content for 150 cm.

2.5 Results and Discussion

2.5.1 Yield results from ground meteorological inputs

The yield simulation analysis utilizing ground meteorological data for selected fields of Argentina *Pampas* in Sinclair model showed a majority of the predicted yield results were within 20% of the actual yield. All of the percentage differences had negative values, which is due to higher values of predicted yields than the actual yields obtained from farmers. Since, Sinclair model predicts potential yield without accounting for other stresses that might be present in the field such as soil nutrient, pests, and weeds as well as other management related issues such as crop row spacing; the higher values of predicted yields are an expected result. The overall percentage differences between the actual yields obtained from farmers and predicted yields with ground-based meteorological station data is within 30%. However, the majority of the fields have differences between actual and predicted yield within 15-20%. The yield obtained from ground-based daily meteorological datasets is considered as ‘Base Yield’ which is used to compare with the yield obtained from SALDAS forcings throughout the analysis (See Table 2.4).

2.5.2 Yield results from all SALDAS forcings

The daily ground-based meteorological inputs required by the Sinclair soybean model, which are daily minimum and maximum temperatures, solar radiation and precipitation values, were replaced with SALDAS forcings meteorological inputs. Although, a majority of the fields had predicted yield values within 30% of the actual yield compared to the farmers, most of the simulated yields had lower percentage difference. Only one farm had higher than 30% difference from actual yield when using ground meteorological data. Whereas the yield values obtained using the SALDAS forcings, five farms showed greater than 30% difference than actual yield (See Table 2.4).

Table 2.4 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield using SALDAS data.

Field Location	Crop Rotation	Zonal Id	Farmer's Yield (g/ m^2)	Simulated Yield [Base] (g/m^2))	SALDAS Yield(g/ m^2)	% Difference (Farmers Vs. Base Yield)	% Difference (Farmer Vs. SALDAS Yield)	% Difference (Base and SALDAS Yield)
Rio Segundo	soy/soy	101	460	514	350	-12	24	32
Rio Segundo	soy/wheat	102	400	453	326	-13	19	28
Rio Segundo	soy/maize	104	484	512	365	-6	25	29
Monte Buey	soy/wheat	401	340	388	489	-14	-44	-26
Marcos Juarez	soy/maize	501	468	493	482	-5	-3	2
Marcos Juarez	soy/wheat	503	401	467	399	-16	1	15
Magiolo	soy/wheat	601	368	455	527	-23	-43	-16
Magiolo	soy/maize	602	397	476	534	-20	-34	-12
Magiolo	soy/soy	604	395	484	524	-23	-33	-8
Pergamino	soy/soy	703	420	514	493	-22	-17	4
Pergamino	soy/maize	704	420	481	521	-14	-24	-8
Pergamino	soy/wheat	705	320	441	499	-38	-56	-13
Rosario	soy/soy	801	400	500	495	-25	-24	1

2.5.3 Yield results from SALDAS precipitation

In separate yield simulations, only the precipitation values from ground meteorological station data were replaced with SALDAS precipitation values. This was done to obtain a separate analysis for SALDAS precipitation. From the yield simulations performed, it was observed that the SALDAS precipitation performed moderately well. Comparison of yield values from SALDAS precipitation with yield values from all ground meteorological data showed that when the SALDAS precipitation was used, the differences were within 15% (Table 2.5). On comparing the farmer obtained actual yield values with the SALDAS precipitation yield values, percentage difference between the yield values were within 15% for six fields and the remaining seven fields were within 20-31%.

Table 2.5 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS precipitation.

Field Location	Crop Rotation	Farmer's Yield (g/ m^2)	Base Yield (g/ m^2)	Yield Simulated with SALDAS Precipitation (g/ m^2)	% Difference [Farmer vs. SALDAS Yield]	% Difference [Base Vs. SALDAS Yield]
Rio Segundo	soy/soy	460	514	452	2	12
Rio Segundo	soy/wheat	400	453	431	-8	5
Rio Segundo	soy/maize	484	512	461	5	10
Monte Buey	soy/wheat	340	388	445	-31	-15
Marcos Juarez	soy/maize	468	493	520	-11	-5
Marcos Juarez	soy/wheat	401	467	514	-28	-10
Magiolo	soy/wheat	368	455	458	-24	-1
Magiolo	soy/maize	397	476	519	-31	-9
Magiolo	soy/soy	395	484	511	-29	-5
Pergamino	soy/soy	420	514	441	-5	14
Pergamino	soy/maize	420	481	466	-11	3
Pergamino	soy/wheat	320	441	403	-26	9
Rosario	soy/soy	400	500	502	-25	0

2.5.4 Yield results from SALDAS minimum temperature

Another set of simulations were performed where only the minimum temperature from the ground meteorological datasets were replaced with SALDAS forcings minimum temperature dataset. The SALDAS forcings minimum temperature data performed quite well, with nine of the fields had yield differences within 30% of the actual yield obtained from farmers (Table 2.6). The comparison with ground meteorological data yield simulation results showed that twelve out of thirteen fields were within 10% difference. The remaining one field was also within 15% difference.

Table 2.6 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS minimum temperature.

Field Location	Crop Rotation	Farmer's Yield (g/ m^2)	Base Yield (g/ m^2)	Yield Simulated with SALDAS Minimum Temperature (g/ m^2)	% Difference [Farmer vs. SALDAS Yield]	% Difference [Base Vs. SALDAS Yield]
Rio Segundo	soy/soy	460	514	489	-6	5
Rio Segundo	soy/wheat	400	453	459	-15	-1
Rio Segundo	soy/maize	484	512	522	-8	-2
Monte Buey	soy/wheat	340	388	416	-22	-7
Marcos Juarez	soy/maize	468	493	479	-2	3
Marcos Juarez	soy/wheat	401	467	502	-25	-7
Magiolo	soy/wheat	368	455	489	-33	-7
Magiolo	soy/maize	397	476	510	-29	-7
Magiolo	soy/soy	395	484	486	-23	0
Pergamino	soy/soy	420	514	462	-10	10
Pergamino	soy/maize	420	481	486	-16	-1
Pergamino	soy/wheat	320	441	457	-43	-3
Rosario	soy/soy	400	500	525	-31	-5

2.5.5 Yield results from SALDAS maximum temperature

Similarly, another set of simulations was performed for the thirteen test fields, in which the maximum temperature dataset from ground meteorological stations were replaced by SALDAS forcings maximum temperature. This performed better than other meteorological forcings. Ten of the thirteen fields simulated yields within 21% of the actual yield (Table 2.7). In comparison to base yields with ground meteorological datasets, eleven of the thirteen fields had differences within 10%.

Table 2.7 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS maximum temperature.

Field Location	Crop Rotation	Farmer's Yield (g/ m^2)	Base Yield (g/ m^2)	Yield Simulated with SALDAS Maximum Temperature (g/ m^2)	% Difference [Farmer vs. SALDAS Yield]	%Difference [Base Vs. SALDAS Yield]
Rio Segundo	soy/soy	460	514	490	-7	5
Rio Segundo	soy/wheat	400	453	356	11	21
Rio Segundo	soy/maize	484	512	494	-2	3
Monte Buey	soy/wheat	340	388	375	-10	3
Marcos Juarez	soy/maize	468	493	503	-8	-2
Marcos Juarez	soy/wheat	401	467	402	0	14
Magiolo	soy/wheat	368	455	491	-33	-8
Magiolo	soy/maize	397	476	467	-18	2
Magiolo	soy/soy	395	484	519	-31	-7
Pergamino	soy/soy	420	514	483	-15	6
Pergamino	soy/maize	420	481	478	-14	1
Pergamino	soy/wheat	320	441	418	-31	5
Rosario	soy/soy	400	500	479	-20	4

2.5.6 Yield results from SALDAS solar radiation

Finally, in a similar manner the simulations were repeated for the SALDAS solar radiation values. Eight of the fields had yield values that differed by more than 20% from actual field values (Table 2.8). Three fields had yield values within 15-20%, and two fields had yield values within 15% of the actual yield. Comparison with the simulated yield from using all the ground meteorological inputs showed that nine of the fields had yield values within 10% and the rest were within 10 -20%.

Table 2.8 Comparison between Sinclair simulated base yields using with ground meteorological data and the simulated yield in which the ground meteorological input is replaced with only the SALDAS solar radiation.

Field Location	Crop Rotation	Farmer's Yield (g/ m^2)	Base Yield (g/ m^2)	Yield Simulated with SALDAS Solar Radiation (g/ m^2)	% Difference [Farmer vs. SALDAS Yield]	%Difference [Base Vs. SALDAS Yield]
Rio Segundo	soy/soy	460	514	534	-16	-4
Rio Segundo	soy/wheat	400	453	534	-33	-18
Rio Segundo	soy/maize	484	512	502	-4	2
Monte Buey	soy/wheat	340	388	437	-28	-13
Marcos Juarez	soy/maize	468	493	519	-11	-5
Marcos Juarez	soy/wheat	401	467	475	-18	-2
Magiolo	soy/wheat	368	455	511	-39	-12
Magiolo	soy/maize	397	476	563	-42	-18
Magiolo	soy/soy	395	484	512	-30	-6
Pergamino	soy/soy	420	514	520	-24	-1
Pergamino	soy/maize	420	481	491	-17	-2
Pergamino	soy/wheat	320	441	481	-50	-9
Rosario	soy/soy	400	500	511	-28	-2

2.5.7 SALDAS soil moisture as initializing inputs

For most of the sample sites, differences were seen in field soil moisture values and SALDAS soil moisture values, with SALDAS estimating much lower values. On average, percentage difference of 50 was obtained. The differences in the values are listed in the table below (Table 2.9).

Table 2.9 Comparison between SALDAS soil moisture values and field observed soil moisture.

Volumetric Soil moisture from fields for November for 150 cm depth	Field No.	Date field samples taken	SALDAS soil moisture (mm) for 200 cm	SALDAS soil moisture (mm) for 150 cm	Field vol. soil moisture (mm) for 150 cm	% Difference [Field vs. SALDAS values]
La Carlota Lot 3	301	2 Nov, 2006	181.50	136.13	239.40	43
La Carlota Lot 10	303	2 Nov, 2006	181.50	136.13	258.22	47
La Carlota Lot 9	304	2 Nov, 2006	181.50	136.13	218.50	38
Maggiolo Lot 27	604	2 Nov, 2006	273.32	204.99	442.57	54
Maggiolo Lot 40	602	2 Nov, 2006	269.18	201.88	391.11	48
Maggiolo Lot 41	605	2 Nov, 2006	270.53	202.89	388.80	48
Marcos Juarez Lot 23	501	3 Nov,2006	272.32	204.24	458.10	55
Marcos Juarez Lot 6	502	3 Nov,2006	272.32	204.24	453.20	55
Monte Buey Lot 1	404	3 Nov,2006	224.70	168.52	476.02	65
Monte Buey Lot 2	402	3 Nov,2006	224.70	168.52	476.55	65
Monte Buey Lot 3	403	3 Nov,2006	224.70	168.52	450.80	63
Pergamino Lot 3	703	1 Nov, 2006	350.15	262.61	527.70	50
Pergamino Lot 4	704	1 Nov, 2006	350.15	262.61	517.10	49
Rafaela Lot 103	202	4 Nov, 2006	352.79	264.59	434.65	39
Rafaela Lot 54	201	4 Nov, 2006	351.28	263.46	522.67	50
Rafaela Lot 3	205	4 Nov, 2006	352.79	264.59	453.10	42
Rosario Lot 4Y5	801	4 Nov, 2006	418.73	314.05	530.80	41
Rosario Lot 2	803	4 Nov, 2006	418.73	314.05	539.40	42
RioSegundo Circle A	101	6 Nov, 2006	455.57	341.68	357.00	4
RioSegundo Lot 32	104	6 Nov, 2006	455.57	341.68	356.85	4
RioSegundo Lot 39	103	6 Nov, 2006	412.72	309.54	212.20	-46
						Average= 50

2.5.7.1 Use of SALDAS moisture values for initializing Sinclair model

A small comparison test was performed to compare the volumetric soil water values calculated from SALDAS with field measured volumetric soil water values for a depth of 150 cm. The comparison of percentage difference between the field-based volumetric soil moisture values and SALDAS volumetric soil moisture values for 21 fields showed that the differences ranged from 37 to 64%. The three fields in Rio Segundo showed different pattern with respect to the other fields. Two fields in Rio Segundo had percentage difference of only 4.25 and one of the fields in Rio Segundo showed lower soil moisture values than SALDAS soil moisture. These three fields in Rio Segundo were considered as outliers.

The comparison was done for the month of November, which is the month of planting for most varieties of soybean. All the fields, except for the three fields in the Rio Segundo which were considered outliers, showed underestimation of soil moisture by SALDAS with an average value of 50%.

To utilize SALDAS soil moisture values to initialize the Sinclair model, a conversion process was utilized to convert the total daily average soil moisture values to the plant available water values. In the Sinclair model, all the values used for soil water initialization are in millimeters of water per given depth. The SALDAS values show promise for initializing ESW (variable that represents soil moisture at the first 15 cm of the top layer where evaporation takes place) and DEEP (variable which is the initial water in 90 cm depth).

The actual available soil water content for these initializing variables can be calculated by negating the permanent wilting point or its comparative replacement factor 'crop lower limit'. The permanent wilting point values were measured from the field samples for each site at the INTA Manfredi Station.

These values were averaged to obtain a single representative value for our area of interest. It is well known that permanent wilting point values differ in different soils but for this study, it was assumed that the soils for our area of interest were more or less homogenous. An average value of 0.15 (or 15%) permanent wilting point was obtained from all the field observations.

The SALDAS values were converted to fraction and plant available water in fraction was calculated by using the following formula:

$$\text{Plant available water} = \text{water content at field capacity} - \text{wilting point} \tag{2.1}$$

The fractional values of plant available water content were again converted to millimeters of plant available water content for a given depth. Using these values, the available water content for 15 and 90 cm depths were calculated for ESW and DEEP respectively. Sinclair model simulations were run to test whether SALDAS soil moisture values could be used to initialize the Sinclair model runs (Table 2.10). Two simulations sets were run for both ESW and DEEP values. The sets are as follows:

a. Simulations run for SALDAS obtained ESW values for plant available water in top 15 cm with expert provided DEEP values.
b. Simulations run for SALDAS obtained ESW values increased by 50% for plant available water in top 15 cm with expert provided DEEP values.
c. Simulations run for SALDAS obtained DEEP values for plant available water for 90 cm depth expert provided ESW values.
d. Simulations run for SALDAS obtained DEEP values increased by 50% for plant available water for 90 cm depth expert provided ESW values.
e. Simulations run for SALDAS obtained DEEP values as well as ESW values for plant available water.
f. Simulations run for SALDAS obtained DEEP values as well as ESW values increased by 50% for plant available water.

The results from the comparison to the base yield show that the percentage difference utilizing the ESW and DEEP values derived from SALDAS soil moisture is below 5% in majority of the fields. Even though SALDAS soil moisture values were observed to be 50% lower than field measured soil moisture values (as shown in Table 2.9), the utilization of ESW and DEEP values derived from SALDAS soil moisture did not produce much difference in the simulated yields.

From this result, we can safely say that that the Sinclair model is not very sensitive to ESW and DEEP initialization factors (Table 2.11). The model runs are more dependent on the accuracy of meteorological variables, especially rainfall data for the soil-water budget calculation. Therefore, as long as the daily rainfall meteorological inputs are accurate, the lower initialization values of ESW and DEEP are not detrimental to the simulation results.

However, too high initialization values for DEEP seem to have some influence in the water budget calculation. It can be observed from the case of the field in RioSegundo Lot 39, in which the SALDAS soil moistures was higher by 46% than the field measured soil moisture values and was considered as an outlier. However, utilizing the SALDAS values for initialization for this field, which was 46% higher, did not produce much difference in the yield. But when the original SALDAS DEEP values were increased by 50%, a much larger difference of 22% in the simulated yield was observed.

The root mean square error (RMSE) calculated between the yields from six sets of simulations using different ESW and DEEP values and base yield (Table 2.12) showed that the RMSE values were lower when only the ESW values were replaced with SALDAS soil moisture. The RMSE values increased when the DEEP values were replaced with SALDAS soil moisture. The results point out that the DEEP value initializations of soil moisture from SALDAS causes more deviated variation in yield results than utilizing ESW values.

Table 2.10 Yield from six sets of simulations using different ESW and DEEP values results obtained.

Location	Field ID	Yield 1: SALDAS ESW 15 cm (g/m^2)	Yield 2: SALDAS ESW 15 cm + 50% Increase (g/m^2)	Yield 3: SALDAS DEEP 90 cm (g/m^2)	Yield 4: SALDAS DEEP 90 cm + 50% Increase (g/m^2)	Yield 5: SALDAS ESW 15 cm + DEEP (g/m^2)	Yield6: SALDAS ESW 15 cm DEEP 90 cm + 50% Increase (g/m^2)	BASE YIELD (g/m^2)
Maggiolo Lot 37	601	468	455	451	459	451	459	455
Maggiolo Lot 27	604	485	480	556	505	556	505	484
Maggiolo Lot 40	602	475	476	482	479	482	479	476
Marcos Juarez Lot 23	501	492	520	482	477	482	477	493
Marcos Jurarez Lot 26	503	468	468	470	521	470	521	467
Monte Buey Lot A4	401	388	388	371	380	371	380	388
Pergamino Lot 3	703	515	514	473	511	473	511	514
Pergamino Lot 5	705	442	442	442	442	442	442	441
Pergamino Lot 4	704	481	481	478	479	478	479	481
Rosario Lot 4Y5	801	500	500	512	517	512	517	500
RioSegundo Circle A	101	514	514	507	519	507	519	514
RioSegundo Lot 32	104	512	512	504	519	505	519	512
RioSegundo Lot 39	103	453	453	436	355	436	355	453

Table 2.11 Percentage differences of the simulated yields with ESW and DEEP from SALDAS with base yields (simulated yield results using all experts provided initializing variables and ground meteorological datasets as inputs).

Location	Field ID	Base Vs. Yield 1 (%)	Base Vs. Yield 2 (%)	Base Vs. Yield 3 (%)	Base Vs. Yield 4 (%)	Base Vs. Yield 5 (%)	Base Vs. Yield 6 (%)
Maggiolo Lot 37	601	-3	0	1	-1	1	-1
Maggiolo Lot 27	604	0	1	-15	-4	-15	-4
Maggiolo Lot 40	602	0	0	-1	-1	-1	-1
Marcos Juarez Lot 23	501	0	-6	2	3	2	3
Marcos Jurarez Lot 26	503	0	0	-1	-12	-1	-12
Monte Buey Lot A4	401	0	0	4	2	4	2
Pergamino Lot 3	703	0	-7	8	1	8	1
Pergamino Lot 5	705	0	0	0	0	0	0
Pergamino Lot 4	704	0	0	0	0	0	0
Rosario Lot 4Y5	801	-3	-3	-2	-3	-2	-3
RioSegundo Circle A	101	0	0	1	-1	1	-1
RioSegundo Lot 32	104	0	0	1	-1	1	-1
RioSegundo Lot 39	103	0	0	4	22	4	22

Table 2.12 Root mean square error (RMSE) calculated between yield from six sets of simulations using different ESW and DEEP values and base yield.

RMSE SALDAS ESW 15 cm and Base Yield (g/m^2)	RMSE ESW 15 cm + 50% Increase and Base Yield (g/m^2)	RMSE SALDAS DEEP 90 cm and Base Yield (g/m^2)	RMSE SALDAS DEEP 90 cm + 50% Increase and Base Yield (g/m^2)	RMSE ESW 15 cm + DEEP and Base Yield (g/m^2)	RMSE ESW 15 cm DEEP 90 cm + 50% Increase and Base Yield (g/m^2)
3.66	7.58	24.63	32.44	24.61	32.44

2.6 Conclusions

The utilization of grid-based datasets such as SALDAS have many benefits for regional yield predictions. One major benefit is that such datasets can facilitate the process of geo-processing enabled efficient regional crop modeling. The other benefit is to provide a source of meteorological variables for any geographical region even in areas where meteorological stations are absent. Similarly, the usefulness of SALDAS type datasets depends upon the nature of the model as well. In this case, Sinclair model which is a semi-mechanistic model was used. The model depends upon daily input of meteorological variables and some initialization parameters for initial soil moisture conditions. The SALDAS forcings as well as soil moisture grids were applicable for use for both conditions. The daily availability of SALDAS datasets was compatible with the daily meteorological variable requirements of the Sinclair model. Therefore, crop models similar to Sinclair can benefit more from SALDAS type datasets. The geo-processing framework developed led to an efficient adaptation of the Sinclair model to use spatially referenced SALDAS grid datasets. A cell-based implementation of crop yield model may be desirable in the future. However, such cell by cell implementation maybe computationally intensive.

Sinclair model simulation of soybean growth and yield simulations were validated with the farmer's reported yields. The results showed the use of SALDAS inputs resulted in less accurate yield prediction, when compared with the yield prediction with respect to the inputs of ground-based meteorological datasets as well as actual yield. The deviation from the baseline yield values seemed to increase when all the ground meteorological datasets were replaced with SALDAS data values. A one on one replacement of ground based datasets with SALDAS seemed to provide yield results within a difference of 10% in most cases. Replacing all the ground meteorological datasets with SALDAS seemed to have a cumulative increase in deviation from the actual yields. The results were not unexpected considering the fact that each SALDAS pixel is $1/8^{th}$ of a degree and were utilized for field level yield predictions. A higher resolution grid datasets may be expected to provide a better accuracy to the yield results. Similarly, the SALDAS soil moisture initializing variables for soil water at 15 cm depth (ESW) and for 90 cm depth (DEEP) were tested. The analysis of SALDAS soil moisture values for initializing Sinclair model showed less sensitivity of the model to these variables. The ESW initializations showed less variation in the yield results than the DEEP initializations as observed from the results of the RMSE analysis. The results show that the SALDAS soil moisture values can be used as an efficient source for initializing soil moisture for model runs. Future work is recommended in utilizing SALDAS type grid datasets to be used for other crop models as well as for different geographical regions.

2.7 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: 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*, 2:1109–1123.

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

Goncalves, L. G. G. de, W. J. Shuttleworth, E. J. Burke, P. Houser, D. L. Toll, M. Rodell, and K. Arsenault (2006), Toward a South America Land Data Assimilation System: Aspects of land surface model spin-up using the Simplified Simple Biosphere, *Journal of Geophysical Research*, 111, D17110, doi: 10.1029/2005JD006297.

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 (2006), Land Information System - An interoperable framework for high resolution land surface modeling, *Environmental Modelling & Software,* 21: 1402-1415.

Larroza, E., L. G. De Goncalves, D. L. Toll, D. Herdies, E. Larroza, R. Borges, M. Bottino, J. Rozante, J. Shuttleworth, M. Rodell, B. A. Cosgrove, and J. Aravequia (2007), Retrospective and real-time forcing data for the South American Land Data Assimilation System initiative, AGU 2007.

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, 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.

O'Hara, C. G. (2008), Arcaml Zonal Processing Script, GeoSystems Research Institute, Mississippi State University, Mississippi State, MS.

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*, 19:119–139.

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*, 21:3487–3508.

Salado-Navarro, L. R. and T. R. Sinclair (2009), Crop rotations in Argentina: Analysis of water balance and yield using crop models, *Agricultural Systems*, 102:11-16.

Sinclair, T. R. (1986), Water and Nitrogen Limitations in soybean grain production, *Fields Crops Research*, 15: 125-141.

Sinclair, T. R. and M. M. Ludlow (1986), Influence of soil water supply on the plant water balance of four tropical grain legumes, *Australian Journal of Plant Physiology,* 13: 329-341.

Sinclair, T. R., L. R. Salado-Navarro, G. Salas, and L. C. Purcell (2007), Soybean yields and soil water status in Argentina: Simulation analysis, *Agricultural Systems*, 94:471-477.

Shrestha, B. (2008), Database Manipulation Script, GeoSystems Research Institute, Mississippi State University, Mississippi State, MS.

Vila, D. A., L. G. de Goncalves, D. Toll, and J. R. Rozante (2009). Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America, *Journal of Hydrometeorology*, 10: 533-543.

CHAPTER III

THE UTILITY OF MODIS AND SIMULATED VIIRS IMAGERIES FOR MONITORING CROP PRODUCTIVITY

3.1 Abstract

This research evaluates the ability of Normalized Difference Vegetation Index (NDVI) from Moderate Resolution Imaging Spectroradiometer (MODIS) and simulated Visible/Infrared Imager Radiometer Suite (VIIRS) imageries to monitor soybean crop bio-productivity. A geo-processing framework was employed to extract time-series Normalized Difference Vegetation Index (NDVI) values from selected fields that provided representative values of crop vigor for each crop type. These extracted daily time-series NDVI values were used to create NDVI time-series plot or curve to study the ability of MODIS NDVI to monitor, track, and evaluate soybean crop growth and phenological stages of development. The time-series plots obtained from MODIS NDVI were compared with predicted phenological events from the Sinclair model to cross-validate the Sinclair model and MODIS daily NDVI values. This served as a baseline to compare the simulated VIIRS with MODIS, in which simulated VIIRS daily NDVI values were compared in a similar framework. The results, from the cross-validation between MODIS NDVI and Sinclair model confirmed the ability of MODIS data to monitor crop growth conditions and also validated the ability of Sinclair crop model to

predict and simulate important plant phenological events. The comparison of MODIS with simulated VIIRS NDVI in the same framework showed that VIIRS data, although having coarser resolution than MODIS, is capable of providing the same level of performance as MODIS for crop growth monitoring capability with respect to providing NDVI time-series curves that conformed to soybean crop phenological events.

3.2 Background and Introduction

Satellite remote sensing has the ability to provide spectral information of the crop canopy and physical plant properties that helps to assess crop vigor and growth, which is related to crop yield (Lobell et al., 2003). Early research in this area found that the percentage of crop cover was closely related to vegetation indices (Tucker, 1979). In that research, when the crop cover increased or decreased the vegetation index values measured had a corresponding change. Thus, due to the observed relationship between the vegetation indices and crop development, crop conditions could be monitored through spectral measurements. Wiegand and Richardson (1990a, 1990b) in their research with different vegetation indices which included NDVI, Perpendicular Vegetation Index (PVI), and Red index, found that although limited, the vegetation indices do have relationships to crop growth and development and can be used to infer leaf area, evapo-transpiration and yield. Of the many vegetation indices, NDVI is the most commonly used index for monitoring vegetation growth.

NDVI is based on the properties of the plant that absorb light in the visible red wavelength and its inherent property that reflect in the infra-red wavelengths. One of the many applications of NDVI for agricultural applications has been in finding relationships between NDVI values and crop biomass and growth processes that can be related with yield. Research conducted by Boken et al. (2002) found that the average NDVI during the heading period correlated highly with wheat yield. Dabrawoska et al. (2002) concluded that there was a strong correlation between cereal yields and VCI (Vegetation Condition Index) calculated from NDVI and brightness temperature from National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) data in the weeks of 16 and 22 of crop growing period. Similarly, Rasmussen (1998) used a NOAA-AVHRR NDVI-based model for predicting crop yield in Senegal. In that research, regression analysis of the various data parameters was performed, and grain yield and time weighted NDVI values were highly correlated. These research efforts all indicate some correlation between NDVI during crop growth stages and crop yield.

Reed et al. (1994) developed a set of metrics from time-series NDVI from AVHRR for characterizing phenological states. The metrics characterizes various phenology states such as onset of greenness, duration of greenness, maximum NDVI, and end of greenness; which can be related to different vegetative photosynthetic activities as well as different stages of plant development (Kastens et al., 1998; Leeuwen et al., 2006). For operational uses, agencies such as Foreign Agricultural Service (FAS) at the United States Department of Agriculture (USDA), Production Estimation and Crop Assessment Division (PECAD) require global NDVI data in the operational process for their global decision making in crop condition assessments and predictions of yield before harvest. Currently, PECAD utilizes NDVI from AVHRR to fulfill their operational needs (Turner, 1998; Bethel and Doorn, 1998).

It is valid to note that in these referenced past research studies for regional monitoring of crop conditions, NDVI from NOAA-AVHRR has been predominantly used. NOAA-AVHRR has a spatial resolution of 1 km and can provide daily global reflectance datasets. The reason for using NOAA-AVHRR in many applications more specifically in regional crop monitoring is largely because it was historically the only available data source for large area monitoring that could provide cloud-free composited NDVI values. The AVHRR NDVI data are usually available as near weekly, bi-weekly or monthly composites. In the future, AVHRR will be discontinued and succeeded by the multi-agency sponsored MODIS and VIIRS sensors.

Since the launch of the NASA Earth Observing System (EOS) AQUA and TERRA satellites in December 1999 and May 2002, respectively, global daily reflectance datasets from MODIS sensor have been available with spectral characteristics similar to AVHRR in red and infrared bands but at a higher resolution of 250 m. MODIS data will provide continuity to the historic NDVI datasets that have been computed since 1979, when NOAA-6 was launched with the inclusion of visible and near-infrared (NIR) spectral bands (Ji et al., 2008). Compared to AVHRR, MODIS can provide better results due to its spatial resolution at 250 m. Although the MODIS NDVI product is considered as a continuity product for AVHRR NDVI, utilization of MODIS NDVI products are still limited and more research is required for validating the potential of MODIS in agricultural applications. Another sensor called VIIRS has been planned to collect visible and infrared imagery at 400 m spatial resolution onboard the US National Polar-orbiting Operational Environmental Satellite System (NPOESS) as a "successor" sensor for MODIS (Ji et al., 2008). The major difference between these sensors are their spatial resolutions and spectral ranges for visible and NIR bands. VIIRS has more similarity with AVHRR radiometry although at 400 m the spatial resolution lies between MODIS and AVHRR. The benefit of data sources such as MODIS is not only the spatial resolution that allows regional level analysis, but also the temporal resolution of daily data availability, which is a significant advantage for detecting crop damages due to sudden weather anomalies. MODIS, which is on both TERRA and AQUA platforms, can acquire spectral data daily two times a day; therefore MODIS provides the benefit of near real-time regional monitoring of vegetation growth stages, which is especially important for seasonal crops that require constant monitoring. In fact, both MODIS, and in the future VIIRS, can provide daily temporal coverage and their spatial resolution are effective for large area (regional and national) monitoring of crop conditions (see Table 3.1).

Table 3.1 Comparison between spatial, temporal and spectral resolution of MODIS, VIIRS and AVHRR.

Sensors	Temporal Resolution	Spatial Resolution	Spectral Ranges (FWHM bandwidth in nm) for Red and NIR (Source: Leeuwen et al., 2006)
VIIRS	1 day	400 m	Red: 600-680 nm NIR: 846-885 nm
MODIS	1-2 days	250 m	Red: 620-670 nm NIR: 841-876 nm
AVHRR (NOAA-17)	1 day	1 km	Red: 589-680 nm NIR: 734-988 nm

Therefore, this study has been undertaken with the objective of exploring, documenting, and understanding more fully the potential of MODIS datasets for monitoring crop growth processes. The research was designed to test the ability of MODIS 250 m NDVI to detect changes in soybean phenology at the farm level for the 2006/2007 soybean growing season in Argentina. Argentina's soybean production is one of the highest in the world with intensive soybean farming in the *Pampas* region. Additionally, the availability of large soybean fields in Argentina provide an opportunity to utilize coarser resolution imageries such as MODIS for field level validation and for obtaining unmixed soybean-based NDVI time-series data for each farm as much as possible within the constraints of the 250 m resolution of MODIS images. The sensitivity of the MODIS datasets to detect changes in NDVI at the farm level, will in large part, determine the level of effectiveness of MODIS datasets for regional applications. The ability of MODIS data to provide continuity for regional crop growth and yield monitoring applications in which AVHRR NDVI has been used successfully by crop scientists will also be tested. Data continuity is important for operational use of long-term NDVI datasets provided by AVHRR for agencies such as PECAD, as well as for the research community. MODIS and VIIRS have been identified as the sensors that can provide datasets with similar spectral and temporal characteristics to AVHRR datasets (Gallo et al., 2005; Ji et al., 2008; Leeuwen et al., 2006).

In this research, NDVI values from MODIS data are used to create soybean crop growth stage curves to study the ability of MODIS NDVI to monitor, track, and evaluate soybean crop growth stages of development by comparing observed values of NDVI with predicted phenological events from the Sinclair model. Since VIIRS is a planned future sensor, the research utilizes simulated VIIRS data as well. The study will help to determine the effect of resolution, in this case comparing the ground resolution of MODIS at 250 m with an effective resolution of VIIRS at 400 m, on the time-series NDVI values that can provide phenological growth characteristic curves for a particular crop during a growing season. The simulated VIIRS daily NDVI values will be compared with MODIS NDVI to test the effectiveness of future VIIRS vegetation index products by comparing the resultant time-series curve depicting soybean growth generated from each source to ascertain their similarities and differences, as well as the degree to which they are in agreement with ground-truth crop data and Sinclair model predictions for the growth process of the soybean crop.

3.3 Methodology

3.3.1 Study Area

The study area lies within the main soybean growing region of Argentina (The *Pampas*), which includes the Cordoba region and other nearby areas. The *Pampas* is a major soybean farming region within Argentina. For this study, various farms engaged in soybean farming were selected for field information on planting dates, and other pertinent ancillary information on farming practices. The major crops in our selected sites were wheat, corn, and soybeans which were planted in a rotational basis. For our study time of interest 2006/2007, large farms were selected to match the 250 m and 400 m resolution of MODIS and VIIRS, in which soybeans were planted for the 06/07 cropping season. These soybean farms were selected as our study areas of interest (AOI) and are within the following coordinates: upper left corner -65.2237, -31.089747 and lower right corner -60.091664, -35.0000. (See figure 3.1)

3.3.2 Data Sources

3.3.2.1 Moderate Resolution Imaging Spectroradiometer (MODIS)

AQUA and TERRA satellite platforms view the entire earth surface every 1-2 days and contain the Moderate Resolution Imaging Spectroradiometer (MODIS), which acquires data on 36 spectral bands at different spatial resolutions of 250 m, 500 m and 1000 m. For this research, only infrared and red bands are utilized, which are available at 250 m resolution. MODIS is designed to follow the heritage of the NOAA-AVHRR sensor series for providing long-term integrated measurements of the land surface (see http://modis.gsfc.nasa.gov). For the purpose of this research, the MODIS product MOD/MYD (TERRA/AQUA) 09GQK gridded data were used. The tile number H12V12 for this gridded dataset contained the area of interest.

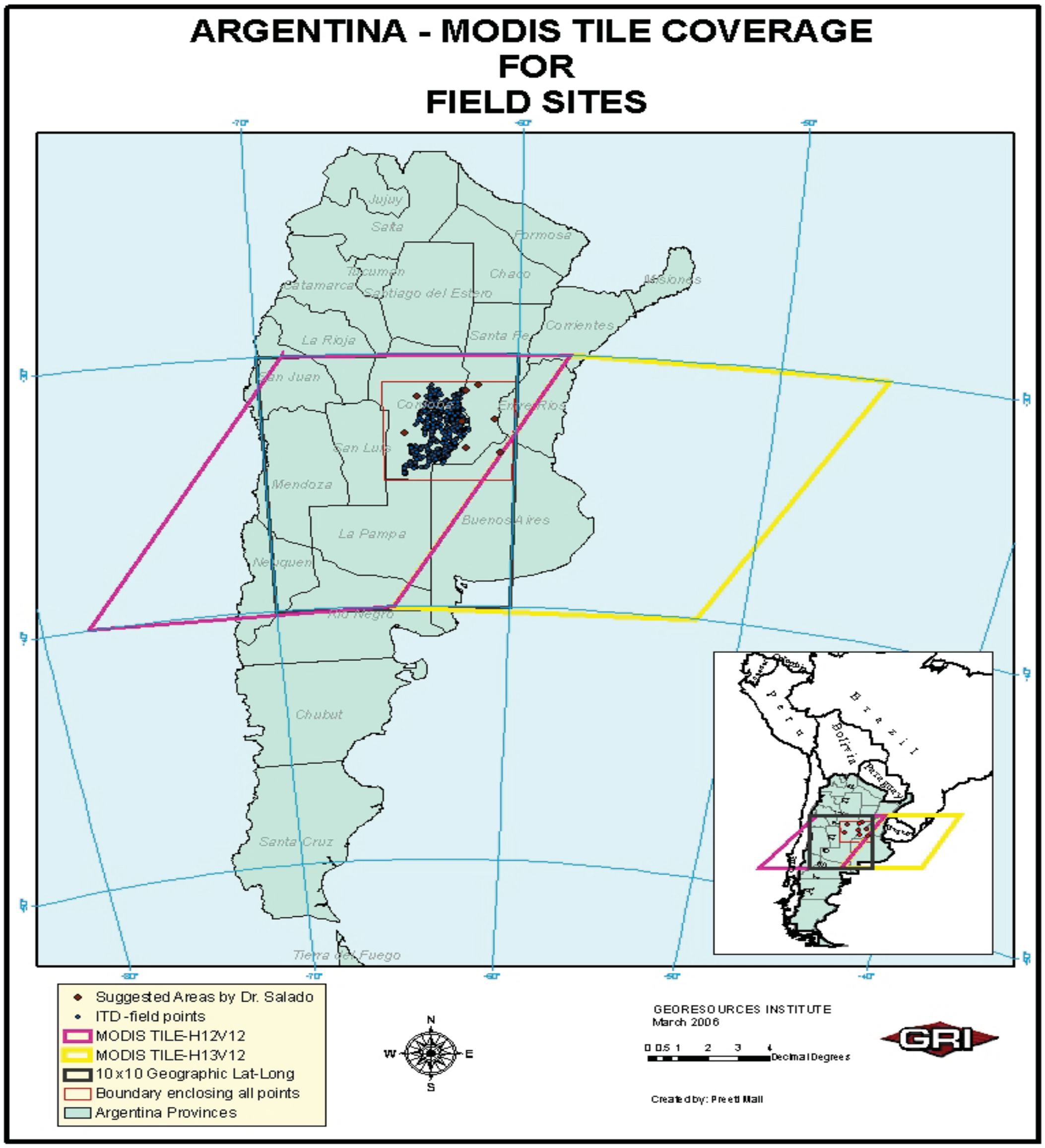


Figure 3.1 The AOI of our interest is shown by the red box inside the larger magenta colored box. The magenta colored box is the MODIS tile H12V12 used for this experiment.

3.3.2.2 Visible Infrared Imaging Radiometer Suite (VIIRS)

VIIRS is a future NASA sensor that is planned to serve as an operational follow-on for primary visible to thermal infrared sensing assets for both civil and defense communities that are currently using NOAA-AVHRR and Defense Meteorological

Satellites Program (DMSP) and MODIS. VIIRS onboard NPOESS Preparatory Project (NPP), with a spatial resolution of 400 m, will collect visible/infrared imagery and radiometric data (Lee et al., 2005). This study evaluates simulated VIIRS dataset to assess its use for agricultural efficiency applications. The MODIS reflectance data is used to obtain simulated VIIRS products using ART (Application Research Toolkit) developed by Science System and Applications, Incorporated (SSAI). The input MODIS datasets which have the resolution of 250 m are converted to resolution of VIIRS with 400 m after simulation.

3.3.3 Preprocessing

3.3.3.1 Field Boundary Delineation and Verification

GPS points for corners of soybean fields selected for the study were provided by Dr. Luis Salado Navarro, a research collaborator in Argentina. These points were used to delineate and demarcate field boundaries. Google Earth's high resolution imageries were used to place these GPS points in the correct field corners. At the same time, Advanced Wide Field Sensor (AWiFS) imagery for the location was also used for properly delineating the soybean fields. These points were then digitized as an ArcGIS polygon shape file. The digitized fields were then verified using land parcel maps obtained from farmers for each field and as well as from the first hand knowledge of Dr. Navarro.

3.3.3.2 Data Subsetting, Reprojection and Format Change

The ART toolbox had a functionality that accesses the functions of MODIS re-projection tool (MRT) which can read files in HDF format and re-project. The original HDF format MODIS data were in Sinusoidal projection. The MODIS red and infrared bands were extracted and re-projected to UTM projection, Zone 20S, WGS 84 and then subsetted utilizing the MRT application through ART encompassing the soybean field sites selected for our study. The HDF files were then converted to Geographic Tagged Image File Format (geoTIFF) files. Then, the subsetted images were utilized by ART toolkit to simulate images with the spectral and spatial resolution of VIIRS imageries.

3.3.3.3 VIIRS Simulation

The VIIRS images were simulated using the ART software. The algorithm requires the input of imagery with higher spectral resolution than the desired resolution of the simulated image (Zanoni et al., 2002). In this study, the desired simulated image was from VIIRS. VIIRS is planned to have a spatial resolution of 400 m and will contain visible and infrared bands. The MOD09GQK is the surface reflectance daily level 2G global 250 m datasets that were used to simulate VIIRS red and infrared bands. The main purpose was to assess the NDVI values of MODIS at 250 m for regional yield productivity monitoring and test the same with simulated VIIRS NDVI at 400 m resolution. The ART toolbox uses MATLAB functions and utilizes a MATLAB script to handle all file input and output paths as well as different configuration settings required for the simulations (Figure 3.2).

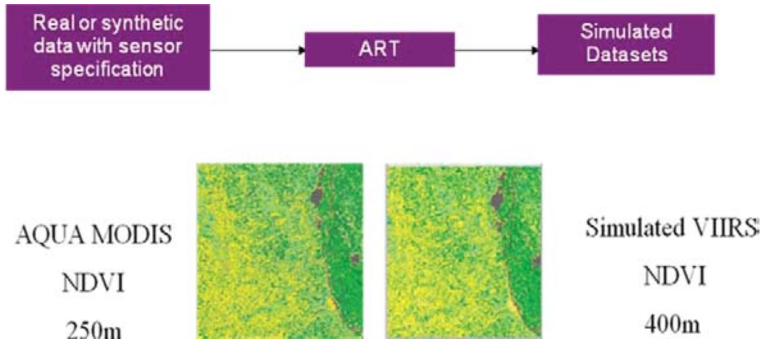


Figure 3.2 Simulated VIIRS NDVI (400 m) from MODIS data (250 m) using Application Research Toolbox (ART).

3.3.3.4 NDVI Calculation

The MODIS and VIIRS (obtained from simulation using the ART toolbox) imageries were then converted to NDVI images. The NDVI was calculated using the following relationship:

$$NDVI = NIR-R/NIR+R \tag{3.1}$$

where NDVI = Normalized Difference Vegetation Index, NIR = Near Infrared Band, R= Red Band. The NDVI value ranges from -1 to +1; the value increases from -1 to +1 with the increase in vegetation.

3.3.3.5 Zonal Value Extraction

The NDVI values were extracted using an Arc-aml script with batch processing capability that was created by O'Hara (2008). First, the NDVI files which were in GeoTiff format were converted to Arc grid format. The fields selected in Argentina from where field data were collected were digitized into polygon zones from the GPS points collected in the fields (Figure 3.3).

These fields were considered as zones that were used to extract the NDVI values for each day of the growing season for soybean in Argentina. Each zone or field site was given a specific code which was used for zonal function within ArcGIS. The zonal function created a separate text output file for each zonal NDVI value for each day. These were rearranged into a spreadsheet for analysis using script for database manipulation written in Matlab by Shrestha (2008). Table 3.2 shows the field zones used for the analysis and the code names given for zonal extraction.

3.3.3.6 Large Data Size and Volume Handling with Batch Processing

Since the time frame for the analysis consisted of multiple sources of data per day for a complete crop growing season for soybean, a large number of datasets had to be processed before it could be analyzed. For all the processing steps, batch processing was applied. The MODIS data were collected daily for both AQUA and TERRA MODIS. The data were processed starting from August 26, 2006 to July 15, 2007 based on the soybean growing time frame in Argentina. A flowchart of the pre-processing steps is given in the Figure 3.4.

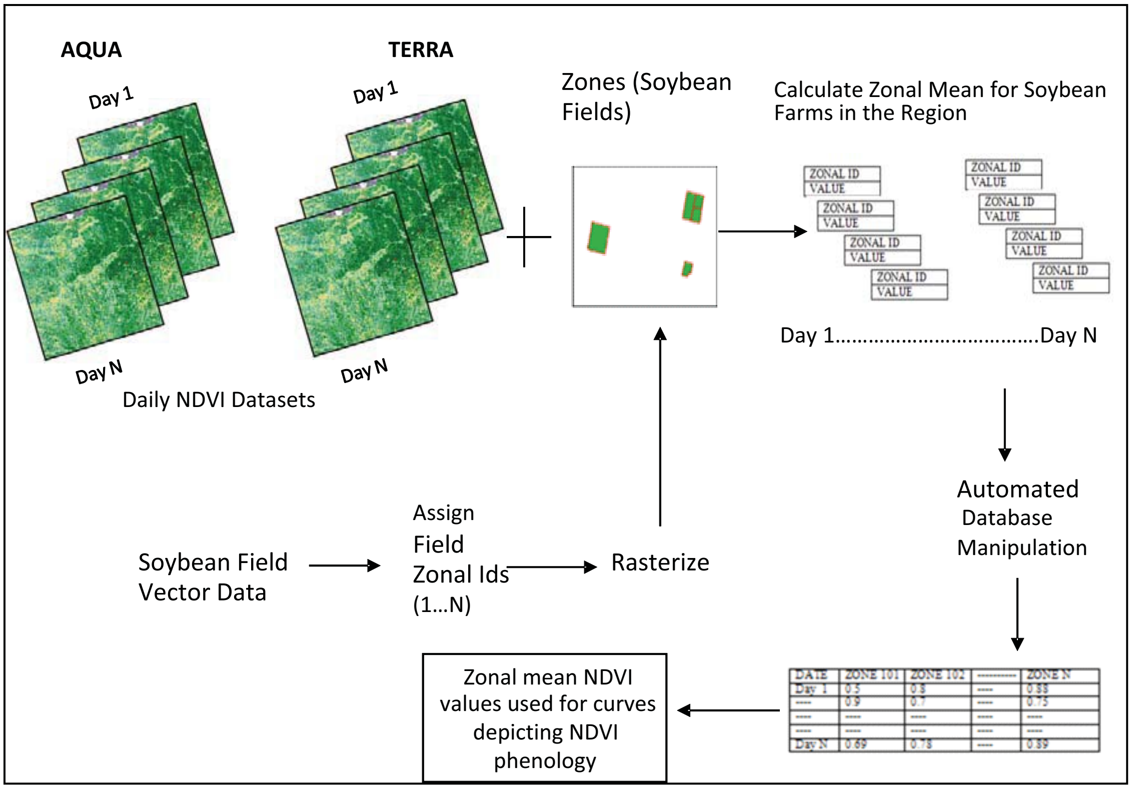


Figure 3.3 Flowchart of the geo-processing methodology used for the analysis.

Table 3.2 Soybean fields for the year 2006/2007 in Argentina.

Area Name	Zone ID (Field No.)	Crop Rotation
Rio Segundo	101	Soybean over Soybean
Rio Segundo	102	Soybean over Wheat
Rio Segundo	104	Soybean over Maize
Marcos Juarez	501	Soybean over Maize
Marcos Juarez	503	Soybean over Wheat
Venado Tuerto	601	Soybean over Wheat
Venado Tuerto	602	Soybean over Maize
Venado Tuerto	604	Soybean over Soybean
Pergamino	703	Soybean over Soybean
Pergamino	704	Soybean over Maize
Pergamino	705	Soybean over Wheat
Rosario	801	Soybean over Soybean
Rosario	803	Soybean over Maize
Rosario	804	Soybean over Soybean

The planting dates in Argentina, for soybean farming, start from October till December and the harvest time is from March until May. A few months of data before and after the growing season were also added for studying the NDVI changes before planting and after harvest. Therefore, for the purpose of analysis, 324 images for AQUA MODIS and another 324 images for TERRA MODIS were required. These images were then used to simulate VIIRS images for both AQUA and TERRA MODIS datasets. Hence, a total of at least 1296 images were processed before the final analysis could be performed. For decreasing computational time required for processing a large number of images, a 64 bit computer was used to store the data and process using ART/MRT tool for MODIS preprocessing and VIIRS simulation.

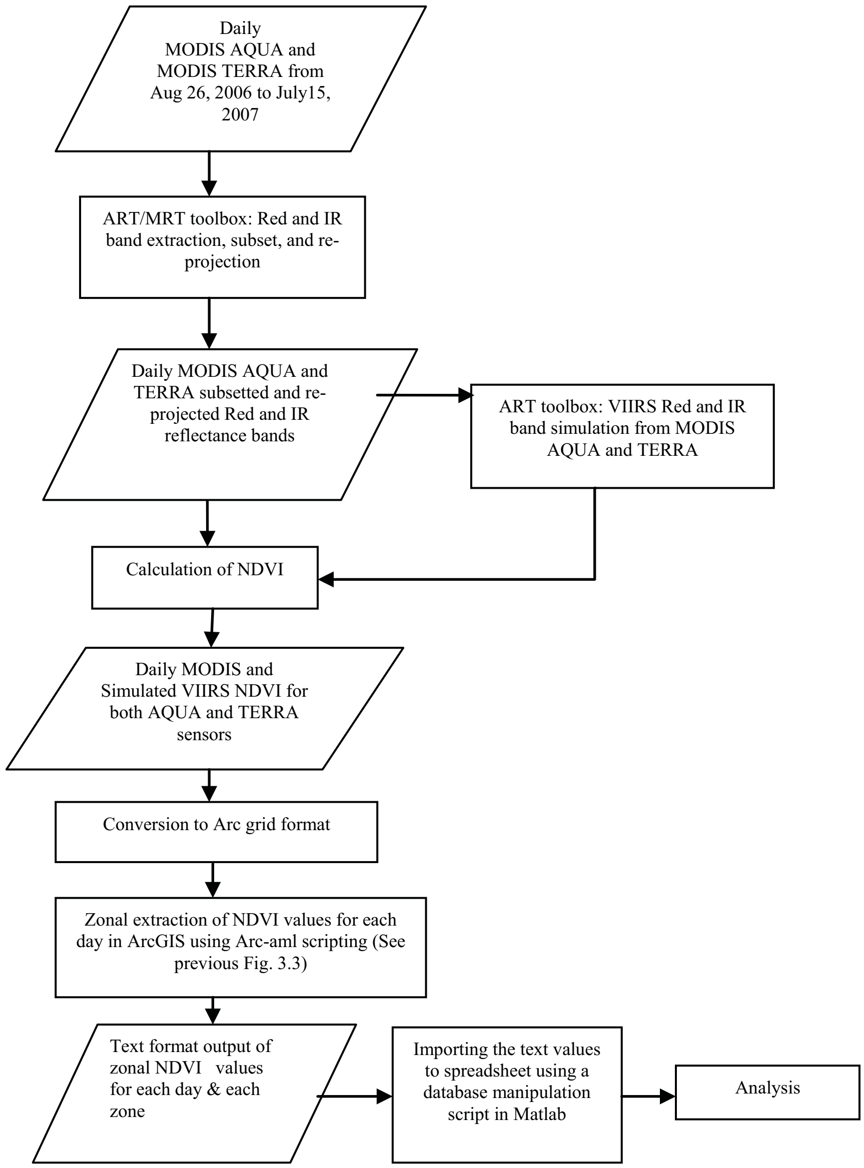


Figure 3.4 Flowchart of the pre-processing steps performed before analysis.

3.4 Analysis and Results

3.4.1 Comparison between MODIS and simulated VIIRS

3.4.1.1 Comparison using scatter plot

Scatter plots were created for 324 days of extracted MODIS and simulated VIIRS NDVI values (from both AQUA and TERRA MODIS), for the fourteen selected soybean field sites in Argentina. The extracted simulated VIIRS from both AQUA and TERRA MODIS showed similar characteristics for the fields which seemed to imply that the NDVI values from AQUA and TERRA MODIS were similar. The NDVI values from MODIS and simulated VIIRS complemented each other and demonstrated a linear relationship (See Figures 3.5 and 3.6). The squared correlation coefficient (R^2) values between the two datasets were high. In majority of the fields the R^2 was found to be 0.9 or higher except for the field 704 in Pergamino in which the R^2 was 0.86 for TERRA MODIS and 0.81 for AQUA MODIS. The exact reason for the lower values of the squared correlation coefficient (R^2) for this particular field is not known; however, the lower value for the field does point out that the simulation of VIIRS from MODIS does not always produce a nearly similar NDVI product of lower resolution at all times. The reason could also be due to some internal algorithm configuration of the ART toolkit for VIIRS simulation.

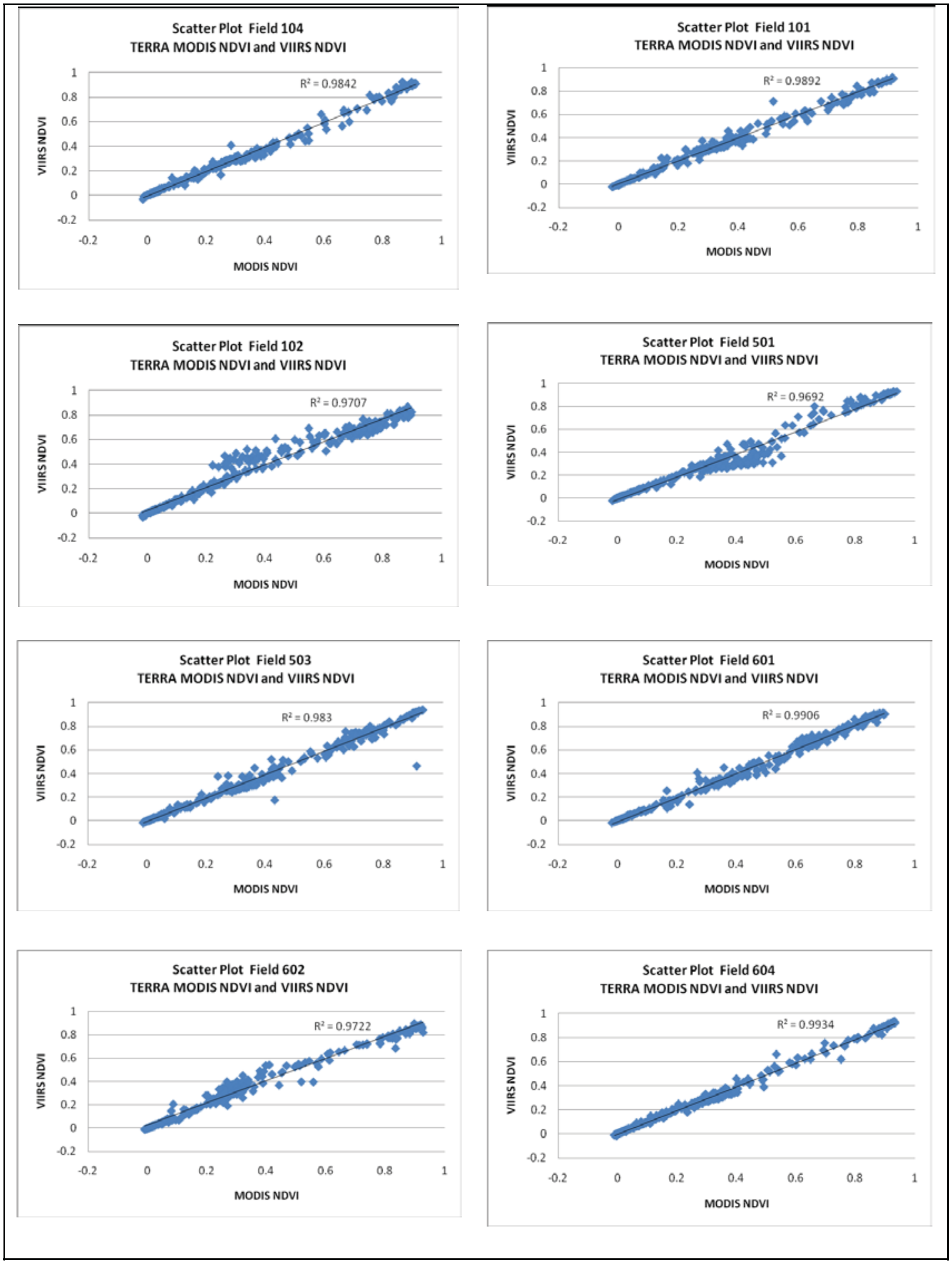


Figure 3.5 TERRA MODIS NDVI and simulated VIIRS NDVI scatter plots for 14 fields in the area for 2006-2007 soybean growing season.

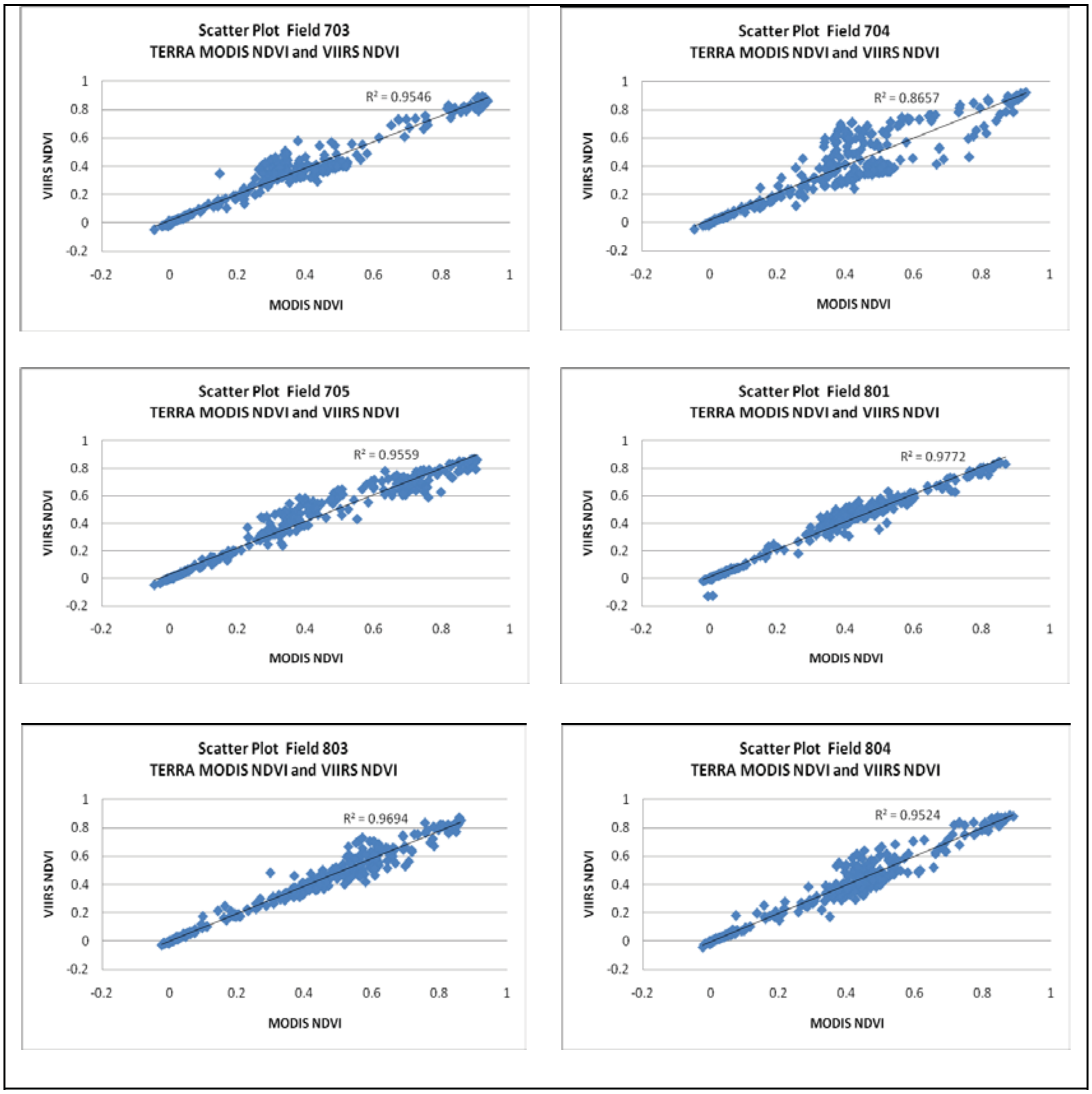


Figure 3.5 (Continued)

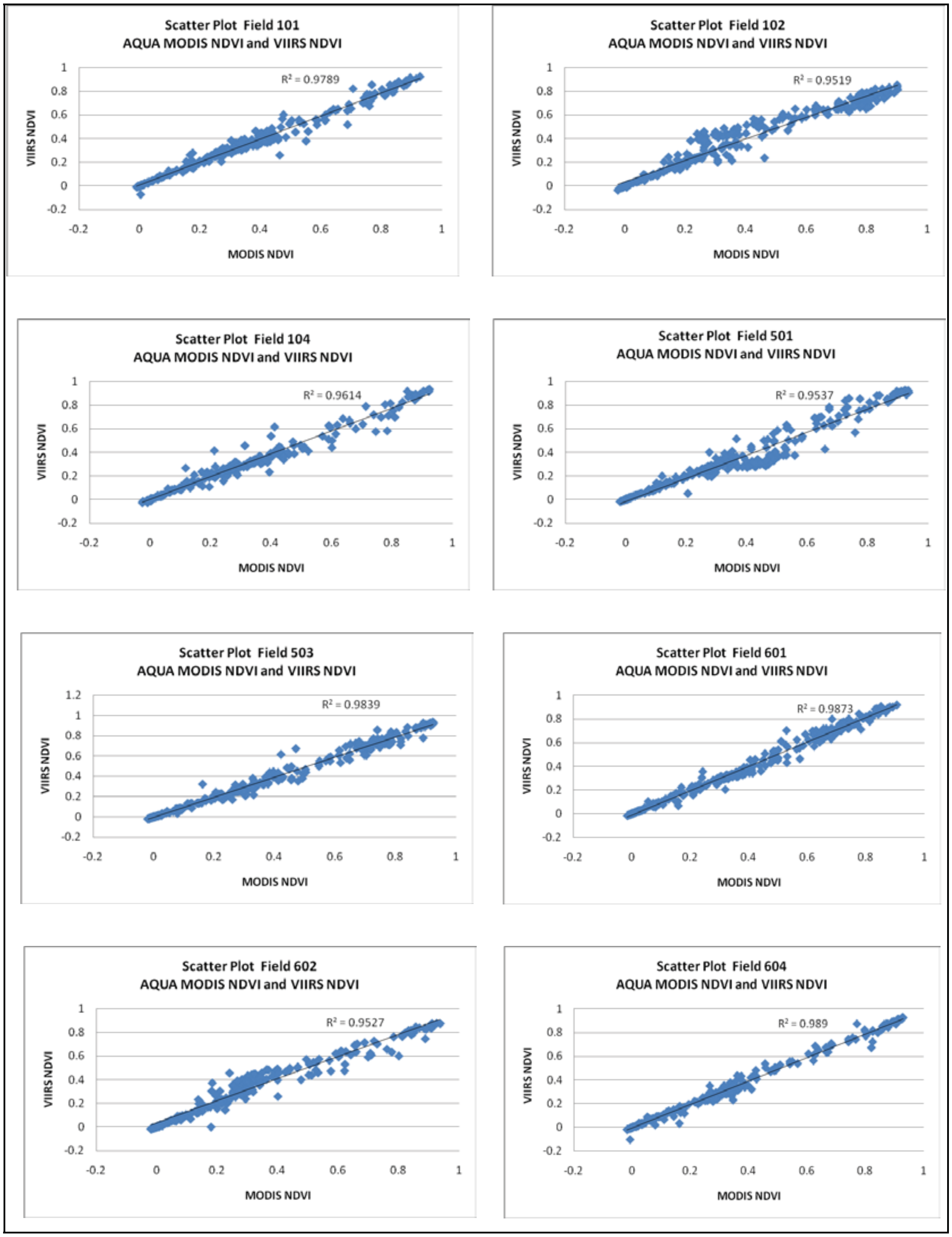


Figure 3.6 AQUA MODIS NDVI and simulated VIIRS NDVI scatter plots for 14 fields in the area for 2006-2007 soybean growing season.

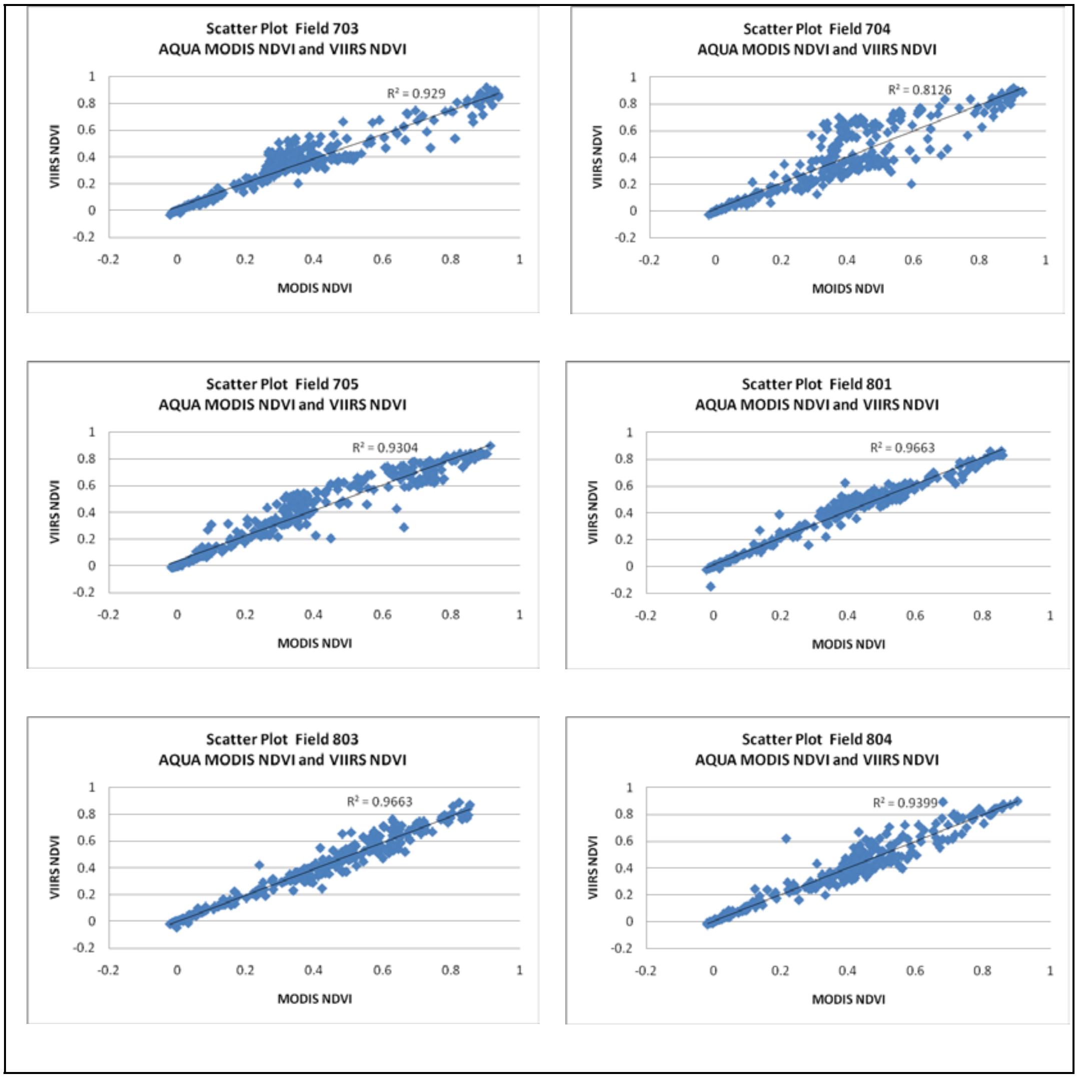


Figure 3.6 (Continued)

3.4.1.2 Validation of VIIRS simulation

The research design required VIIRS simulated from MODIS to be compared with MODIS. The simulated VIIRS from MODIS was compared against the VIIRS simulation from AWiFS imagery. The objective was to validate the simulated VIIRS from a different sensor of a higher resolution and radiometric properties. This validation process was required to reduce any bias in the comparison of NDVI from VIIRS simulated from MODIS with NDVI from MODIS.

AWiFS has a spatial resolution of 56 m. For the comparison, a cloud free AWiFS imagery was selected. The imagery on the acquisition date of December 26, 2006 was found to have no cloud cover. The simulations from AWiFS to VIIRS were provided by SSAI, Stennis Space Center, MS. Similarly, the VIIRS simulation from MODIS using MOD02 dataset for the same date was also provided by SSAI, Stennis Space Center. The NDVI values were calculated from the simulated images. Both the simulated VIIRS from MODIS and from AWiFS were projected in the same projection and subsetted to an area (552x476 pixels) with common coverage. Before comparison, one image was co-registered with the other image using image to image registration using ground control points. Thus, co-registered images were then compared using simple statistics and image to image correlation. The table below (Table 3.3) shows the statistics calculated between the two images:

Table 3.3 Statistical differences between VIIRS images simulated from MODIS and AWiFS.

Image	Min NDVI	Max NDVI	Mean NDVI	Standard Deviation
VIIRS simulated from MODIS (MOD02)	-0.38531	0.8563	0.5624	0.1851
VIIRS simulated from AWiFS	-0.5923	0.8741	0.5899	0.1937

VIIRS simulated from MODIS (MOD02) and VIIRS simulated with AWiFS was found to have a very high value of correlation coefficient (r) at 0.91. The statistical output (Table 3.3) and the correlation coefficient show a good agreement between the AWiFS simulated VIIRS and MODIS (MOD02) simulated VIIRS. Both of the compared simulated VIIRS were of 375 m in resolution and not 400 m resolution, the resolution that were used for the analysis. Since, the comparison was radiometric in nature and did not have a spatial aspect to it, the resolution difference of 25 m for the validation of the product was assumed to be not so significant with respect to the focus of the validation criteria. The simulated VIIRS from MOD02 were not atmospherically corrected. Similarly, AWiFS imagery was also not atmospherically corrected.

Therefore, the MOD02 NDVI product was compared with MOD09 NDVI product (atmospherically corrected). Any difference in the NDVI value could be considered as the difference due to atmospheric correction. For the comparison, four scenes of calculated NDVI from MOD02 reflectance values and MOD09 reflectance values were chosen within the days of soybean growing season in Argentina. The dates chosen were: Dec 26, 2006; Feb 17, 2007; Feb 2, 2007; and Jan 14, 2007. The images were subsetted to a smaller subset of 639 pixels by 597 pixels. Image to image correlation was calculated between the MOD02 NDVI and MOD09 NDVI imageries. The correlation coefficient in all the test dates were found to be higher than 0.8 (Table 3.4). The MOD09 atmospherically corrected NDVI had higher values than MOD02 NDVI which were not atmospherically corrected, which is a valid result as the atmospheric constituents are expected to reduce the reflectance values from the target. From this small test, we can infer that the atmospheric interference reduces the NDVI values by approximately 0.2.

Table 3.4 Correlation coefficient value between image to image comparison between MOD02 (atmospherically uncorrected level 1 image) vs. MOD09 (atmospherically corrected level 2 image).

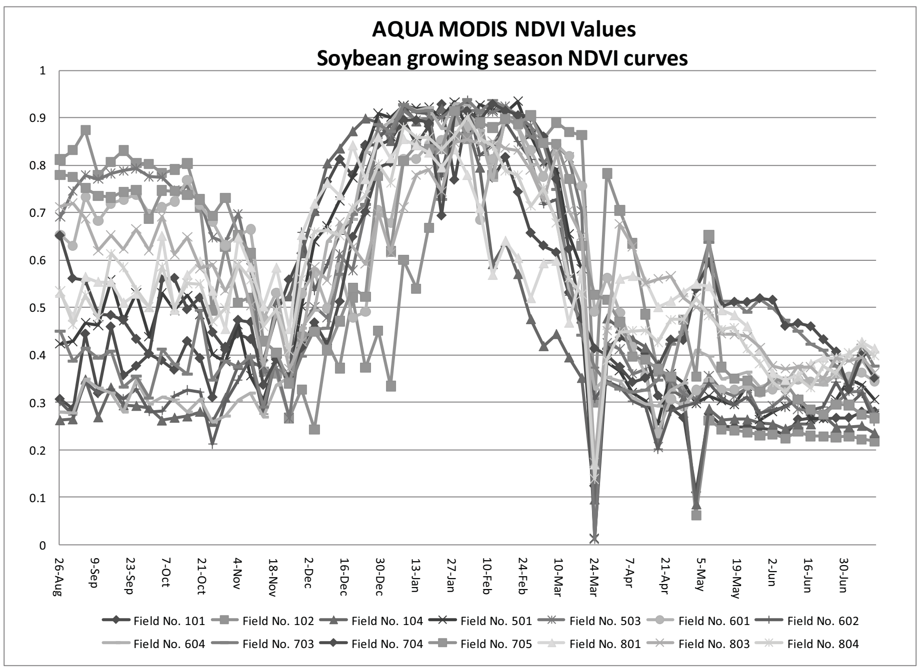
Image Date	Correlation Coefficient
2006_Dec 26	0.81
2007_Feb 17	0.81
2007_Feb 2	0.88
2007_Jan 14	0.97

3.4.1.3 Validation of time-series curves based on daily NDVI with respect to Soybean Phenology for MODIS and simulated VIIRS

In order to specifically test the effectiveness of the MODIS NDVI time-series data to monitor specific crop growth stages and to test the ability of VIIRS NDVI for its ability to replace MODIS data, time-series curves from both MODIS and VIIRS NDVI were created for selected soybean fields and compared with specific phenology dates for those fields. Among the dates of specific phenological stage required for the study, only the planting dates could be obtained from the farmers. The rest of the specific time value for the phenological stages such as emergence, reproductive stages, maturation etc. had to be computed as well as simulated from soybean growth simulation model called Sinclair.

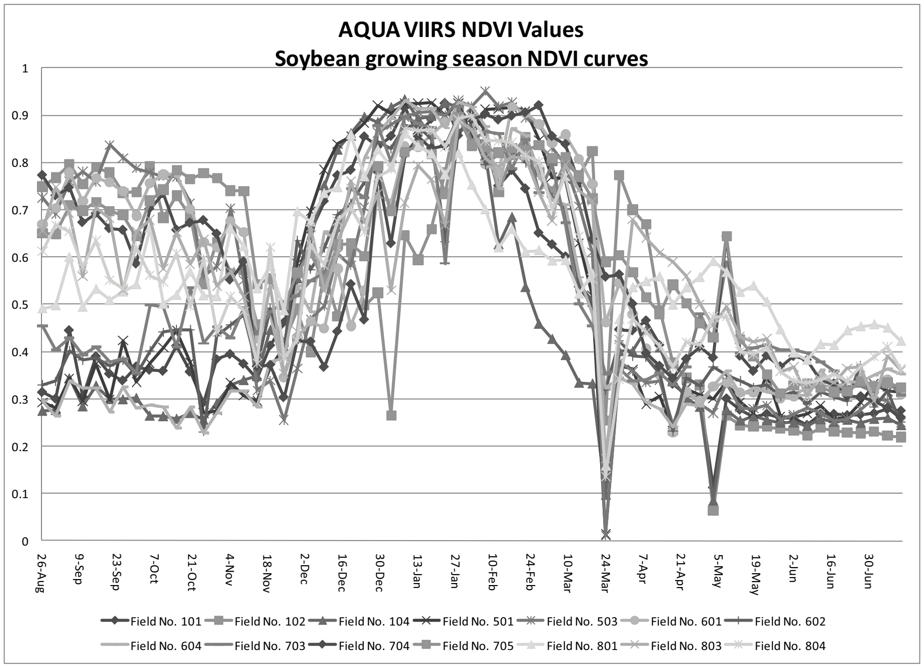
3.4.1.4 NDVI time-series depicting soybean growth

NDVI time-series curves were created for both daily MODIS NDVI and VIIRS NDVI. Since the zonal extracted NDVI values were from images without any filtering or compositing process for different fields for the soybean growing season, cloudy pixels were also present. NDVI values are sensitive to water and presence of clouds in the images lowers the NDVI values. Therefore, in order to remove the NDVI values that were affected by cloud reflectance a macro language was created in Excel to select only the maximum NDVI values within a five-day period. This resulted in smoother NDVI time-series curves. The five-day maximum NDVI values from August 26, 2006 to July 15, 2007 were plotted for AQUA MODIS, TERRA MODIS, AQUA simulated VIIRS and TERRA simulated VIIRS. Still some values with low NDVI compared to days before and after were detected. These values were the result of presence of cloud cover remaining for more than 5 days. These values were considered as outliers, thus removed. Outliers were identified as values that had significantly different values than the values of the neighboring days. Some values less than 0.001 were detected at a point in the time-series when all the neighboring values were in the range of 0.9, thus were removed. And some ‘NODATA’ values (-9999) as well as values that were out of NDVI range of ‘-1’ were also detected and removed. Other than these removals, the data was not filtered or smoothed and the raw NDVI values were used for analysis and for the creation of NDVI time-series curves. The time-series plots thus created, depicted the growth pattern of soybean growth in Argentina. The time-series plots also showed similarities between MODIS NDVI temporal curve and VIIRS temporal curve. Both the MODIS and simulated VIIRS curves were able to show phenological characteristics of the soybean crop growth stages (See figures 3.7 a,b,c,d).

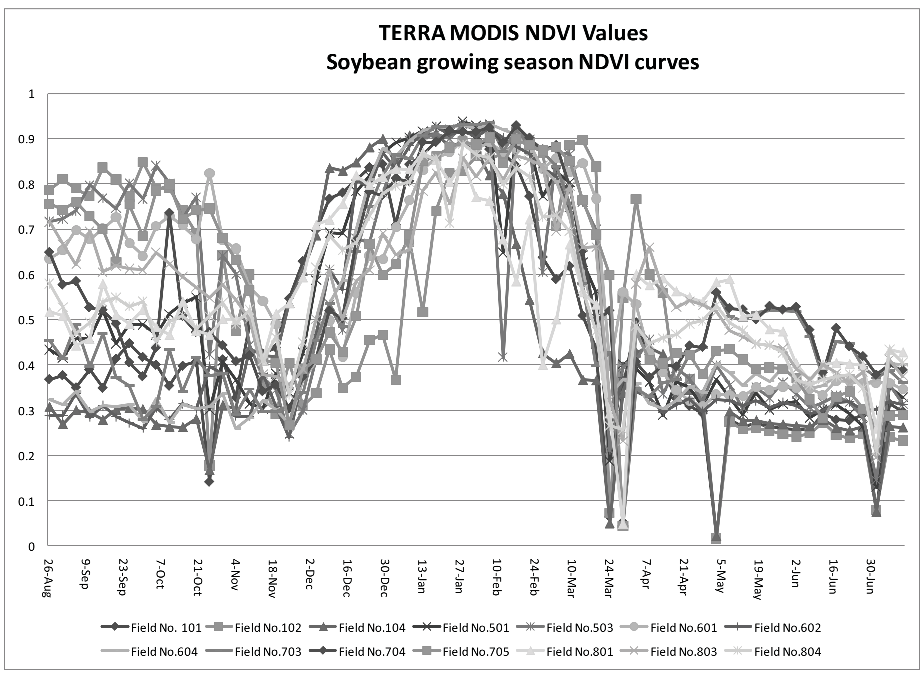


(a)

Figure 3.7 NDVI curves depicting soybean growth characteristics for the fourteen soybean test fields in Argentina from AQUA MODIS (a), VIIRS simulated from AQUA MODIS (b), TERRA MODIS (c), VIIRS simulated from TERRA MODIS (d).

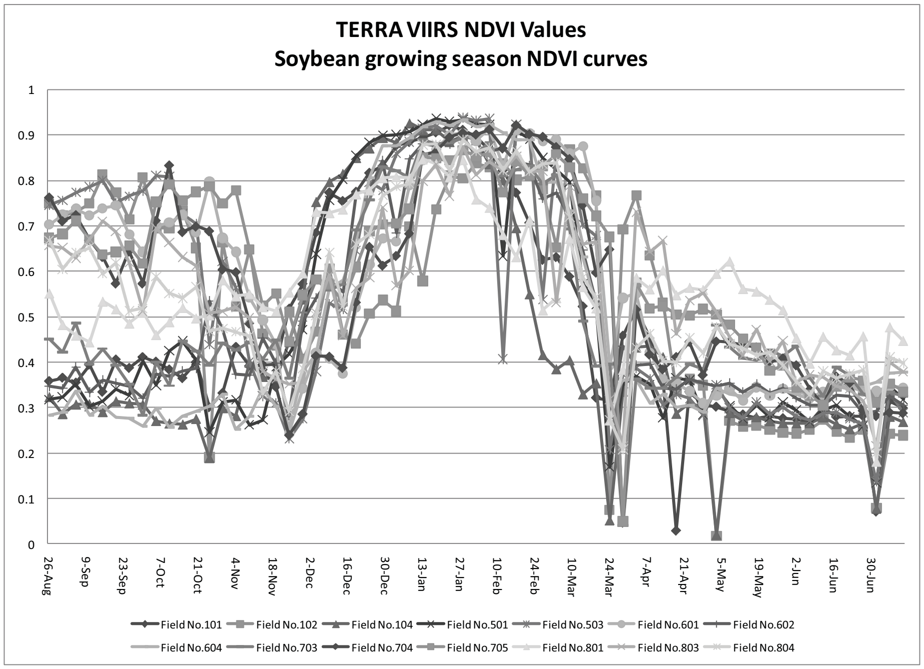


(b)



(c)

Figure 3.7 (Continued)



(d)

Figure 3.7 (Continued)

In the figures (3.7 a,b,c,d), the time-series curves for the 14 fields were able to depict the soybean growth curve for the Argentine soybean growing season. The planting time frame for soybeans is from October until mid-December. In that time frame all the soybean fields, from which the NDVI values were extracted, showed a clear decrease and then a significant linear increase in NDVI values and then reached a plateau stage around January. At the time of mid-March, the NDVI values start to decrease again. This behavior of time-series NDVI values corresponded with the soybean growth and maturation time in Argentina.

The NDVI time-series curves for the soybean fields also showed similarities to the metrics characterizing vegetation phenology developed by Reed et al. (1994). The time-series curve could be related to the temporal metrics and the NDVI-value metrics demonstrated by Kastens et al. (1998) after Reed et al. (1994). The four major stages in temporal metrics can be related to specific time of soybean phenological stages which is as given in the table (3.5) below:

Table 3.5 "Temporal Metrics" (Reed et al. 1994) and the corresponding phenological stages.

Temporal metrics	Phenological Stages
1. Time of onset of greenness	Emergence
2. Time of end of greenness	Maturity
3. Duration of greenness	Emergence to Maturity
4. Time of maximum greenness	Time of optimum growth after which no leaf development occurs can be considered the time up to the termination of leaf growth starts.

3.4.1.5 Cross Validation Approach

A cross validation approach was implemented to compare the MODIS and VIIRS NDVI time-series curve with the Sinclair model simulated crop growth stages and vice versa. Planting date, emergence date, termination of leaf growth and maturity dates were chosen as specific growth stages at the specific area in time-series curves based upon the temporal metrics developed by Reed et al. (1998) (See table 3.5).

Since planting dates were the only specific dates obtained from the farmers for the selected soybean fields; based on the planting date, the emergence dates were estimated using growing-degree-day (GDD) concept. The value of 150 GDD was considered as an optimum value of degree days required for soybeans to emerge after planting, based upon personal consultation with Dr. Thomas Sinclair (author of the Sinclair model). The GDD values were calculated from the day of planting for each field separately from the minimum and maximum temperature value obtained from the data from the local weather station and the base temperature of 8ºC. The following formula was used for calculating the GDD:

$$\text{GDD per day} = (\text{min temp} + \text{max temp}/2) - 8^{\circ}\text{C} \tag{3.2}$$

The calculated GDD values for each day after planting were then cumulated until it reached 150, and that day was considered the day of emergence for the particular soybean field (See Table 3.6 for the calculated emergence values).

From the simulation of the Sinclair model, specific dates for the time of first reproductive stage (R1), termination of leaf growth (TLG) and maturation (Mat) were obtained. For the simulation of the Sinclair model, daily values on minimum temperature, maximum temperature, solar radiation and precipitation were obtained for each field site. These are the input variables required for the model to simulate the daily water balance and biomass and finally yield. The model also calculates development rate based on day length and is able to simulate phases such as time of R1 reproductive stage, the time of termination of leaf growth and the time of maturity for the soybean plant (See Table 3.7 for the simulated times for R1, TLG and Mat dates).

Thus, the obtained dates for various phenological stages of soybean growth were plotted against the NDVI time-series curves to see whether the time-series curves actually depicted the soybean growth stages. In all the cases of 11 soybean fields studied, it was consistently observed that the dates of various phenological stages matched well with the "temporal metrics" as defined by Reed et al. (1994) (See Table 3.5) that can be observed in a time-series NDVI curve. It was found that after the date of calculated emergence, a time lag occurred before a genuine increase in NDVI could be detected that could be considered as the onset of greenness. This was observed in most of the cases. The model-predictions for time of termination of leaf growth seemed to be stage that best corresponded with the observation or simulation derived time-series NDVI curves. In most of the cases, the TLG dates were at the plateau of the NDVI curve and in some cases the NDVI values seemed to start to decrease from the point of TLG.

The NDVI time-series from the 400 m resolution VIIRS dataset matched well with the NDVI time-series from 250 m MODIS dataset. Both the MODIS and simulated VIIRS time-series NDVI curves depicted the soybean growth characteristics (See Figure 3.8). At the same time, for soybean planted after wheat, it was observed that the NDVI values were higher before plantings which showed that the crop rotations could also be detected from both MODIS and VIIRS NDVI time-series curve.

Table 3.6 Values of various phenological stages for the soybean fields and the corresponding representative NDVI value from Terra MODIS and LAI values simulated from Sinclair Model.

Field Site	Planting Date (PD)	MODIS NDVI	Simulated LAI	Emergence Date (EMG)	MODIS NDVI	Simulated LAI
Marcos Juarez 501	31-Oct-06	0.4	0	13-Nov-06	0.3	0.33
Marcos Juarez 503	24-Nov-06	0.24	0	4-Dec-06	0.4	0.30
Magiolo 604	3-Nov-06	0.26	0	15-Nov-06	0.32	0.32
Magiolo 602	3-Nov-06	0.29	0	15-Nov-06	0.38	0.32
Magiolo 601	13-Dec-06	0.4	0	23-Dec-06	0.64	0.33
Pergamino 703	3-Nov-06	0.3	0	16-Nov-06	0.31	0.45
Pergamino 705	15-Dec-06	0.42	0	25-Dec-06	0.66	0.38
Pergamino 704	3-Nov-06	0.4	0	16-Nov-06	0.34	0.45
Rosario 801	10-Oct-06	0.46	0	21-Oct-06	0.47	0.31
Rio Segundo 101	10-Oct-06	0.37	0	22-Oct-06	0.39	0.31
Rio Segundo 104	11-Oct-06	0.26	0	23-Oct-06	0.28	0.32

Table 3.7 Values of various phenological stages for the soybean fields and the corresponding representative NDVI value from Terra MODIS and LAI values simulated from Sinclair Model.

Field Site	**Date of Reproductive Stage 1 (R1)**	**NDVI**	**LAI**	**Date of Termination of Leaf Growth (TLG)**	**NDVI**	**LAI**	**Date of Maturity Stage (Mat)**	**NDVI**	**LAI**
Marcos Juarez 501	10-Dec-06	0.69	1.09	2-Feb-07	0.9	7.079	15-Mar-07	0.6	0.09
Marcos Juarez 503	30-Dec-06	0.77	1.28	10-Feb-07	0.93	7.104	25-Mar-07	0.21	0.09
Magiolo 604	14-Dec-06	0.58	1.01	4-Feb-07	0.92	7.126	16-Mar-07	0.6	0.08
Magiolo 602	14-Dec-06	0.56	1.01	4-Feb-07	0.9	7.030	16-Mar-07	0.61	0.09
Magiolo 601	18-Jan-07	0.85	1.08	19-Feb-07	0.88	4.973	6-Apr-07	0.53	0.09
Pergamino 703	11-Dec-06	0.54	1.2	3-Feb-07	0.93	7.144	14-Mar-07	0.5	0.09
Pergamino 705	18-Jan-07	0.86	1.11	16-Feb-07	0.89	4.942	4-Apr-07	0.47	0.09
Pergamino 704	11-Dec-06	0.51	1.24	3-Feb-07	0.91	7.059	13-Mar-07	0.65	0.09
Rosario 801	15-Nov-06	0.4	0.82	14-Jan-07	0.86	7.158	17-Feb-07	0.6	0.08
Rio Segundo 101	24-Nov-06	0.54	1.10	22-Jan-07	0.91	7.007	9-Mar-07	0.6	0.09
Rio Segundo 104	25-Nov-06	0.52	1.11	23-Jan-07	0.88	7.020	10-Mar-07	0.42	0.09

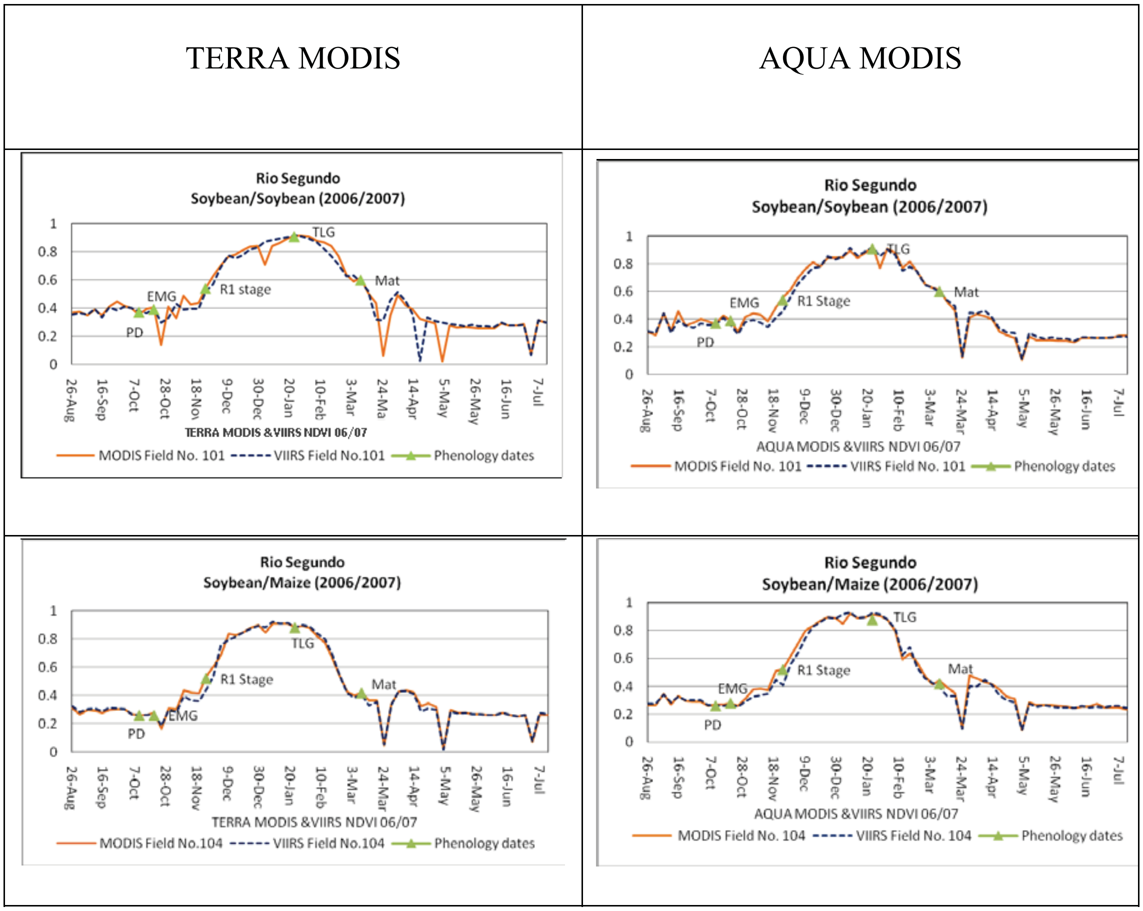
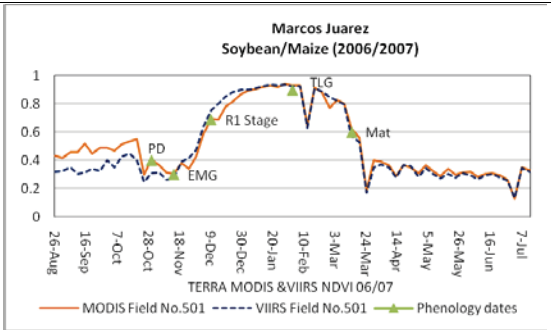
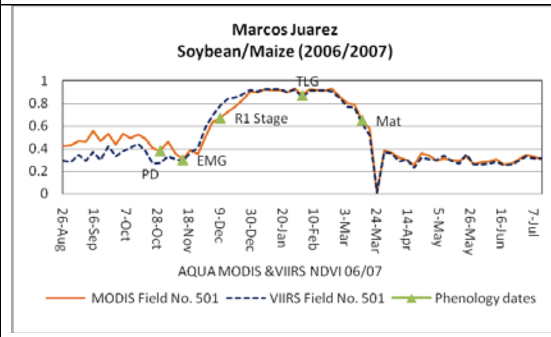


Figure 3.8 Charts showing the NDVI curves for various selected soybean fields in Argentina from TERRA MODIS and VIIRS simulated from TERRA MODIS and specific phenological stages for each field.

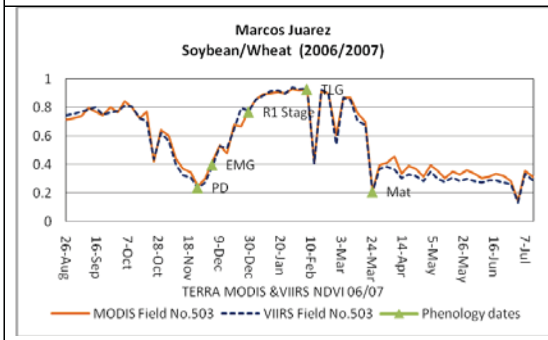
<table>
<tr><th>TERRA MODIS</th><th>AQUA MODIS</th></tr>
<tr><td>



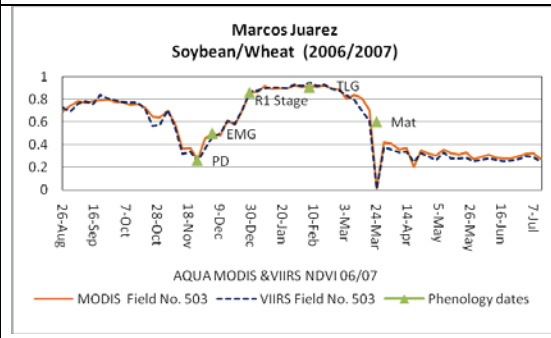
</td><td>



</td></tr>
<tr><td>



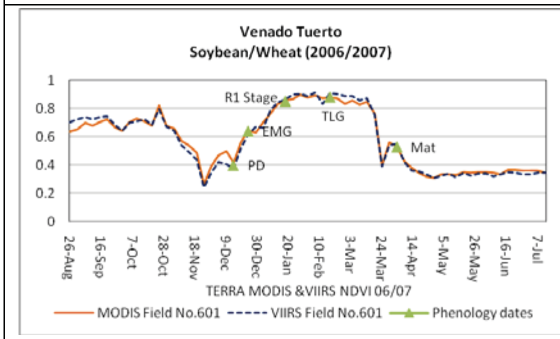
</td><td>

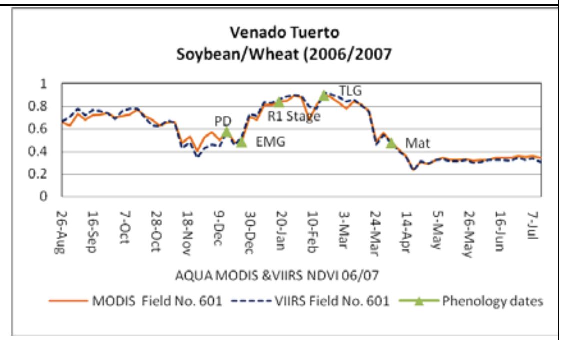


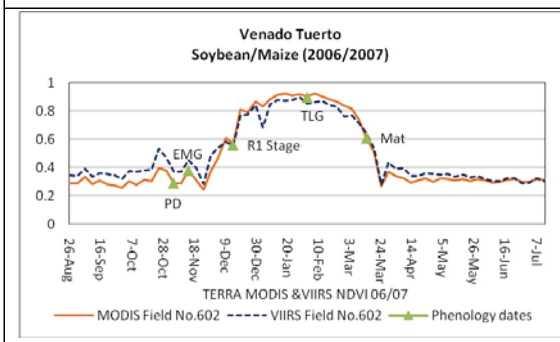
</td></tr>
</table>

Figure 3.8 (Continued).

TERRA MODIS	AQUA MODIS







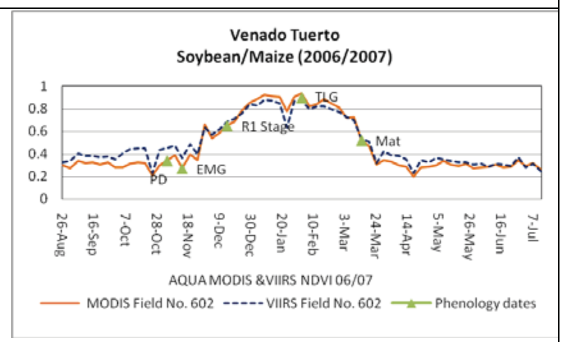


Figure 3.8 (Continued).

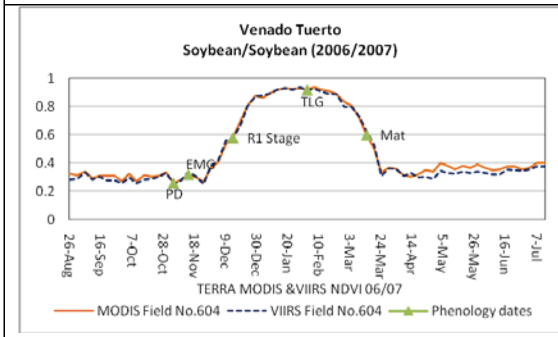
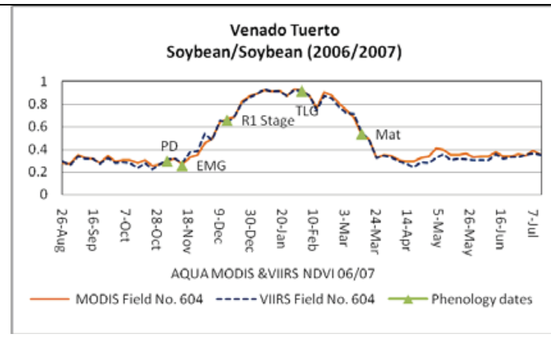
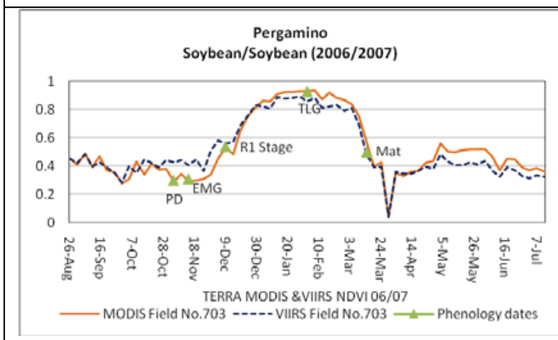
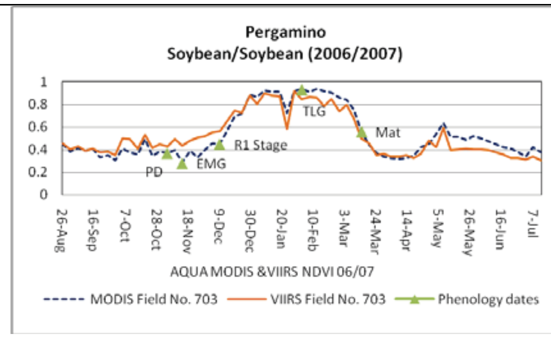
TERRA MODIS	AQUA MODIS
	
	

Figure 3.8 (Continued).

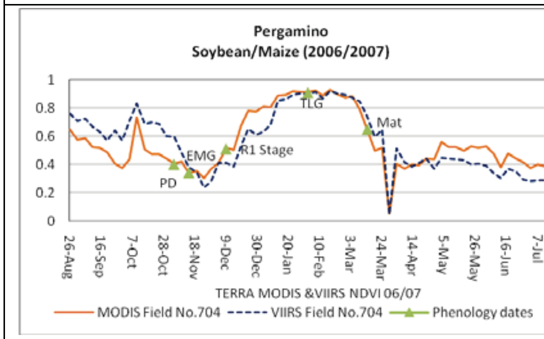
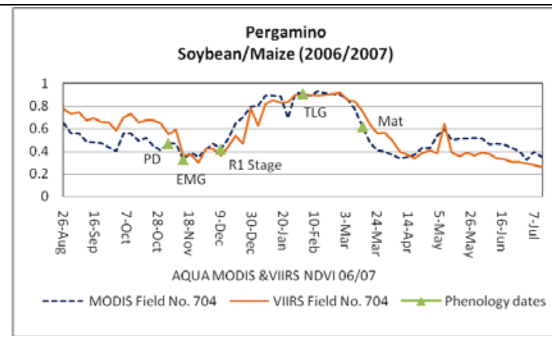
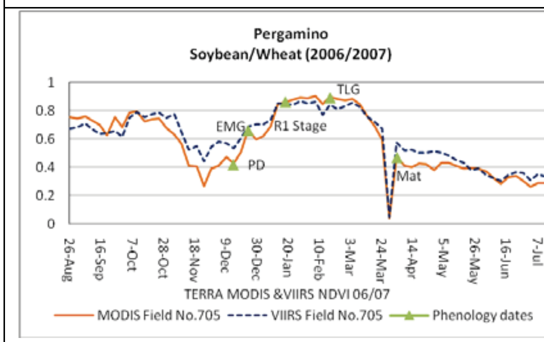
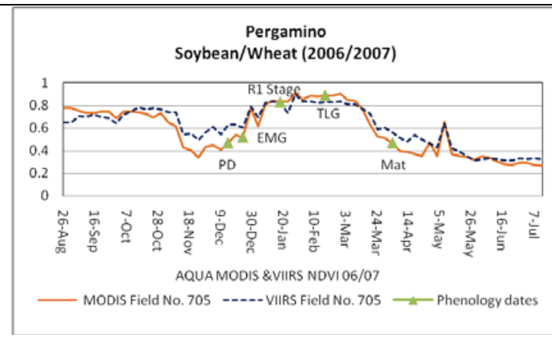
TERRA MODIS	AQUA MODIS
	
	

Figure 3.8 (Continued).

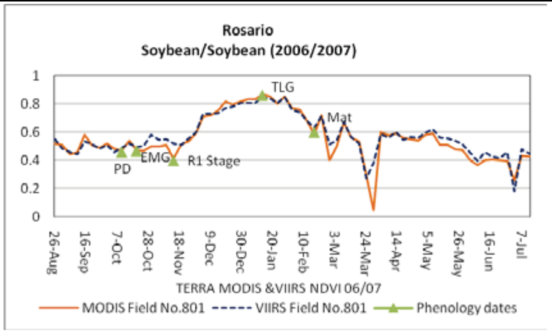
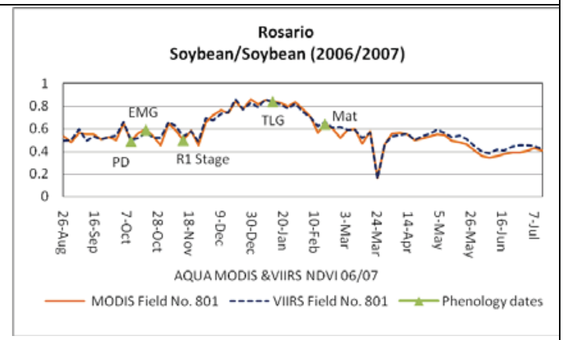
<table>
<tr><th>TERRA MODIS</th><th>AQUA MODIS</th></tr>
<tr><td>
</td>
<td>
</td></tr>
</table>

Figure 3.8 (Continued).

3.5 Results and Discussions

MODIS time-series for soybean fields were tested for their effectiveness in monitoring growth and productivity. The NDVI time-series curves created from MODIS and simulated VIIRS depicted the growth curve of soybean based upon the metrics developed by Reed et al. (1994). The validation of simulated VIIRS from MODIS comparing it with simulated VIIRS from AWiFS, both atmospherically uncorrected, showed good agreement between the simulated VIIRS from two different simulation sources. Since our analysis was performed with atmospherically corrected gridded MOD09, comparisons were done between MOD09 and MOD02 datasets. The agreement was high, with 0.8 or higher image to image correlation. The NDVI for atmospherically corrected imageries have higher values than the NDVI from non-atmospherically corrected imageries.

Specific dates for various phenological stages were simulated from the Sinclair model based upon the planting dates obtained from the farmer. The emergence dates were calculated based on the growing-degree-day concept using the daily min and max temperature values for each fields studied. The other phenological stages were simulated from the Sinclair model. Thus, obtained dates were compared with the NDVI values for the same dates in the time-series curves obtained from MODIS and simulated VIIRS. In all the cases, the NDVI values showed good agreement with the specific phenology dates for both MODIS and simulated VIIRS. For both cases, the NDVI values increased and decreased depending on crop condition or specific growth stage. The correlation between MODIS and VIIRS NDVI were 0.9 and higher in the majority of cases, and no significant differences were found in the NDVI time-series between AQUA MODIS NDVI and TERRA MODIS NDVI. Therefore, representative values of NDVI selected from TERRA MODIS was used to compare with the specific phenology dates and the LAI simulated from the Sinclair model. In this study, for the emergence dates, the NDVI values ranged from 0.28 to 0.66. The NDVI values seem to be slightly higher at this point. These higher NDVI values may be due to the presence of crop residue and weeds and because of crop rotations and no-till agricultural practices adopted in Argentina (Sinclair et al., 2007). The NDVI values of 0.6 during emergence for two of the fields were for soybeans planted after winter wheat. The fresh crop residue of winter wheat may have resulted in the higher NDVI values for these two fields. However, these fields with soybean over wheat also had the lowest simulated LAI values at 4.9 from the Sinclair model. Both of these fields have planting dates in mid-December, therefore the result could be due to the lesser amount of growth time required for the cultivar to mature causing the simulation model to predict lower LAI values. However, even if the LAI values for these fields were low, the NDVI values were found be at 0.8.

For the R1 reproductive stage, the values ranged from 0.4 to 0.8. At the stage of the termination of leaf growth (TLG), the NDVI values ranges from 0.88 to 0.92, this is also the range of the maximum NDVI as well as the maximum LAI for the plants during growth simulation from Sinclair model. At the crop maturity stage, the NDVI values were in the range of 0.2 to 0.6.

The research was unable to find a correlative relationship between the NDVI at various growth stages and the simulated LAI from Sinclair model. However, for all the stages, irrespective of the simulated LAI and the NDVI value at emergence, the NDVI values seem to increase up to the point of TLG and then start to decrease from that point.

The well known fact on the limitations of MODIS and VIIRS imageries for providing highly accurate crop-based data due to their resolution was observed. However, the benefits of large area coverage in a single scene and the ability to monitor vegetation growth at the farm level albeit with less accuracy as shown in this research results provides encouraging pointers to their use for regional level application.

3.6 Conclusions

The results obtained seem to point that, even though the NDVI values from these sensors could not provide the accuracy at canopy level LAI simulations for the field, the overall monitoring capability of both MODIS at 250 m and VIIRS at 400 m can be considered acceptable for analysis at the field level. Furthermore, both MODIS and VIIRS resolutions seem quite sensitive to crop residue on the ground as well as to crop growth. The background reflectance of crop residue seemed to have affected the crop reflectance. However, for soybean planted after winter wheat, both MODIS and simulated VIIRS NDVI time-series were sensitive to the reflectance of winter wheat before soybean planting dates. Thus, both MODIS and VIIRS resolution showed good application potential for crop rotation monitoring as well. Therefore, considering the regional applicability of these sensors, the ability of both MODIS and VIIRS to provide large coverage and high temporal resolution enables deliverance of data and data products well suited for regional or large area agricultural monitoring.

3.7 References

Bethel, G., and B. Doorn (1998), USDA remote sensing technical and systems support for operational worldwide agricultural analysis, *Proceedings of the First International conference on Geospatial Information in Agricultural and Forestry*, Lake Buena Vista, Florida, 1-3 June 1998.

Boken, V. K., and C. F. Shaykewich (2002), Improving an operational wheat yield model using phenological phase based Normalized Difference Vegetation Index, *International Journal of Remote Sensing*, 23: 4155-4168.

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*, 2:1109–1123.

Gallo, K., L. Ji, B. Reed, J. Eidenshink, and J. Dwyer (2005), Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data, *Remote Sensing of the Environment*, 99:221-231.

Kastens, D. L., K. P. Price, E. A. Martinko, and T. L. Kastens (1998), Assessing wheat conditions in Kansas using an eight-year biweekly enhanced AVHRR dataset and crop phenological indices, *Proceedings of the First International conference on Geospatial Information in Agricultural and Forestry*, 1-3 June, Lake Buena Vista, Florida.

Ji L., K. Gallo, J. C. Eidenshink, and J. Dwyer (2008), Agreement evaluation of AVHRR and MODIS 16-day composite NDVI data sets, *International Journal of Remote Sensing*, 29: 4839-4861.

Leeuwen, W. J. D. Van, B. J. Orr, S. E. Marsh, and S. M. Hermann (2006), Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications, *Remote Sensing of the Environment*, 100: 67-81.

Lobell, D. B., G. P. Asker, 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.

O'Hara, C. G. (2008), Arcaml Zonal Processing Script, GeoSystems Research Institute, Mississippi State University, Mississippi State, MS.

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, 19:119–139

Reed, B. C., J. H. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant, and D. O. Ohnlen, (1994), Measuring phenological variability from satellite imagery, *Journal of Vegetation Science*, 5:703-714.

Sinclair, T. R., L. R. Salado-Navarro, G. Salas, and L. C. Purcell (2007), Soybean yields and soil water status in Argentina: Simulation analysis, *Agricultural Systems,* 94:471-477.

Shrestha, B. (2008), Database Manipulation Script, GeoSystems Research Institute, Mississippi State University, Mississippi State, MS.

Thomas, L. F., S. D. Miller, C. F. Schueler, J. D. Hawkins, J. J. Turk, K. Richardson, and J. Kent (2005), NPOESS VIIRS: Dramatically improved operational capability, *Proc. SPIE/ Volume* 5658, DOI: 10.1117/12579019.

Tucker, C. J., J. H. Elgin, and J. E. McMurtrey (1979), Temporal spectral measurements of corn and soybean crops, *Photogrammetric Engineering and Remote Sensing*, 45:643-653.

Turner, J. T. (1998), Operational use of remote sensing and other geospatial information technologies for crop forecasting at USDA Foreign Agricultural Service, *Proceedings of the First International Conference on Geospatial Information in Agricultural and Forestry*, 1-3 June, Lake Buena Vista, Florida.

Wiegand, C. L., and A. J. Richardson (1990), Use of spectra indices to infer leaf area, evapotranspiration and yield: I. Rationale, *Agronomy Journal*, 82: 623-629.

Wiegand, C. L., and A. J. Richardson (1990), Use of spectra indices to infer leaf area, evapotranspiration and yield: II. Results, *Agronomy Journal*, 82: 630-636.

Zanoni, V., B. Davis, R. Ryan, G. Gasser, and S. Blonkski (2002), Remote sensing requirements development: A simulation-based approach, *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Proceedings Volume: XXXIV part 1, 10-15 Nov, Denver, USA.

CHAPTER IV

IMPROVEMENT IN PLANTING DATE ESTIMATION THROUGH THE USE OF NDVI DERIVED FROM SATELLITE IMAGERY

4.1 Abstract

In this chapter, a method of estimating soybean crop planting date is evaluated using daily fused composited Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) temporal curves depicting farm level soybean growth patterns. Planting date is an important initializing variable within a crop model. Accurate planting dates can greatly increase the accuracy of crop yield estimations. A cross-platform sensor data fusion method was used to combine MODIS data from AQUA and TERRA platforms to obtain daily cloud-free composited NDVI values. The analysis showed that the NDVI time-series curves of late-season soybean varieties (double cropped soybeans planted right after wheat harvests) do not provide a clear NDVI increase after emergence, as the no-till method was used predominantly in Argentina and the pre-crop growth causes these non-crop signals to be mixed with actual crop growth signals. Therefore, the estimation of planting date using the NDVI was limited to early season soybeans. With the obtained actual planting dates, it was possible to verify that actual plant emergence [as predicted by 150 cumulative growing degree

days (CGDD)] was not detectable and at least approximately 2 weeks of subsequent growth was required for detection of a "jump" in the greenness value on the phenology curve. Analysis of 10 field trials resulted in an estimated CGDD of 356 prior to observed jump. Comparing this estimation method to actual planting dates, the analysis was able to predict planting date in average of ±3.4 days. Further studies and refinements to the analysis are required to develop quantitative analysis methods for validated estimations of planting date as detection of greenness increase will likely vary among sensor systems, crop varieties, and potentially geographic locale and, as the analysis has shown, even within the crop planting seasons.

4.2 Background and Introduction

Most crop yield models require various initializing values for different parameters and variables in addition to the driving variables which are mostly daily weather values, for crop growth simulation process (Bouman et al., 1992; Mass, 1988). Accurate estimations of these initializing parameters are needed for reliable yield estimation and thus minimizing errors in the simulation results. For regional yield predictions, a lot of these initializing variables are given values derived from approximation based upon previous years' data, or utilizing a default value or through trial and error (Dorigo et al., 2007; Hansen and Jones, 2000). However, this method creates uncertainty and causes error in the model output (Dorigo et al., 2007). Planting date is one such variable that most crop models require for initialization of the model (Sinclair et al., 2007; Fang et al., 2008; Moen et al., 1994). Planting date is also an important variable since other variable values are also initialized for the day of planting such as initializing values of soil water content.

Soybean development is sensitive to temperature as well as day length; therefore planting of the seeds at the right time is important for obtaining optimum yield (Sinclair et al., 2007; Zhang et al., 2001). Hence, planting date is related to the maturity group and has an effect on the development rate and phenological growth stages of the plant (Loomis and Connor, 1992).

In the field level, the actual dates of planting depend mostly on the farmer's decision that is based upon the weather and other local conditions. And as an initializing variable for crop yield modeling, dates for planting that has been determined at the local level is a difficult estimate.

The effects of planting date on various fields in Argentina were observed during simulation analysis of soybean from Sinclair soybean yield model. The simulation results for predicted phenological stages of soybean showed [see Fig. 4.1] that earlier plantings resulted in a longer duration between reproductive stage 1 and stage of termination of leaf growth. The duration between reproductive stage 1 and stage of termination of leaf growth is lessened in late plantings. If the planting date occurred in early to late October, the time period between the R1 stage and the TLG stage was higher compared to planting dates occurring in late November and early December. The duration between the R1 and TLG stages were significantly less in fields where late planting in December occurred, compared to the duration between these stages for early October plantings.

This result demonstrated the effect of planting date on the crop development phases. Similarly, for the Sinclair soybean yield model LAI growth is highly sensitive to sowing and an earlier sowing date resulted in the peak LAI resulting earlier in the season (Sinclair et al., 2007). Various crop models have been found to be sensitive to planting dates. A sensitivity analysis on the various initializing variables for corn yield estimation for CERES maize model showed planting date as the second most sensitive variable after nitrogen application (Fang et al., 2008). These results show that accurate planting date is important for crop development, growth and yield. A study by Moen et al. (1993) showed that planting dates affected regional yield predictions and using a single planting date resulted in less accuracy of maize simulation model than utilizing seven planting dates within the area.

For regional level analysis, planting date estimation becomes very important because it is almost impossible to correctly obtain the exact planting dates for a large area of interest. Operationally, for regional analysis at the USDA FAS/PECAD, the planting date is calculated based on crop reports which are subjective and can deviate up to 30 days (Doorn, B., personal communication, 2005).

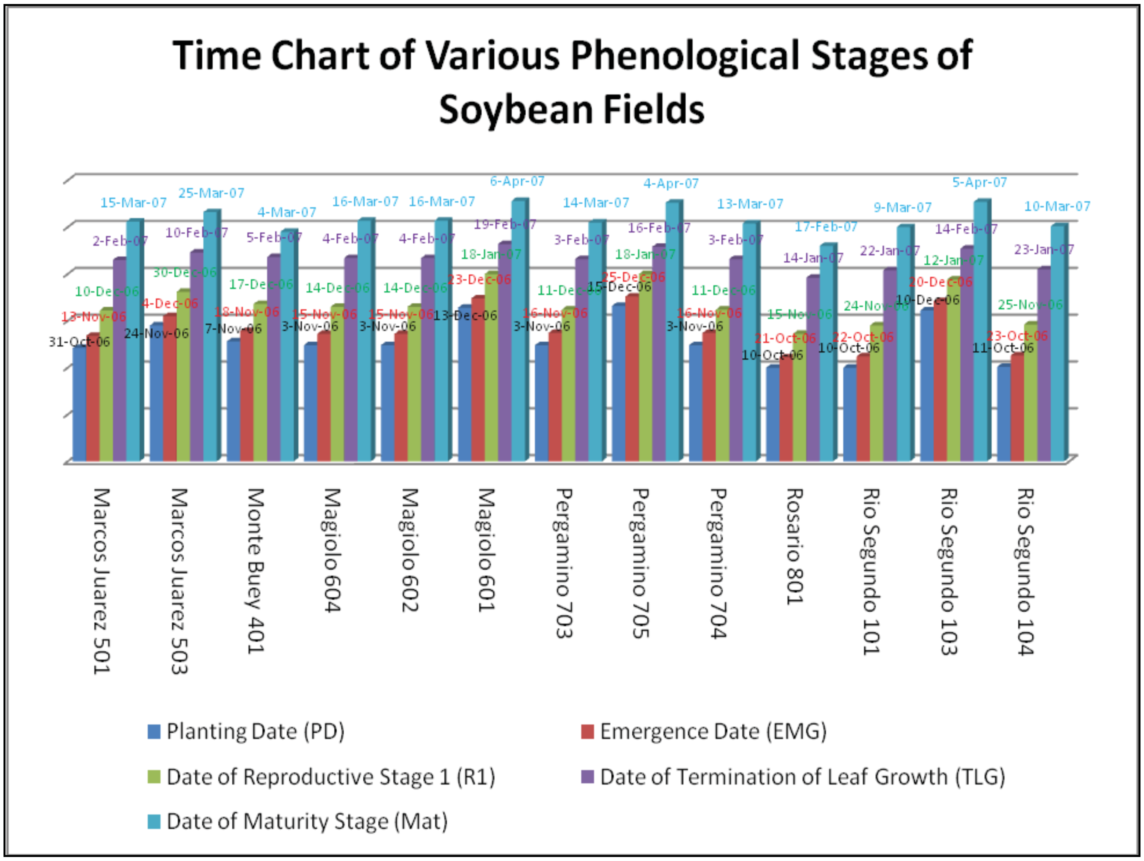


Figure 4.1 Time chart of various phenological stages for the soybean fields as calculated by the Sinclair crop model.

These deviations result in increased inaccuracies in crop yield estimations that FAS/PECAD provides to the end users, which are government agencies and national policy makers that are involved in making decisions on local, national and global food trade and commodity pricings (Doorn, B., personal communication, 2005; NASA, 2003). These subjective estimations are possible only in situations where crop reports are available and an infrastructure to collect reports from farmers already exists.

Therefore in this chapter, an alternate method utilizing remote sensing for planting date estimation is investigated. Vegetation indices from remote sensing imageries can provide values that are representative to the vegetation cover on the ground

(Hatfield et al., 2008). NDVI is the most commonly used index for monitoring crop growth and also an easily available product through various data archiving and distributing agencies. For the objective of estimating planting date, the temporal resolution of imageries is critical, since the required variable is a time value. Time-series vegetation index values, which can be obtained from low resolution imageries with high temporal visits such as AVHRR and MODIS, can be utilized to obtain vegetation index temporal curve with information on phenological characteristics (Reed et al., 1994). The temporal curves can provide the ability to detect the onset of greenness that can be related to crop emergence. The detection of emergence can be used as an indicator to estimate the planting date. The main problem in utilizing high temporal resolution data such as AVHRR and MODIS is the presence of cloud cover. A compositing process is used to select highest NDVI values for a given time range, since pixels on days with clouds have low NDVI values. The time period for AVHRR composites ranges from bi-weekly to monthly datasets, which is an insufficient level of temporal coverage for detecting the subtle changes in greenness. A fusion-based vegetation index compositing as described in Shrestha et al. (2008) and O'Hara et al. (2008) utilizes both AQUA and TERRA MODIS NDVI values to obtain near-daily NDVI values. The ability to detect the NDVI changes daily can provide an efficient method for detecting the onset of greenness and for estimating a more accurate planting date. Therefore, in this research the methodology for processing AQUA and TERRA MODIS NDVI values to obtain near daily cloud free NDVI values is discussed and the relative success with which these data were used for estimating planting date and the potential advantage of such products for crop growth monitoring and yield modeling is demonstrated.

4.3 Methodology

4.3.1 Data Collection and Preprocessing

The study is performed with imageries from MODIS sensors which are onboard both TERRA and AQUA Earth Observation System (EOS) satellites. MODIS has 36 spectral channels. Among the spectral channels within MODIS imagery, near infrared and red channels have a spatial resolution of 250 m (Huete et al., 2002, Justice et al., 1998). The next five channels have resolution of 500 m and the remaining channels have resolution of 1000 m. MODIS AQUA and TERRA satellites are able to provide global coverage of the earth two times a day. The datasets were collected for two different study areas. A preliminary test was performed for the Mississippi-Arkansas area for MODIS image compositing. Then, for the planting date estimation study, the *Pampas* in Argentina was chosen. A number of soybean fields in the *Pampas* were selected for field verification and planting dates.

For the preliminary testing, MODIS AQUA level 2G datasets with 250 m resolution were downloaded from Land Processes – Distributed Active Archive Center (LP-DAAC) website for a month for the area encompassing Mississippi and Arkansas. MODIS AQUA reflectance datasets (MYD09GQK), quality datasets (MYD09GGK) and angle datasets (MYD09GGAD) were collected. These, MODIS level 2G, images downloaded from LP-DAAC were already processed for radiometric, atmospheric and geometric corrections. The projection for the images was converted from World Sinusoidal to Universal Transverse Mercator (UTM) projection, Zone 15, WGS 84 Datum. The projected images were then subsetted to the boundaries of Mississippi and

Arkansas. A tool called HEG (HDF-EOS to GIS Format Conversion Tool) was used for the preprocessing of these imageries.

For the estimation of planting dates, the MODIS reflectance datasets (MOD/MYD09GQK), quality datasets (MOD/MYD09GGK) and angle datasets (MOD/MYD09GGAD) were collected for both AQUA and TERRA MODIS through NASA-Stennis-based Institute for Technology Development (ITD). The study time period for our study was 2006/2007 soybean growing period in Argentina. The images for this time period were processed using the ART toolbox that had a functionality that accesses the functions of MODIS re-projection tool (MRT) which can read files in HDF format and re-project. The original HDF format MODIS data were in Sinusoidal projection. The MODIS red and infrared bands were extracted and re-projected to UTM projection, Zone 20 S, WGS 84 and then subsetted utilizing the MRT application through ART encompassing the soybean field sites selected for our study. The HDF files were then converted to Geographic Tagged Image File Format (GeoTIFF).

4.3.1.1 Field Boundary Verification

Through a field expert and research collaborator in Argentina (Dr. Luis Salado Navarro), GPS points for corners of soybean fields were obtained. These points were used to delineate and demarcate field boundaries. Google Earth's high resolution imageries were used to place these GPS points in the correct field corners. At the same time, Advanced Wide Field Sensor (AWiFS) imagery for the location was also used for properly delineating the soybean fields. These points were then digitized as an ArcGIS polygon shape file. The digitized fields were then verified using land parcel maps obtained from farmers for each field and as well as from the first hand knowledge of the field expert (Chapter 2, Figure 2.3).

4.3.1.2 Field Based Datasets Collected

Planting dates for the selected soybean fields were obtained from the farmers through the help of local crop scientist Dr. Luis Salado Navarro. Similarly, meteorological datasets required for this study (daily minimum and maximum temperatures) were also obtained.

4.3.2 Analysis and Results

4.3.2.1 NDVI Calculation and Fusion Based Compositing

The re-projected and subsetted MODIS images were then converted to NDVI images. The NDVI was calculated using the following equation:

$$NDVI = NIR-R/NIR+R \tag{4.1}$$

where NDVI = Normalized Difference Vegetation Index, NIR = Near Infrared Band, R= Red Band.

The NDVI value ranges from -1 to +1; the value increases from -1 to +1 with the increase in vegetation. The clouds are in the lower end of the NDVI value range. Since the images had cloud coverage, although daily data were obtained, daily NDVI values could not be obtained. Therefore, vegetation index compositing was performed to remove the cloud cover. However, traditional maximum value compositing process usually can provide only data for 8-10 periods. Therefore, a fusion based vegetation index compositing as described by Shrestha et al. (2008) and O'Hara et al. (2008) that utilizes both AQUA and TERRA MODIS NDVI was used. This method is able to provide near daily NDVI values.

In the compositing process, the reflectance datasets (MOD09GQK) were used in conjunction to quality datasets (MOD09GGK) as well as angle dataset (MOD09GGAD). The surface reflectance quality cube has been derived from the quality image datasets obtained from LP-DAAC. The information content from the original quality dataset was re-coded to represent various cover types which are cloud, land, water, snow and no data values. In the algorithm, land pixels were given the first preference while choosing the NDVI values during the compositing process.

In the maximum value compositing process, off nadir pixels are more likely to be selected as these pixels tend to have higher NDVI values (Zhu and Yang, 2003). Therefore removing the pixels with large angle values is required. For testing the optimum view angle for compositing process, a test was performed for the imageries from Mississippi and Arkansas area. At first, the NDVI values with view zenith angle lesser than 55º were selected in the compositing process. However, the resultant images showed striping effects (Mali et al., 2005). Therefore, a comparative analysis was performed to find the best view zenith angle for the compositing process. On visual comparison of composited images using zenith angle of 55º degrees to 42º, it was found that at around the view zenith angle of 48º, the striping effect was removed. The best visual was obtained at 42º. However, in order to obtain an optimum zenith angle so that a majority of pixels will be considered while compositing, correlation coefficient between composited NDVI imageries for angles of 55º to 42º and the composited imagery of 42º were calculated (Figure 4.2). At the zenith angle of 48º, the correlation coefficient crossed 0.99. Therefore, the zenith angle of 48º was considered as a cutoff point and used for angle constraints in the compositing process.

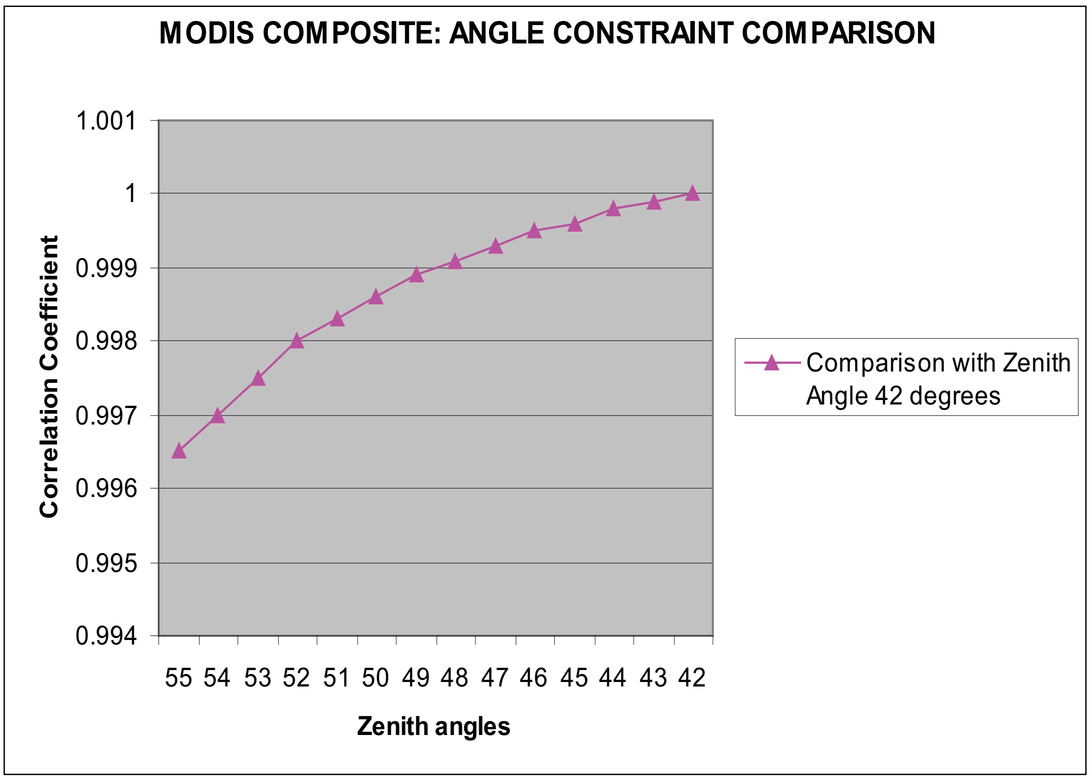


Figure 4.2 Comparison between various zenith angle constraints for compositing process. The zenith angle of 48º provided a good cutoff point.

4.3.2.2 Difference between AQUA and TERRA MODIS mean NDVI values

The cross-platform fusion algorithm utilizes both AQUA and TERRA MODIS data. Since AQUA and TERRA satellites orbit a location on Earth at different times in a day, a preliminary test for our study site was performed to test if there were any significant differences between the AQUA and TERRA MODIS NDVI values. For the testing, two sampled t-test at 95% confidence interval was performed for the AQUA and TERRA zonal NDVI mean values for nine locations for the time period of total 32 days for our selected area of interest in Mississippi and Arkansas. At 95% confidence interval, the p value for all the zones were greater than 0.05, thus the null hypothesis that for all the zones there is no significant difference between the means of AQUA and TERRA NDVI values was accepted.

4.3.2.3 Cross Platform Fusion and Composite Creation

The details of the algorithm are described in O'Hara et al. (2008) and Shrestha et al. (2008). The cross platform fusion and compositing was performed for the selected area of interest in the Argentine *Pampas* that consisted of soybean fields selected for our field verification purposes. The fusion process utilizes rule-based approach for selecting pixels and creating a subset of observations that meet view angle and quality criteria. The rule-based algorithm is based on the fact that pixels with lower zenith angles contains less noise, and also utilizes the associated metadata that can differentiate between land, water, cloud and snow observations to mask out the pixels that doesn't represent land observations. Then, a temporal filtering is used to find the pixels that fulfill the required criteria. The criteria that that were enforced were that NDVI value is the maximum and the zenith angle for the pixel must be less than 48^{o} and also that the underlying pixel is a classified land pixel based on the associated image metadata.

A six-day temporal window was considered for compositing using filtering and fusion good observations, i.e., the application first checked the current day data; and if it did not fulfill the constraints, checked up to five days of NDVI data for the pixel under consideration for both AQUA and TERRA MODIS NDVI. Fusion Quality Confidence Codes (FQCC) values were calculated for each pixel of the fused image to quantify the quality and confidence of the pixel under consideration. Higher confidence codes were given for NDVI chosen on the day of interest that matches the view zenith and quality code criteria and less confidence is given to NDVI values chosen away from the day of interest. If no pixels match the zenith angle and quality control, then the pixel with highest NDVI is chosen. The chosen pixel could be from either AQUA or TERRA MODIS. Using these criteria, the algorithm was able to create cloud free NDVI images for each day that could be within six days of interest.

4.3.2.4 Zonal Processing

The zonal processing was done in ArcGIS environment utilizing Arc-aml functions with batch processing capability that was created by O'Hara (2008). The daily fused NDVI composite images for 2006/2007 soybean growing season were utilized as a time-series data cube and using the soybean fields as zones, the NDVI values were extracted for each day. The extracted time-series NDVI values for each field were then used to detect the onset of greenness (See Figure 4.3).

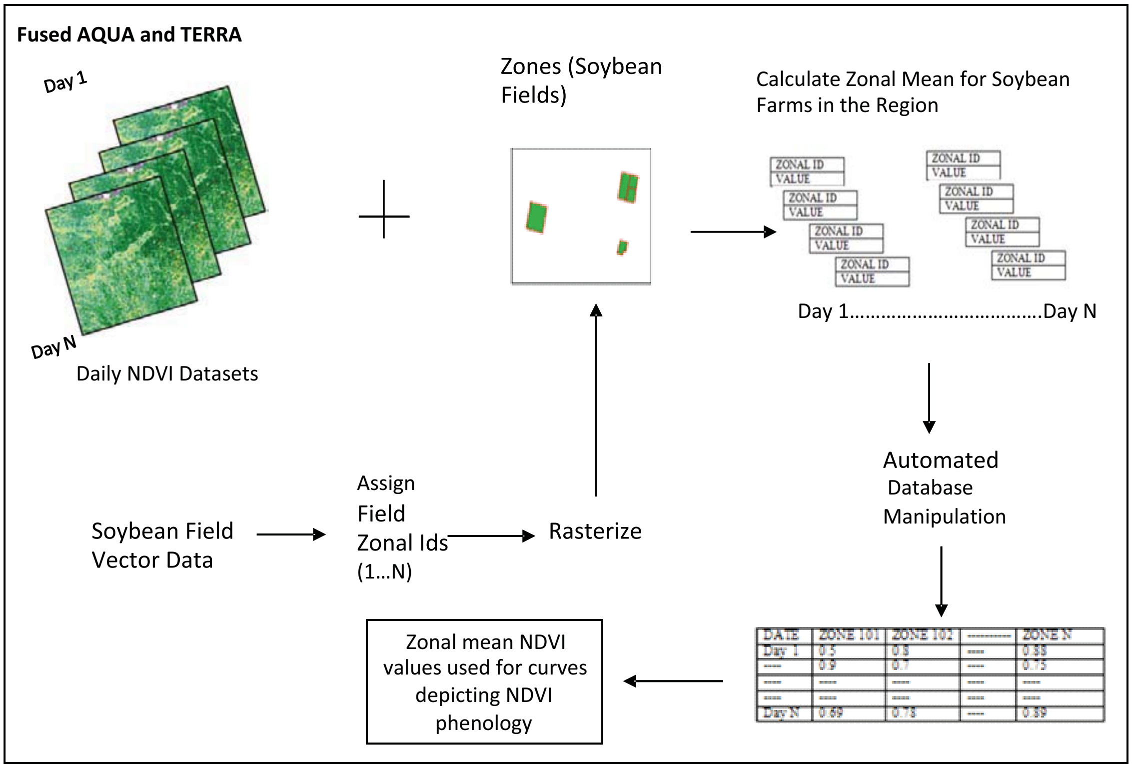


Figure 4.3 Flowchart of the geo-processing methodology used for the analysis.

4.3.2.5 Planting Date Estimation Process

Once the daily NDVI values are calculated, the planting date can be estimated from an easy method utilizing daily minimum and maximum temperature values for each field studied. Utilizing these daily temperature values obtained from the local weather stations, growing degree day (GDD) values were calculated. The GDD values were cumulated to obtain cumulative GDD values. In the time-series NDVI values, the start of the continuous increase of NDVI was determined by observation of the steady rise of NDVI values greater than 0.3 in the beginning of the soybean growth per field. This point was considered as the tentative emergence date. Utilizing this emergence date the planting date was estimated by back propagating the cumulative growing degree days.

The emergence date was also calculated based on the planting date observed from the farmers. This emergence date was considered the actual emergence date and was used to compare with the estimated emergence date detected from the daily NDVI values (Figure 4.4).

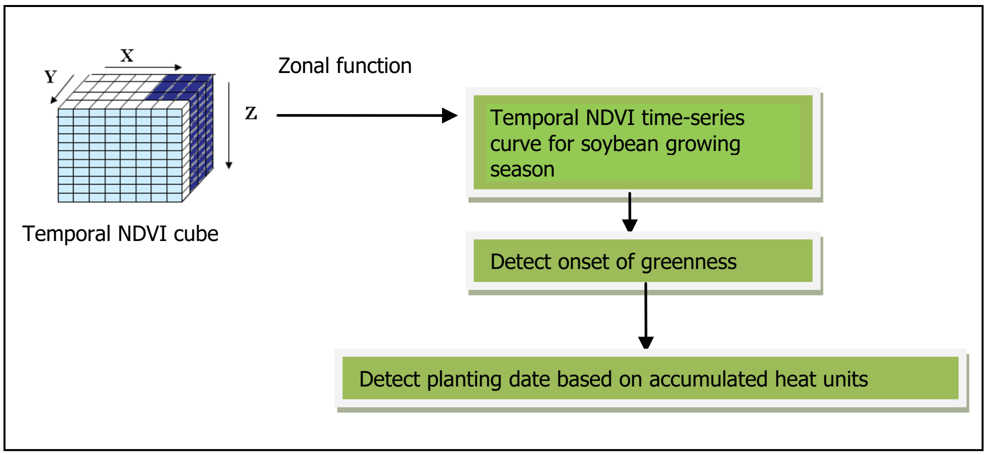


Figure 4.4 Planting date estimation using Vegetation Index and Temporal map algebra (Mali et al., 2006).

4.4 Results and Discussion

The planting dates were estimated based on the relationship between the growing degree days (GDD) and the soybean phenology. The GDD values were calculated for each field using minimum and maximum temperature values for each day and the base temperature of 8^{o}C. Figures 4.5 and 4.6 show the NDVI phenology curve for the soybean fields. In the field of soybean planted over soybean (Fig. 4.5), the maturity of the soybean corresponds to the GDD. However, for the field with soybean planted over winter wheat (Fig. 4.6), where the soybean planting is late for a late maturing variety, the GDD does not correspond with the maturation as observed from the phenology curve although the increase in initial phases of soybean growth corresponds with the increasing GDD. Therefore, even though the planting date may be related with the GDD for the soybean growing season, the other phenology stages may not produce the required results based solely on the relationship between GDD and soybean phenology. This observation is in agreement with a similar conclusion by Daubenmire (1947) that the "heat unit" requirement of a given process is constant only for that range within which a direct proportionality exists between the growth rate and temperature (Wang, 1960, pp. 787).

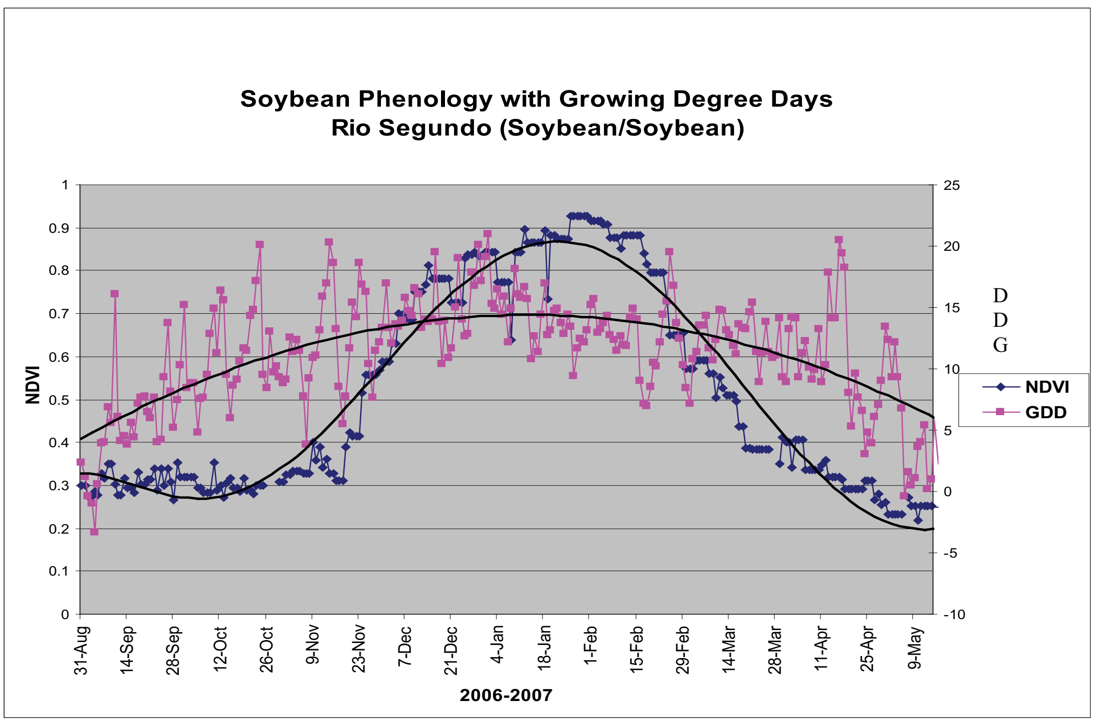


Figure 4.5 Graph showing soybean phenology and growing degree days for soybean planted over soybean field in Rio Segundo.

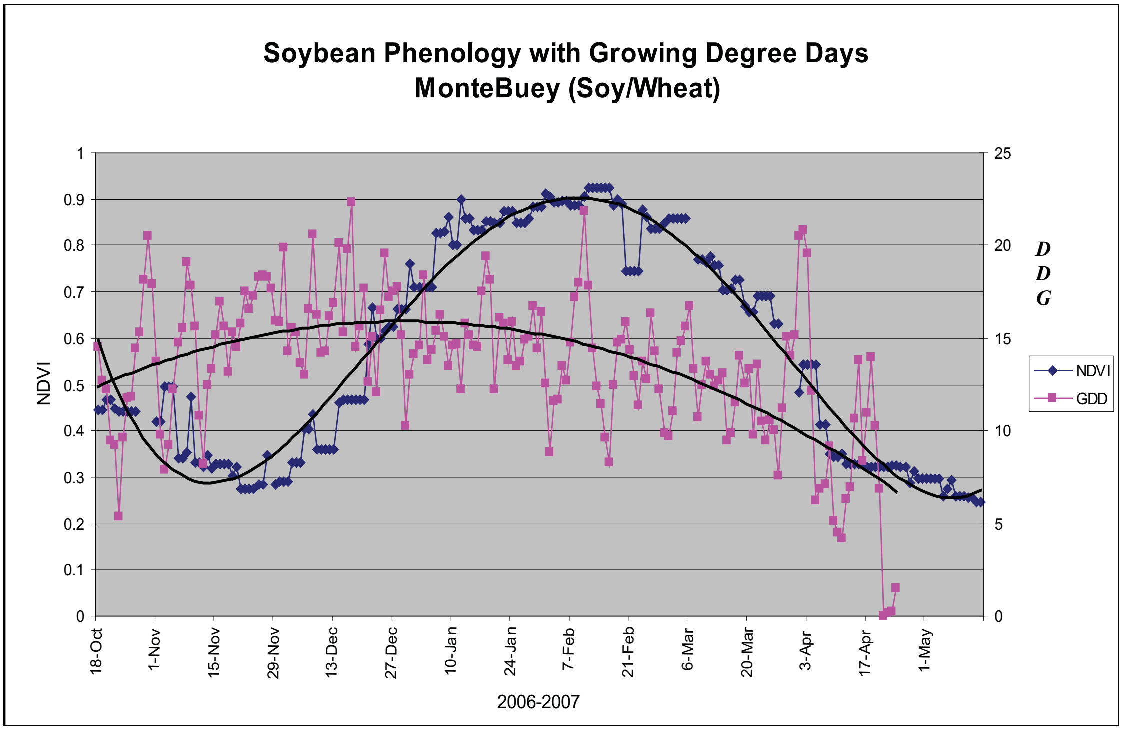


Figure 4.6 Graph showing soybean phenology and growing degree days for soybean planted over wheat field in MonteBuey.

The analysis revealed a clear difference in the NDVI values for early-season soybeans and late-season soybeans at the beginning of the growing season. The analysis found that, in the case of early-season soybeans which are planted on early October to early November, the start of NDVI increase could be easily detected. The time-series NDVI for early plantings showed a clear demarcation of pre-planting NDVI and post-planting NDVI values, and a value of NDVI 0.3 and higher could be considered as the point of possible onset of greenness. In the case of fields where plantings took place in late November to late December, the clear demarcation between the pre-planting NDVI and post-planting NDVI was not found. It could be inferred that the unclear demarcation of the point of rise of NDVI values (for late planting varieties of soybeans) for the early growing phase of the plant is due to the presence of mulch on the ground. A majority of farmers in Argentina practice no-till agriculture along with intensive crop rotations (Salado-Navarro and Sinclair, 2009). The late soybean plantings are done in areas where a crop has recently been harvested before the current soybean plantings. Due to no-till practice the mulch remaining on the ground resulted in a higher NDVI values before and after plantings which resulted in unclear demarcation of possible point of emergence date. Therefore, the planting date estimations were limited for the soybean fields with early plantings.

The results (Table 4.1) showed that start of continuous linear increase in NDVI values did not occur from the actual plant emergence date, but in an average of two weeks later and in some cases it was up to 3 weeks. From the calculated emergence date it was observed that, based on the concept of 150 CGDD days, the climatic factors in Argentina (particularly temperature) caused the emergence to occur in an average of 12 days for the early planting soybean varieties. Since the observed NDVI increase occurs approximately 2 weeks later and in some case 3 weeks, it can be inferred that the MODIS NDVI detects the start of the vegetative growth around the various vegetative growth stages probably from V2 to V4. The CGDD values for the date when the NDVI jump was detected for the various fields were found to range from 289 to 406, the average of which is 356 CGGD. This average value of 356 CGGD was used for predicting the planting dates for back propagating the cumulative CGDD after the NDVI increase detection.

Table 4.1 Estimated planting dates of early planted (October to early November) soybeans.

Field Locations	Crop Rotation	Actual Planting date	Emergence Date (150 CGDD)	NDVI Jump Detection	Corresponding CGDD	Days of NDVI jump after emergence	NDVI jump after planting	Predicted Planting date based on NDVI jump and CGDD	Difference between actual and predicted planting dates
Rio Segundo	Soybean/ Soybean	10-Oct-06	22-Oct-06	13-Nov-06	406	22	34	13-Oct-06	3
Rio Segundo	Soybean/ Maize	11-Oct-06	23-Oct-06	17-Nov-06	451	25	37	19-Oct-06	8
Maggiolo	Soybean/ Wheat	3-Nov-06	15-Nov-06	27-Nov-06	289	12	24	27-Oct-06	-7
Maggiolo	Soybean/ Maize	3-Nov-06	15-Nov-06	29-Nov-06	312	14	26	29-Oct-06	-5
Marcos Juarez	Soybean/ Maize	31-Oct-06	13-Nov-06	25-Nov-06	338	12	25	29-Oct-06	-2
Rosario	Soybean/ Soybean	10-Oct-06	21-Oct-06	7-Nov-06	377	17	28	11-Oct-06	1
Monte Buey	Soybean/ Wheat	7-Nov-06	18-Nov-06	3-Dec-06	366	15	26	08-Nov-06	1
Pergamino	Soybean/ Maize	3-Nov-06	16-Nov-06	2-Dec-06	351	16	29	02-Nov-06	-1
Pergamino	Soybean/ Soybean	3-Nov-06	16-Nov-06	29-Nov-06	306	13	26	29-Oct-06	-5
Monte Buey	Soybean/ Wheat	7-Nov-06	18-Nov	3-Dec-06	366	15	26	08-Nov-06	1
								Average	+/- 3.4 days

From the results obtained it could be concluded that the NDVI "increase detection day" is only later in the vegetative stage rather than sooner. The estimation of the planting date based on the methodology described previously resulted in estimation of planting dates as close as one day for some fields. Overall, the method was able to predict planting date within an average of ±3.4 days.

4.5 Conclusions

A new method is described for providing an alternate method of estimating crop planting date utilizing time-series vegetation index curves for crop growth modeling. The day of planting is a sensitive factor that can affect the growth rate of the plant and phenological dates, but is highly subjective to field conditions and a difficult model input parameter to obtain. Historically, it is estimated by using crop calendars and previous crop reports, but this often is inaccurate. Our analysis showed that the date of termination of leaf growth increased or decreased from the R1 reproductive stage based upon the planting date. The changing environmental patterns cause the field level decisions by the farmers to be dynamic in nature, which makes the estimation of such field level variables more difficult.

The utilization of daily fused NDVI from MODIS AQUA and TERRA images provided the required temporal resolution for the estimation of planting date. A small test was performed to determine significant differences in the mean value of selected zones for a test site in Mississippi and Arkansas. The t-test results showed no significant differences between the mean NDVI values. The NDVI values are sensitive to the sensor geometry. Since each MODIS scene provides sensor geometry information, the test described above revealed that for obtaining the best possible NDVI values during the pixel selection criteria, the values is best chosen from the pixels with the zenith angle of 48^{o} which is optimum during the compositing process. For the early planting varieties, ten fields were tested using the method, out of which in four of the fields the estimated planting dates were within a day of the actual planting date. On average, the estimated planting dates were within ± 3.4 days. The results are encouraging, although for conditions where the previous crop mulch exists, the detection of NDVI increase after emergence becomes difficult. The method used is still qualitative and dependent upon some pre-existing knowledge of field conditions. A mathematical model-based analysis that can extract a change in the NDVI trajectory to demark increases may provide an efficient alternative quantitative process. The daily NDVI values provided from the cross-platform fusion compositing method was important, as it provided the daily temporal resolution that tremendously helped in selecting NDVI values for any date as the analysis required. The lack of daily values would have changed the methodology through which the detection process would be performed and obviously decrease the accuracy of the method as well. Overall, the method described has provided encouraging results although further refinements and more research are required.

4.6 References

Bouman, B. A. M., H. van Keulen, H. H. van Laar, and R. Rabinge (1996), The 'School of de Wit' crop simulation models: A pedigree and historical overview, *Agricultural Systems*, 52:171-198.

Delecolle, R. (1992), Remote sensing and crop production models: Present trends, *ISPRS Journal of Photogrammetry and Remote Sensing*, 47:145-161.

Dorigo, W. A., R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman (2007), A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling, *International Journal of Applied Earth Observation and Geoinformation*, 9:165-193.

Fang, H., S. Liang, G. Hoogenboom, J. Teasdale, and M. Cavigelli (2008), Corn yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model, *International Journal of Remote Sensing*, 29: 3011-3032.

Fischer, A. (1997), Coupling satellite data with vegetation functional models; review of different approaches and perspectives suggested by the assimilation, *Remote Sensing Reviews*, 15: 283-303.

Hansen, J. W., and J. W. Jones (2000), Scaling-up crop models for climate variability applications, *Agricultural Systems*, 65:43-72.

Hatfield, J. W., A. A. Gitelson, J. S. Schepers, and C. L. Walthall (2008), Application of spectral remote sensing for agronomic decisions, *Agronomy Journal*, 100:117-131.

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.

Justice, C. O., E. Vermote, J. R. G. Townshend, R. Defries, D. P. Roy, D. K. Hall, V. V. Salomonson, J. Privette, G. Riggs, A. Strahler, W. Lucht, R. Myneni, Y. Knjazihhin, S. W. Running, R. R. Nemani, Z. Wan, A. R. Huete, W. van Leeuwen, , R. E. Wolfe, L. Giglio, J. P. Muller, P. Lewis, and M. J. Barnsley (1998), The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research, *IEEE Transactions on Geosciences and Remote Sensing*, 36:1228–1249.

Loomis, R. S., and D. J. Connor (1992), Crop ecology: productivity and management in agricultural systems, Cambridge University Press, 538p.

Maas, S. (1998), Using satellite data to improve model estimates of crop yield, *Agronomy Journal*, 80:662-665.

Mali, P., C. G. O'Hara, and V. Anantharaj, (2006), Consideration and comparison of different remote sensing inputs for regional crop yield prediction model, *In Proceedings of ASPRS Annual Conference 2006*, Reno, NV.

Mali, P., C. G. O'Hara, B. Shrestha, and V. Vijayraj (2005), Use and analysis of temporal map algebra for vegetation index compositing, *Analysis of Multi-Temporal Remote Sensing Images, 2005 International Workshop,* 16-18 May, Biloxi, MS, pp. 90-94.

Moen, T. N., H. M. Kaiser, S. J. Riha, (1994), Regional yield estimation using a crop simulation model: Concepts, methods and validation, *Agricultural Systems*, 46: 79:92.

Moulin, S., A. Biondeau, and R. Delecolle (1998), Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing*, 19:1021–1036.

NASA (National Aeronautics and Space Administration) (2003), Decision Support Tools Evaluation Report for FAS/PECAD, Version 1.0.

O'Hara, C. G., B. Shrestha, and P. Mali, (2010), Fusion of cross-platform satellite data and metadata for high-temporal resolution cloud-free compositing, *International Journal of Remote Sensing* (In revision).

O'Hara, Charles G. (2008), Arcaml zonal processing script, GeoSystems Research Institute, Mississippi State University, Mississippi State, MS.

Reed, B. C., J. H. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant, and D. O. Ohnlen, (1994), Measuring phenological variability from satellite imagery, *Journal of Vegetation Science*, 5:703-714.

Salado-Navarro, L. R. and T. R. Sinclair (2009), Crop rotations in Argentina: System analysis of water balance and yield using crop models, *Agricultural Systems*, 102: 11-16.

Sinclair, T. R. (1986), Water and nitrogen limitations in soybean grain production, *Fields Crops Research*, 15: 125-141.

Sinclair, T. R., L. R. Salado-Navarro, and J. M. Bennett (2007), Assessment of soybean model predictions in Argentina. (Unpublished Report).

Sinclair, T. R., L. R. Salado-Navarro, G. Salas, and L. C. Purcell (2007), Soybean yields and soil water status in Argentina: Simulation analysis, *Agricultural Systems*, 94: 471–477.

Wang, J. Y. (1960), A critique of the heat unit approach to plant response studies, *Ecology*, 41: 785-790.

Zhang, L., R. Wang, and J. D. Hesketh (2001), Effects of photoperiod on growth and development of soybean floral bud in different maturity, *Agronomy Journal*, 93:944-948.

Zhu, Z. and L. Yang (1996), Characteristics of the 1 km AVHRR data set for North America, *International Journal of Remote Sensing*, 17: 1915-1924.

CHAPTER V

FINAL SUMMARY AND RECOMMENDATIONS

The objectives of this study were defined for grid-based meteorological datasets and utilizing remote sensing for providing an efficient and improved regional soybean yield modeling and growth monitoring. For enhanced regionalization of crop modeling the research tested various data sources to provide inputs, especially those that required ground-based data collection, with a broader objective of enabling the Sinclair model for regional level use. In chapter one of this dissertation, methods were employed to replace ground-based meteorological data sources to provide 1) forcings from daily SALDAS meteorological datasets and 2) initializing conditions via SALDAS soil moisture values used to initialize the water content value for the water balance module in the Sinclair model. In chapter two of the dissertation, MODIS data and simulated VIIRS were tested for plant growth monitoring through geoprocessing and analysis that compared extracted growth curves to crop model growth stage predictions. In chapter three, remote sensing data streams from AQUA and TERRA MODIS were employed along with grid-based meteorological datasets to develop a semi-empirical method for detecting and refining planting date, a critical variable in crop modeling. The outcome of the research suggests that the combination of remote sensing data and grid-based meteorological datasets provide data streams of high utility for improved regional soybean crop modeling and growth monitoring.

In the case of utilization of SALDAS type datasets, the availability of various crop yield modeling methodologies and models with different level of input requirements creates uncertainty in the utilization of such datasets for crop yield models as a major data source. Therefore, the applicability of SALDAS type grid-based meteorological datasets for crop modeling will depend upon the utilization of model type and various parameters within the model.

The study area chosen for this research, Argentina, provided both complexity as well as opportunities for this research. The multi-crop rotation patterns and no-till agriculture for most part of Argentina resulted in mixed signals of NDVI values. The sensitivity of the NDVI to background reflectance due to previous crop mulch caused NDVI values to differ from the actual plant-based values especially during early crop growth periods. The study found sensitivities in the time-series NDVI curves in the areas where previous crop harvest had shortly preceded soybean plantings. This resulted in difficulty in detecting planting dates for fields where wheat harvest had immediately preceded the soybean crop. Despite the complexity resulting from the intensive crop rotation and no-till agriculture, the MODIS and VIIRS NDVI time-series curves were able to provide monitoring capabilities even at the field level. Therefore, in areas with simpler crop management systems, MODIS and VIIRS NDVI time-series data might be expected to fare better.

Following are some recommendations:

1. The current research was performed with 1/8th degree resolution SALDAS datasets in which each variable was tested separately as well as combined as inputs to the soybean yield model. The soil moisture datasets were also tested for providing initialization values to the model. Given the low resolution, the results could be considered as encouraging for utilizing SALDAS and similar gridded meteorological datasets for improving crop modeling methods. Gridded datasets are available for even higher resolution up to 1 km. Utilizing higher resolution gridded data may provide better results. Therefore, future studies with the utilization of higher resolution of gridded data are recommended.
2. Further research is also recommended for testing SALDAS and similar gridded meteorological datasets for other yield prediction models, as well as for other crop types. Testing of gridded data such as SALDAS for various other crop models for different crop types can provide more validity for such datasets to be used so that integrated use of remote sensing and gridded data can be used with greater confidence for the benefit of regional yield estimation.
3. In this research, both MODIS and VIIRS were sensitive to crop growth patterns. If the prevalent crop rotation patterns are known, time-series MODIS and VIIRS NDVI data seemed to have the potential to provide information on multi-year crop rotation patterns. MODIS and VIIRS resolutions were sensitive to crop residue on ground. Thus, both MODIS and VIIRS resolution showed good application potential for crop rotation monitoring. For Argentina and similar places where no-till agriculture along with double cropping is the predominant

agricultural method, information on previous cropping patterns is important to estimate the amount of reduced water loss in evapo-transpiration which is important for water balancing in crop yield models. In Sinclair model, this reduction in evapo-transpiration is represented by CSEVP value. A value of 0.3 CSEVP is given when previous crop mulch exists and the crop has been identified as maize. The value of 0.3 indicates a 70% reduction in evapo-transpiration due to maize stubbles in relation to having no crop stubbles in the field. The research results show that MODIS time-series NDVI value has potential for classifying crop types using multi-year crop time-series NDVI and for also detecting crop mulch. Further research on the applicability of these time-series datasets for providing multi-year crop rotation based information and for detection of crop mulch is recommended.

4. The research utilized daily fused time-series NDVI for the study and found NDVI to be sensitive to crop mulch. Further research on other types of vegetation indices that are less sensitive to crop mulch and more sensitive to the vegetation are recommended. This might help in finding better onset of greenness.
5. For the planting date estimation, the method used is still qualitative and dependent upon some pre-existing knowledge of field conditions. A mathematical or quantitative approach based upon the change in trajectory or slope of the NDVI values during the growing period of the soybean as future work is recommended.
6. Finally, further research on the utilization of Enhanced Vegetation Index (EVI) to obtain daily time-series values are also recommended.