

University of Redlands

InSPIRe @ Redlands

MS GIS Program Major Individual Projects

Theses, Dissertations, and Honors Projects

5-2008

Using GIS to Predict Corn Yields in Colombia

Manuel Francisco Lemos

University of Redlands

Follow this and additional works at: https://inspire.redlands.edu/gis_gradproj



Part of the [Agricultural Economics Commons](#), and the [Geographic Information Sciences Commons](#)

Recommended Citation

Lemos, M. F. (2008). *Using GIS to Predict Corn Yields in Colombia* (Master's thesis, University of Redlands). Retrieved from https://inspire.redlands.edu/gis_gradproj/177



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material may be protected by copyright law (Title 17 U.S. Code).

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Honors Projects at InSPIRe @ Redlands. It has been accepted for inclusion in MS GIS Program Major Individual Projects by an authorized administrator of InSPIRe @ Redlands. For more information, please contact inspire@redlands.edu.

University of Redlands

Using GIS to Predict Corn Yields in Colombia

A Major Individual Project submitted in partial satisfaction of the requirements
for the degree of Master of Science in Geographic Information Systems

by
Manuel Francisco Lemos

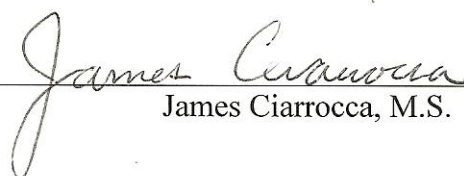
Mark P. Kumler, Ph.D., Chair
James Ciarrocca, M.S.

May 2008

Using GIS to Predict Corn Yields in Colombia

Copyright © 2008
by
Manuel Francisco Lemos

The report of Manuel Francisco Lemos is approved.



James Ciarrocca, M.S.



Mark P. Kumler, Ph.D., Chair

May 2008

ACKNOWLEDGEMENTS

¿Hay en el cielo de Colombia, un coleccionista de nubes?
Is there in the Colombian sky, a collector of clouds?
Pablo Neruda

Members of the University of Redlands MS GIS program, especially Dr. Mark Kumler; thank you for taking the time to understand my project and guide it to accomplish both the University's and my own goals. I know how time-consuming it was to review my paper during the year and the extra dedication you put on it. Jim Ciarrocca, thank you for your sincerity and your effort to make the program a professional and valuable experience. You encouraged me to explore my interests and find a way to unite GIS and business, as I want to. Theresa, thank you for making the MSGIS program feel more like home and your patience for printing and scanning. Lynn Flewelling, thank you for your dedication in finding the right words for my thoughts and fixing thousands of misplaced prepositions.

Helena, thank you for teaching the value of GIS with a great passion, and the true business meaning of "keeping your word". It is a good moment to say how much I like to do what I do. I know I am here because of the opportunities you gave me, your gentle suggestion to pursue my masters degree, and your never-ending support.

Members of the Procalculo Prosis staff, thank you for helping me acquire the almost impossible-to-acquire data for this project. Your help was invaluable. Sandra, thank you for backing me up during the last year.

Momi and Ti, I hope you know how deeply grateful I am for everything you have done and still do for me. I love you. As always, I thank my family for their unconditional support.

Lindi, it is easy now to look back and say it was a great decision. Although it was not easy, I am grateful to you for giving us the change to take the risk. I hope you know how incredibly happy I am to be with you again... and forever.

Bryan, buddy, you can always count on me (especially for that one last beer). Alejandro, ole, muchas gracias por todo, la pasamos bueno. Laura, Alex, Patrick, Jesus, Julie, Becca, Melissa, Andrew, Phil, and Ruth, I am lucky to find friends like you. I will miss you but I will always thank you for making this year an unforgettable experience. I hope to see you often. Focus...

ABSTRACT

Using GIS to Predict Corn Yields in Colombia

by
Manuel Francisco Lemos

Crop yield prediction can play an important role in developing the agriculture sector in Colombia. Remote sensing and GIS have proven to be an effective mechanism for this purpose in developed economies. This project created a proof-of-concept application for the Colombian Ministry of Agriculture and other related governmental institutions. The project used existing methodologies including the classification of satellite imagery, interpolation of climate data into continuous surfaces, the extraction of Normalized Difference Vegetation Index, and the computation of multiple linear regressions. ESRI ArcGIS provided the interface, software, tools and functions to build the application, and to integrate and automate the application's functionalities.

Cloud coverage in the imagery and the lack of specialized data affected the accuracy of the crop yields estimates. Nevertheless, the application predicts corn yields with an estimated accuracy of 71% when cloud coverage is minimal. The application can use both Landsat and Spot preprocessed images, and in less than six minutes yield predictions for areas inside Cordoba, a major corn producing state in Colombia.

Table of Contents

1.	Introduction.....	1
1.1.	Problem Statement	1
1.1.1.	Economic Considerations	2
1.2.	Client	2
1.3.	Needs Analysis.....	3
1.3.1.	Key Government Agriculture Customers	3
1.4.	Justification	4
1.5.	Crop and Study Area	5
1.5.1.	Selection Criteria	5
1.5.2.	Corn Production in Colombia	5
1.5.3.	Area of Study (AOS) Selection.....	7
1.6.	Proposed Solution	10
1.6.1.	Functional Specifications.....	10
2.	Background and Literature Review	13
2.1.	Crop Yield Prediction.....	13
2.2.	GIS and Remote Sensing Based Crop Yield Prediction Models	14
2.2.1.	NDVI.....	15
2.3.	GIS Based Crop Prediction Systems	15
2.3.1.	Developed Countries.....	16
2.3.2.	Developing Countries	16
2.3.3.	Colombian GIS Crop Prediction Systems.....	17
2.4.	Accuracy.....	18
3.	Data.....	19
3.1.	Remotely Sensed Imagery.....	19
3.1.1.	Landsat	20
3.1.2.	Spot	21
3.1.3.	Terra.....	22
3.1.4.	Other Imagery Sources	22
3.1.5.	Spectral Bands Selection.....	22
3.2.	Crop Yield Statistics.....	23
3.3.	Weather Data.....	25
3.4.	Ancillary Data	26
4.	Methodology.....	29
4.1.	Spatial Information Database	29
4.1.1.	Database Design.....	29
4.1.2.	Metadata Edition.....	30
4.1.3.	Spatial Frame Definition.....	30
4.1.4.	Raster Data Population	33
4.1.5.	Vector Data Population.....	34
4.1.6.	Weather Data Population	35
4.2.	Imagery Processing	36
4.2.1.	Image Classification.....	36
4.2.2.	NDVI Calculation	41
4.2.3.	Interpolation.....	42

4.3.	Multiple Linear Regression (MLR) Analysis.....	45
5.	Model-GIS Integration and Results	51
5.1.	General Considerations	51
5.2.	AOI Selection Model	52
5.3.	Interpolation Model.....	53
5.4.	NDVI Model	54
5.5.	Classification Model	55
5.6.	Integration Model.....	57
5.7.	Result.....	59
6.	Conclusions and Recommendations	61
6.1.	Recommendations	61
7.	References.....	63

Table of Figures

Figure 1-1 - Colombian farming landscape (Author)	1
Figure 1-2 - General corn phenology (Adapted from Ministerio de Agricultura et al., 2001)	6
Figure 1-3 - Corn planting and harvesting percentages calendar in Colombian hot climates (Ministerio de Agricultura et al., 2001)	7
Figure 1-4 - Cordoba location in Colombia, South America	8
Figure 1-5 - Corn production in Cordoba, 2005	9
Figure 3-1 - Terra, Landsat, and Spot images over the study area	20
Figure 3-2 - Landsat (A) and Spot (B) scene grids over Cordoba, Colombia	21
Figure 3-3 - Corn production in Cordoba (1987 -2005) Traditionally cropped (A) Technically cropped (B)	24
Figure 3-4 - Comparison between traditionally and technically cropped corn in Cordoba (1987-2005).....	25
Figure 3-5 - Original weather reading tables by station (IDEAM, 2007b).....	26
Figure 4-1 - An ArcGIS Model for Agriculture – Draft (Adapted from ESRI, 2003)	30
Figure 4-2 - Test Image One, Landsat	32
Figure 4-3 - Test Image Two, Spot.....	33
Figure 4-4 - January 1988 temperature interpolations using original locations in black (A) and corrected locations in red (B)	36
Figure 4-5 - Classification of Test Image One identifying clouds in the AOI	37
Figure 4-6 - Three different classifications of Test Image Two for 15, 20 and 25 classes.....	38
Figure 4-7 - Erdas Imagine Signature Editor of Test Image One	40
Figure 4-8 - Windows arrangement required for simulating ground truth from ancillary data.....	40
Figure 4-9 - Weather stations location extending beyond Cordoba boundary	43
Figure 4-10 - IDW interpolation of temperature data including null values (A) and removing null values (B)	44
Figure 5-1 - ArcGIS provides the software and the correspondent user interface.....	51
Figure 5-2 - AOI Selection model diagram	52
Figure 5-3 - Result from the AOI Selection Model using Test Image Two	53
Figure 5-4 - Interpolation model diagram.....	54
Figure 5-5 - Result from the Interpolation Model using rainfall readings in August 2008	54
Figure 5-6 - NDVI model diagram	55
Figure 5-7 - Result from the NDVI Model using Test Image Two	55
Figure 5-8 - Classification model diagram	56
Figure 5-9 - Result from the Classification Model using Test Image Two	57
Figure 5-10 - Regression model diagram.....	57
Figure 5-11 - Integration model diagram.....	58
Figure 5-12 - Areas planted with corn displayed by the estimated yield of each cell	58
Figure 5-13 - Corn Yield Report inside AOI in August, 2006	59

List of Tables

Table 3-1 - Spectral band comparison among Landsat 7, Spot 5, and Terra	23
Table 3-2 - Processed weather stations' readings	26
Table 4-1 - Soil use raster values and description	34
Table 4-2 - Subtypes for Roads and Populated Areas feature classes	35
Table 4-3 - Band's names for NDVI calculation from different sensors.....	41
Table 4-4 - Summary of rainfall estimation errors by interpolant (best result highlighted)	45
Table 4-5 - Summary of temperature estimation errors by interpolant (best result highlighted)	45
Table 4-6 - Data used in multiple regression analysis	46
Table 4-7 - Results of a MLR of yield rate on rainfall, temperature and NDVI (2001- 2005)	47
Table 4-8 - Results of a MLR of yield rate on rainfall and temperature (1988 -2005)	48
Table 4-9 - Results of a MLR of NDVI on rainfall and temperature (2001-2005)	48
Table 4-10 - Data including estimated NDVI for years 1997 to 2000.....	49
Table 4-11 - Results of a MLR of yield rate on rainfall, temperature and estimated NDVI (1997 -2005).....	50
Table 5-1 - Available signature files.....	56

List of Acronyms

ANOVA:	Analysis of Variance
AOI:	Area of Interest
AOS:	Area of Study
AVHRR:	Advanced Very High Resolution Radiometer
BNA:	Bolsa Nacional Agropecuaria (Colombian Agricultural Exchange Board)
CCAP:	Crop Condition Assessment Program of Statistics
DEM:	Digital Elevation Model
ESRI:	Environmental Systems Research Institute
ETM+:	Enhanced Thematic Mapper Plus
GIS:	Geographic Information Systems
HRG:	High Resolution Geometric
IDEAM:	Instituto de Hidrología, Meteorología y Estudios Ambientales (Colombian National Institute for Hydrology, Meteorology and Environmental Sciences)
IDW:	Inverse Distance Weighting
IGAC:	Instituto Geográfico Agustín Codazzi (Colombian National Geographic Entity)
IT:	Information Technologies
MARS:	Monitoring Agriculture with Remote Sensing
MLR:	Multiple Linear Regression
MODIS:	Moderate Resolution Imaging Spectroradiometer
NASS:	National Agriculture Statistics Service
NDVI:	Normalized Difference Vegetation Index
NOAA:	National Oceanic and Atmospheric Administration
ROI:	Return Over Investment
USDA:	United States Department of Agriculture

1. Introduction

Increasing agricultural yields in Colombia requires implementing better technologies. Geographic Information Systems (GIS) proved to be a valuable investment in agro-production in other developing countries, but in Colombia lack of awareness of the potential benefits of GIS and the limited political involvement is delaying progress. The objectives of Procalculo Prosis, the Environmental Systems Research Institute (ESRI) distributor in Colombia, include closing that gap. From a business/academic approach, this project used ESRI ArcGIS software to build a proof-of-concept GIS application focused on showing GIS advantages to the diverse Colombian agriculture industry (see Figure 1-1).



Figure 1-1 - Colombian farming landscape (Author)

This chapter expands on the problem encountered, the vision and sales strategy proposed by Procalculo Prosis, the GIS advantages for decision-makers, the needs and customs of the Colombian agriculture sector, and the proposed solution for the problem.

1.1. Problem Statement

After reviewing the statistics prepared by the Colombian Ministry of Agriculture (Ministerio de Agricultura, 2004), it is easy to see the importance of the agricultural sector to Colombia's economy. Agricultural activities in the country provide income for 27% of the national population, offer 23% of national employment, and contribute 14%

of the gross national production. It is a key economic sector which, however, lags socially; 43% of the rural population lives in poverty. Although the economic participation of the agricultural sector is one of the highest compared to other Latin-American countries, poverty and poor adoption of technology severely hinders the rural sector's global competitiveness.

The Colombian government stated that the main justifications for implementing better technologies are to transform Colombian agriculture to be globally competitive, and to strengthen the agriculture as an element that contributes to the weakening of the drug traffic and terrorism in the country (Ministerio de Agricultura, 2004).

1.1.1. Economic Considerations

Colombia is the number one producer of coca plants in the world. The fact that the country is fighting a war against drug production is widely known. The associated risk, not equally well known, is that farmers may switch to illegal farming if the legal crops cannot sustain farming as a long-term business. The economic incentive of growing illicit crops expands illegal agriculture and fuels the armed conflict. The country needs to support farmers who produce legal crops.

A forthcoming foreign trade agreement with the United States amplifies the complexity of the situation. This trade agreement is important in developing the Colombian economy and generating mutual benefits to both countries. However, its effect on non-competitive products, such as corn and soy, will be devastating due to dumping practices and low internal fixed prices. The only alternative is to make agricultural business more competitive. This requires investments within the sector to compensate for the weather advantages and the subsidies common in other countries (Ministerio de Agricultura, 2004).

Attracting foreign investment to Colombia's agriculture, where the laws of supply and demand fuel the volatility of agricultural products' prices, requires tools that help assure revenue to the investors. Crop yield models may help mitigate the uncertainty in a business where weather and natural calamities play a crucial role in determining the food chain supply. Specifically in Colombia, where extreme weather affects the country at least once a year, the need for tools such as prediction models and management systems becomes obvious.

1.2. Client

Procalculo Prosis S.A. (2007) is a Colombian company, founded in 1968 to implement Information Technologies (IT) in the country. In 1988, Procalculo Prosis initiated the distribution of ESRI products in Colombia and, in 2000, won Leica-Erdas exclusive distribution. During the last twenty years, the core business of Procalculo Prosis S.A. has been to distribute, develop, integrate, deploy, train, support and customize GIS in all related fields, applications, and systems in Colombia. The focus on GIS permitted Procalculo Prosis to win the contract to provide GIS software and services to the Instituto Geográfico Agustín Codazzi (IGAC, for the Spanish Acronym of Colombian National Geographic Entity). IGAC is the Colombian government producer of geographic data; it defined the national standards for GIS and began to use ESRI technologies as its main GIS technologic platform. Based on this success, Procalculo Prosis has continued to grow

its market share in the government to become the leading vendor in the country (Procalculo Prosis S.A., 2006).

Market analysis, marketing, person-to-person service, pre-sales attention, long-term relationship management, and the search for quality all fuel the sales process and, consequently, the success of the company. As the GIS market leader, Procalculo Prosis strategic sales focus concentrates on the government sector and large companies. Supporting key customers to achieve a competitive advantage or a strategic goal using GIS is the main pillar of Procalculo Prosis' commercial strategy (Procalculo Prosis S.A., 2006). These customers serve as pioneers and they often motivate other customers to adopt GIS technologies.

Procalculo Prosis often develops business relationships, first by demonstrating the advantages of GIS to a prospective customer's IT staff. IT staff normally embrace the technical benefits of GIS, but achieving the same involvement level in the decision-making employees can take months and requires frequent assistance by technical presentations, demonstrations, return over investment (ROI) analyses, and other pre-sales support. The ultimate goal is to demonstrate that GIS is an important tool that will support the decision-making process in the organization.

Customer's internal GIS dissemination and the required sales process may take months or even years. Keeping Procalculo Prosis' leading market position requires not only resilient dedication, but also deep understanding of both the market and the customer's needs. Promoting the development of projects like this one, which helps new customers maximize their ROI, is part of the ongoing effort to satisfy customers (Procalculo Prosis S.A., 2006).

1.3. Needs Analysis

By 2006, GIS applications developed in Colombia had successfully reached many key economic sectors, including petroleum, utilities, land management, environment, and education (Procalculo Prosis S.A., 2006). Compared to similar countries, the deployment of geographic information technologies in Colombia is satisfactory. Still, Colombia's circumstances and needs offer a wide spectrum for new and creative GIS implementations. Specifically, the agricultural sector at central government level has been particularly difficult to reach with GIS because of the low level of GIS awareness within almost every important agriculture organization. Some large private agro-producers use GIS for day-to-day operations, but in general, the country is overdue to take advantage of the full potential of GIS to create significant opportunities to implement GIS in Colombia. The agriculture sector has faced ongoing problems and government institutions have implemented policies with questionable success. However, the tools required to control, analyze, and promote these policies are still slowly deployed when compared to developed countries. Procalculo Prosis expects to help by bringing GIS to key government agriculture customers.

1.3.1. Key Government Agriculture Customers

Since 1979, the Bolsa Nacional Agropecuaria (BNA, for the Spanish acronym of Colombian Agricultural Exchange Board) has developed and maintained a derivatives market of agricultural products and services. Its main shareholders are: the Ministry of

Agriculture (38.06%); the domestic private sector (59.14%); and the foreign private sector (2.8%) (Bolsa Nacional Agropecuaria, 2007). One of BNA's goals is to assure, through derivatives financial instruments, stable prices for both producers and customers. Moreover, supply estimation should be a key component of the BNA functions.

The Colombian Ministry of Agriculture is the government entity responsible for formulating, coordinating, and evaluating the policies that promote the sustained development of agriculture in Colombia (Ministerio de Agricultura, 2004). However, the Ministry needs better methods to estimate production, and to prevent and measure the impact of natural phenomena, such as El Niño and La Niña. The Ministry of Agriculture also needs a prediction model to fulfill its goal of assessing the effectiveness of its strategies, attracting foreign investment in the industry, and measuring the impact of natural phenomena.

1.4. Justification

Researchers developed many links between GIS and agriculture thirty years ago. This relationship covers a wide set of fields and applications. Andersen, Pandya-Lorch, and Rosegrant (1999) affirmed that increasing agricultural productivity is the most common global need associated with the small-scale farmers. The use of IT, biotechnology, and precision farming can help small farmers improve price stability, access finance and markets, obtain technical assistance in quality improvement and diversification, and improve organizational strength. Colombia should be no exception; small-scale farmers (the average size of an individual farm in Colombia is around two hectares) account for 40% of the country's agro-production.

Longley et al. (2005) stated that GIS could benefit investigation and management by storing enormous amounts of data, integrating the data from different sources, recording the events and processes, supporting the control and evaluation of strategies and policies, and effectively communicating results and useful information to farmers. Specifically in agriculture, a GIS supports the government decision-making and controlling processes. Syam & Jusoff (1999), Bouman (1995) and Pinter et al. (2003) listed some supported topics, such as enforcement of production quotas, restructuring of farm systems, subsidies allocation, crop area assessment and monitoring, disaster management/prevention, creation of data inventories such as land uses and soil characteristics, weather prediction, crop insurance, monitoring crop rotation techniques, and projecting soil loss from individual farms to country level.

Inside the wide agriculture management field, many authors agreed that yield forecasting at regional scales is critical to regional, national, and worldwide economies (Aronoff, 2005; Bouman, 1995; Hayes, O'Rourke, Terjung, & Todhunter, 1982) mainly because it allows decision-makers to control food management processes and evaluate and improve farming strategies (Yang & Anderson, 2000). Underdeveloped countries are especially susceptible because of their increased economic dependence on agriculture and the socioeconomic impact of food shortages.

Syam & Jusoff (1999) considered agriculture a changing phenomenon governed by the complicated interaction of a crop with the environment. Based on this idea, they proposed integrating remote sensing, crop phenology, and weather properties into a GIS, which

could support operational applications. In addition, they justified the use of GIS and remote sensing technologies to obtain accurate results in the dynamic field of agriculture. They compared different projects' costs using GIS and remote sensing technology against conventional cost and concluded that an increase in the number of projects based on these technologies maximize the benefits. They stated that the four benefits of implementing GIS are: scientific, technological, methodological, and economic. For example, finding new facts is scientific, increasing the work productivity is technological, increasing the accuracy is methodological, and reducing the costs is economic. As a conclusion, they highlighted the fact that satellite images provide the most beneficial method of acquiring data.

1.5. Crop and Study Area

This section initially explores the criteria used to select corn as the objective crop, and Cordoba as the area of study. This selection was a necessary step to define the acceptable scale (state), minimize the amount of data required, and reduce the time needed to execute the models in the prototype. Secondly, the section presents crop characteristics and a small socio-economic description of the study area.

1.5.1. Selection Criteria

Crop selection took into account the size of the area planted to increase the probability of data acquisition. This step also considered the existence of extended local crop research to validate the results, and the crop's economic importance to assure a high impact in the client's customers. According to the Colombian Ministry of Agriculture (Ministerio de Agricultura, 2004), corn planted lands (14.35%) are second only to coffee (20.8%) in Colombia. In addition, corn is the second most traded commodity (after rice) in the BNA (Bolsa Nacional Agropecuaria BNA, 2007). Coffee was excluded from this analysis, because the Federación Nacional de Cafeteros de Colombia (Colombian National Federation of Coffee Growers), a special institution concerned with coffee development, commercialization, and management, is already using GIS technologies.

The selection of the state of Cordoba as the area of study (AOS) was based on corn production statistics. Cordoba has more land planted with corn than any other state in the country. In addition, this study area has an acceptable availability of cloud-free imagery and ancillary data compared to the rest of the country. Furthermore, the state's social needs are interesting for the customers.

1.5.2. Corn Production in Colombia

Colombian corn production is classified into two types: traditionally-cropped and technically-cropped. Traditionally-cropped production refers to planted areas smaller than five hectares. Technically-cropped refers to the crop system in areas greater than five hectares. In general, technically-cropped areas cover plain terrains, use improved seeds, and mechanized systems for terrain preparation, crop, and fertilization. Eighty five percent of the country's planted area is traditionally-cropped thus creating employment for 190,000 families (Ministerio de Agricultura, Departamento Administrativo Nacional de Estadística, Federación Nacional de Arroceros Fedearroz, & Usocoello, 2001).

Corn grain characteristics divide Colombian corn into twenty varieties: nine with white grain and eleven with yellow grain, with different leaf colors. Humans consume mainly the white grain varieties, while the yellow grain is used primarily as animal food. The number of varieties allows corn to be grown in a wide range of ecological conditions. In general the ideal conditions for corn plants are a temperature range of 13 to 29 degrees Celsius, annual rainfall between 2,900 and 6,000 mm, altitude from 0 to 3500 meters, and soils with “medium textures, granular structure, loose, fertile, with good drainage, and a ph between 5.5 and 7.0” (Ministerio de Agricultura et al., 2001). Figure 1-2 illustrates the general corn phenology.

		Establishment	Vegetative development	Flowering	Harvest		
Illustration							
Number of days							
Optimal growing conditions	Temperature (°C)	17-22	18-30	18-30	18-30	20-32	
	Water requirements (mm)	80-120	180-230	130-170	120-150	90-120	
Extreme Thresholds	Temperature (°C)	Below 15	Below 15	Below 15	Above 35	Below 15	
	Water requirements (mm)	Below 70	Below 150	Below 150		Above 35	

Figure 1-2 - General corn phenology (Adapted from Ministerio de Agricultura et al., 2001)

Servilla & Towner (2000) affirmed that crop selection relies on the identification of the best crop varieties’ performances for a given region’s climate. Because of the wide-ranging climate regions in Colombia, a farmer may plant different corn varieties over the course of the year. Different corn varieties have different leaf colors, time need for maturing (from two to eleven months), planting schedules, and other phenological characteristics. Consequently, the lack of a known corn planting/harvesting calendar complicates corn yield predictions using remote sensing, because the spectral signature required for classification is different at every stage and many crops share similar signatures during different phenological stages (United Nations, 2006). In the Cordoba’s hot climate, farmers generally plant corn twice a year, between April and May, and again between August and September (see Figure 1-3), but their calendars are not consistent because of the differences in corn types and seeds, as well as various market factors.

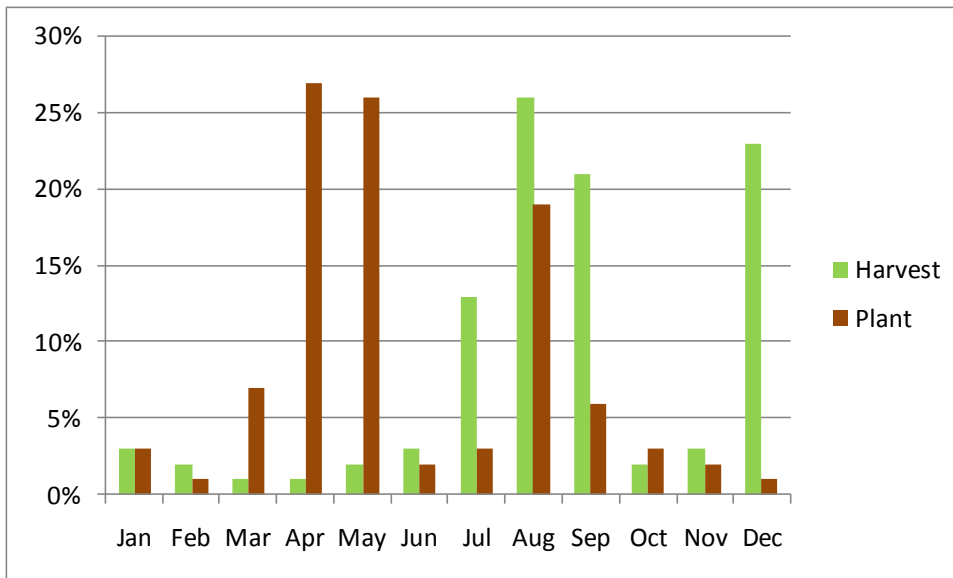


Figure 1-3 - Corn planting and harvesting percentages calendar in Colombian hot climates (Ministerio de Agricultura et al., 2001)

Finally, corn productivity in Colombia averages 5.5 tons/hectare, compared to an average of eight tons/hectare in the United States (Ministerio de Agricultura, 2004). Moreover, seeds and fertilizers increase production costs in Colombia by approximately 50%. With the approval of the foreign trade agreement between the two countries, the Colombian market is expected to receive an additional 1.5 million tons of corn from the United States, reducing local corn price by 10% (Ministerio de Agricultura, 2004; Ministerio de Agricultura & IICA, 2006).

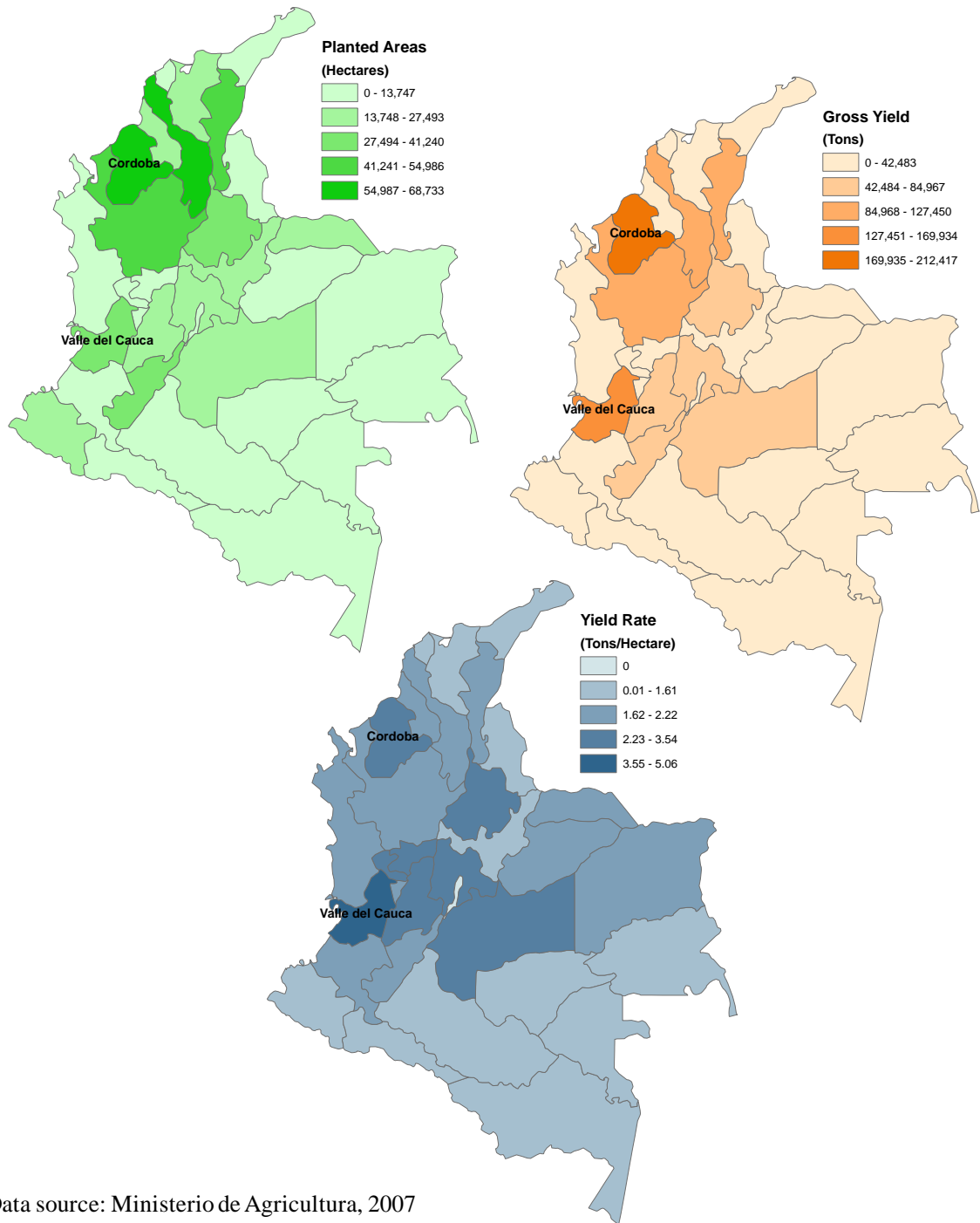
1.5.3. Area of Study (AOS) Selection

Cordoba is a coastal state in northwestern Colombia, located between 7° 22' and 9° 26' north latitude, and between 74° 47' and 76° 30' west longitude (See Figure 1-4). The state is divided into twenty-eight administrative municipalities, one of which, Monteria, is the state capital. The climate is generally tropical, hot, and rainy. The area of Cordoba is about 2,500,000 hectares. These characteristics are ideal for the development of pastures, extensive cattle farming, and highly commercial but transitory agricultural products, such as cotton or corn.



Figure 1-4 - Cordoba location in Colombia, South America

The Colombian Ministry of Agriculture collects and publishes corn production figures for Colombia (Ministerio de Agricultura, 2007). In 2005, Cordoba had the greatest area planted in corn (68,739 hectares), as well as the greatest total yield (212,417 tons). However, the Valle del Cauca state had a higher yield per planted hectare (5.05 tons versus Cordoba 3.09); this is likely the result of better soils, as well as the use of more efficient technical cropping systems (Viloria de la Hoz, 2004) (see Figure 1-5).



Data source: Ministerio de Agricultura, 2007

Figure 1-5 - Corn production in Cordoba, 2005

In contrast, Vilorio de la Hoz (2004) describes the social situation in Cordoba as similar to the rest of the country (see Section 1.1). Cordoba has a population of 1,380,000 inhabitants, growing at a rate near 2.1% annually, with 50% of the population living in rural areas. The state is one of the poorest, with 30% of the population considered by the government to be “poor” and an additional 35% living in “miserable conditions”

(collectively, 65% of the population living below the government designated “line of poverty”). This situation illustrates the great need for agriculture investment and development required within Cordoba.

1.6. Proposed Solution

Predictions can be used to test scientific understanding, but in this project the objective of the prediction was “to serve a decision making process” (Sarewitz et al., 2000). GIS are powerful decision-making tools that can measure and communicate many of the customers’ possible activities. In consequence, the proposed solution was a GIS prototype, which estimates corn crop yield in the Colombian state of Cordoba. The prototype uses information from remote sensing imagery to identify cultivated areas, and includes a linear regression model, which estimates yields based on temperature, rainfall, and vegetation health.

For this prototype, the non-functional requirements were: single-user, reliability, performance (image display and processing times), ease of use, and portability (stand-alone). The area of study and product data used was representative of customers’ needs. The technology required for this GIS was not new or particularly complex. Off-the-shelf software, such as ESRI ArcGIS and Erdas Imagine, provided sufficient functionality to construct the prototype and to integrate processes such as data preparation, spatial information storage, imagery classification, and modeling. The application did not require additional programming.

1.6.1. Functional Specifications

Four categories classified the functional specifications. The categories were: application, presentation, data storage, and data access.

1.6.1.1. Application Requirements

- Import, prepare, compose, interpret, and classify unprocessed imagery
- Create and store classification parameters (detailed training sites and schemes)
- Use training sites to classify image pixels into specific planted areas (by product) and produce accuracy statistics for the image classified
- Read different data formats and organize the information in a geographic database
- Retrieve the data from the geodatabase
- Combine data
- Compress the data
- Provide analysis tools to validate and understand the data
- Provide a tool to build the crop yield estimation model
- Run the model that estimates crop yield based on rainfall, temperature and vegetation health
- Produce reports that helps the decision-making process
- Integrate both vector and raster data

1.6.1.2. Presentation Requirements

- Provide the interface to permit the user to analyze, modify, classify, and change the data and parameters of the model
- Receive estimation parameters directly from the user
- Display/print a report that includes a map of the area of interest, a parameter/data summary, and accuracy estimative

1.6.1.3. Data Storage

- Store data in raster and vector forms
- Store the parameters/results of and estimation
- Store the related metadata

1.6.1.4. Data Access

- Retrieve relevant information for an area of interest based on the metadata or attributes
- Easily modify the information
- Create and delete data

2. Background and Literature Review

This chapter describes how the statistical methods, field procedures, and technologies required for the project's implementation have been developed and extensively used by others. This chapter also justifies the prototype's use of those methods, procedures, and technologies in Colombia, where extreme circumstances such as poor data availability and high levels of cloud coverage reduce potential accuracy. Four sections outline this chapter. The first section explores crop yield prediction. The second section focuses on GIS and remote sensing prediction alternatives. The third section presents the most relevant experiences using GIS and remote sensing around the world. The last section presents alternatives to understanding and measuring accuracy, data availability, and cloud coverage. However, this background does not consider other human and economic factors that could influence crop yield predictions, such as GIS specialist skills, institutional collaboration, or budget restrictions.

2.1. Crop Yield Prediction

Bouman (1995) defined yield prediction as the estimation of what the yield of a given crop will be. Crop prediction models seek to simulate crop development based on physical, chemical, and physiological processes. Sarewitz, Pielke, & Byerly (2000) stated that increased computer power and extensive new data about the environment might dramatically improve prediction analysis. Crop yield estimations rely on various methods ranging from expert knowledge (non-parametric), crop-growth simulation models, trend analysis, regression analysis, and statistical models.

Since Bleasdale & Nelder (1960) first estimated the explicit relationship between areas planted and crop yield, many researchers have added other parameters to agricultural prediction. Watson's study (1963), as cited in Prasad, Chai, Singh, & Kafatos (2006), explained the relationships between weather variables and crop development. The study of Gallaguer and Biscoe (1978), as cited in Bouman (1995), presented one of the most common prediction models. Known as the 'light interception' model, it produces a good indicator of potential yield by considering solar radiation, light use efficiency, and fraction light interception, which is a function of Leaf Area Index; the ratio of one-half the total area of leaves to the total surface area containing that vegetation (Steffen, 2007).

Regression analysis is a powerful method to extract relationships from data and easily adapts to additional, new, or more accurate data (Rogerson, 2006). Consequently, it can be used to help solve crop prediction problems. For example, the World Meteorological Organization (1982) used it to summarize the relationships between climate factors such as solar radiation, temperature, rainfall and crop yields. Nualchawee's study (1984), as cited in Syam & Jusoff (1999), related yields to agronomic variables such as maturity, density, vigor, etc. The study of Garcia-Paredes, Olson, and Lang (2000), as cited in Prasad et al. (2006), added soil type as a predictor variable.

Other more complex models consider additional variables such as phenology, "the study of vegetation dynamics in terms of climatically-driven changes that take place over a growing season" (Steffen, 2007), as well as photosynthesis processes, respiration, soil erosion, and productivity to achieve better accuracy (Bouman, 1995). Some examples of

those models are: Water Response model, EPIC, CERES and CROP-GRO (based on the DSSAT system), WOFOST, SIMTAG, ARCWHEAT, and SUCROS (Bouman, 1995).

2.2. GIS and Remote Sensing Based Crop Yield Prediction Models

GIS can handle and analyze data from different scales, sources, intents, coordinate systems, or formats (Longley et al., 2005). Similarly, in agriculture, GIS has proven to be useful to create homogeneous land units where parameters such as weather, soil properties, and agricultural strategies can be simplified and considered uniform inside specified area (Bouman, 1995). Logically, models and GIS converged. The models proved to be useful in predicting crop yields in different scales and planet locations, and GIS easily integrated different fields under one common term: geographic location. Various cultural motives and technical requirements influenced the integration of models with GIS and remote sensing techniques. Heath's study (1990), as cited in Bouman (1995), enumerated among the cultural motives the lack of consistency between regions and countries, human subjective appreciation, and off-time. Heath included among the technical requirements the reduction of uncertainty and the conclusion that multi-spectral imagery data could effectively be used to estimate planted area.

On the other hand, remote sensing has been effective in monitoring crop-growing conditions and estimating crop yields since the 1970s. Yield estimation uses remotely sensed data in different but complementary ways. It derives parameters that are directly related to yield, such as area (Maas, 1988). It has been used to estimate biometric parameters, such as the Normalized Difference Vegetation Index (NDVI), which are incorporated as other model variables to improve accuracy (Sadler & Russell, 1995; Syam & Jusoff, 1999). Today, an obvious solution to crop yield estimation is calculating cultivated areas from satellite imagery analysis.

It is possible to estimate differences in regional crop production using Landsat MSS imagery (Badhwar, 1984; Bauer, Cipra, Anuta, & Etheridge, 1979; Williams, 1990). The study of Colwell et al. (1977), as cited in Yang & Anderson (2000), found and specified the direct correlation between Landsat imagery and crop yields using traditional sampling techniques. Wiegand, Richardson, & Kanemanesu's study (1979), as cited in Pinter et al. (2003), related Leaf Area Index measurements of winter wheat to a normalized vegetation index derived from Landsat MSS data. The studies of Tucker et al. (1980) and Wiegand & Richardson (1984), as cited in Pinter et al. (2003), found, using linear regressions, that grain yields relate significantly to the red and near-infrared bands data, nine spectral measures, five vegetation indices, and three individual bands from Landsat MSS data. The study of Hayes et al. (1982), as cited in Prasad et al. (2006), presented a crop yield model highly dependent on NDVI. Wiegand & Richardson's study (1984), as cited in Pinter et al. (2003), defined the equations for relating spectral observations to crop growth and yields. Bouman's study (1992), as cited in Bouman (1995), improved the accuracy of SUCROS crop models with remote sensed data. The study of Moulin, A. Bondeau, and R. Delécolle (1998), as cited in Pinter et al. (2003), combined satellite imagery and crop-growth models.

Although special considerations like scale phenomena, spatial correlations, or error normalization influence how regression analysis and geographic sciences complement each other (Rogerson, 2006), regression analysis methods helped to prove the

relationships between remotely sensed imagery and crop yields. The study of Tucker et al. (1980), as cited in Pinter et al. (2003), established that 64% of the grain yield variability could be explained by the variations of remotely sensed imagery. The study of Murthy et al. (1995), as cited in Syam & Jusoff (1999), concluded that the correlation existing between yield and NDVI depends on the phenological stage of the crop. Hence, the relationship depends on the date of the image taken. The study of Garcia-Paredes, Olson, and Lang (2000), as cited in Prasad et al. (2006), explained corn and soybean productivity using multiple linear regression (MLR) on soil classification data. Shanahan et al. (2001) used vegetation indices and showed that the best growth stage to estimate corn yields is midway through the grain fill period, with residual sum of squares (R^2) near 0.80.

Servilla & Towner (2000) described how historic weather information could define crop selection and development. Prasad et al. (2006) predicted corn and soybean yields using a non-spatial linear regression method with breakpoints. They used NDVI, soil moisture, surface temperature, and rainfall as the predecessor variables required to obtain predicted values. They assessed the accuracy comparing the predicted yields against observed yields with residual sum of squares (R^2) greater than 0.78.

2.2.1. NDVI

As observed in the last section, many researchers found a correlation between NDVI and crop yields. NDVI is the most widely used vegetation index (Towson University, 2007). Vegetation indices are the product of transforming raw image data into an entirely new image by calculating an algorithm. The algorithms highlight relationships among multiple bands of multi-spectral imagery.

Like most other vegetation indices, NDVI is a measure of vegetation health or “greenness” (USGS, 2007). The chlorophyll in leafy green vegetation and the density of green vegetation on the surface affect the absorption of visible light, especially red light, and the reflection of near-infrared energy. Consequently, NDVI derives from the ratio of the difference to the sum between measured reflectivity in the red and near-infrared portions of the electromagnetic spectrum. This ratio maximizes the contrast between vegetation and soil and compensates external effects like sunshine. “Healthy vegetation will have a high NDVI value. Bare soil and rock reflect similar levels of near-infrared, red, and so will have NDVI values near zero. Clouds, water, and snow are the opposite of vegetation in that they reflect more visible energy than infrared energy, and so they yield negative NDVI values” (USGS, 2007).

$$NDVI = \frac{Near-infrared - Red}{Near-infrared + Red}$$

“NDVI equation produces values in the range of -1.0 to 1.0, where vegetated areas will typically have values greater than zero and negative values indicate non-vegetated surface features such as water, barren, ice, snow, or clouds.” (Towson University, 2007)

2.3. GIS Based Crop Prediction Systems

Given the objectives of this project, it was relevant to explore some of the crop prediction systems that agriculture policy-makers have deployed around the world. This section

presents state-of-the-art systems, the systems created by underdeveloped countries, and a similar application used in Colombia.

2.3.1. Developed Countries

The research mentioned earlier in this chapter sustained the development of prediction systems to solve national crop yield predictions needs. Monitoring Agriculture with Remote Sensing (MARS) is the program of the European Union Joint Research Centre for the improvement of agricultural statistics (Bouman, 1995; Hanuschak & Delince, 2004). The National Agriculture Statistics Service (NASS) program was developed by the United States Department of Agriculture (USDA) (Allen, Hanuschak, & Craig, 2002; Yang & Anderson, 2000). Canada established the Crop Condition Assessment Program of Statistics (CCAP) (Aronoff, 2005). These three programs were among the first to develop applied crop growth modeling based on remote sensing.

In their executive summary, Hanuschak & Delince (2004) described the three main applications of remote sensing inside NASS. The first application had used Landsat imagery since 1978 as samples in the agricultural estimation statistics, with complete coverage in the United States. Crop acreage estimation is the main application of NASS. The second application uses satellite imagery to improve the statistical precision of crop acreage indicators in some states. The third application is the formation of a public-use GIS data file called the Cropland Data Layer.

“The Cropland Data Layer is the crop specific categorization of the best available set of Landsat (30 meter resolution) digital imagery for the crop(s) season of interest. Users have recently used the Cropland Data Layer to aid in watershed monitoring, soils utilization analysis, agribusiness planning, crop rotation practices analysis, animal habitat monitoring, prairie water pothole monitoring, and in the remote sensing/GIS value added industry.” (Hanuschak & Delince, 2004)

Reichert & Caissy (2002) described how CCAP has employed the Advanced Very High Resolution Radiometer (AVHRR) sensor for vegetation to forecast crop yields and land conditions since 1989. Its forecasts predict, within an accuracy of 5.6%, the statistics produced two months later from field reports; CCAP found significant discrepancies only in years with extreme weather conditions. An important aspect of this project is the methodology developed to remove clouds. The CCAP uses seven-day composed NDVI images; if any additional cloud pixel is present, the system automatically removes it from the analysis. In addition, the system is the perfect example of seamless integration of remote sensing imagery and GIS software. The customization allows CCAP to automatically prepare and distribute reports that include graphs, maps, and comparison charts, and to run individual analyses over the Internet. Specifically, reports that include a comparison chart against a 15-year average plotting NDVI (Aronoff, 2005).

2.3.2. Developing Countries

Syam & Jusoff (1999) summarized how Malaysia, a country highly dependent on its agriculture, started the National Resource and Environment Management program (NAREM) in 1996. The program's objective is to operate a GIS that supports and integrates national resource planning by collecting, classifying, storing, and retrieving

natural resource data. The Malaysian program proved that despite vast cloud coverage and thick rain, the system is helpful in planning and controlling agricultural development.

The MARS FOOD program, developed by the Joint Research Centre of the European Commission, supports Eastern African countries (Rojas, Delince, & Leo, 2006), assessing countrywide yield prospects for maize and sorghum by analyzing two indexes: NDVI and Water Requirement Satisfaction Index, which is tightly linked to rainfalls. The same program is helping Argentina, Bolivia, Brazil, Chile, Paraguay, and Uruguay develop similar programs, and share their experiences (Grilli, Ravelo, Santa, & Zanvettor, 2006; Ravelo, Casa, Ovando, & Planchuelo, 2006; Ravelo, Rojas, Negre, Cherlet, & Planchuelo, 2002).

Brazil is a pioneer in Latin America with its Geosafra program (Conpanhia Nacional de Abastecimineto CONAB, 2007). The program, promoted in 2004 by the Brazilian Ministry of Agriculture, Agribusiness, and Supply, seeks to complement the Brazilian Information System for Agro-Business (SIGA Brazil) with more precise prediction data. The program uses Landsat and Ciber-2 imagery to quantify the cultivated area. The system estimates yields from a combination of area and meteorological information captured by the National Oceanic and Atmospheric Administration (NOAA) satellites. One key objective is to distribute farming information to producers so they can make better decisions about what and when to farm.

All these programs developed a GIS with additional functionalities that help to analyze the information available, and to generate reasonably accurate crop yield estimation. With international help and political commitment, the programs overcame the cost of satellite imagery, the unpredictability of weather, and the extent of cloud coverage.

2.3.3. Colombian GIS Crop Prediction Systems

The United Nations Office of Drugs and Crime's Global Illicit Crop Monitoring Program is the only operational GIS crop yield prediction system in Colombia (United Nations, 2006). Since 1999, they have produced annual reports that estimate the coca cultivation areas and production. The objectives of the program are to define the methodologies for data collection and to monitor and communicate the extent of illicit crop in seven countries, including Colombia. In collaboration with the Colombian government during 2003, they established SIMCI II, an expansion program. SIMCI II provides GIS tools to monitor problems associated with illicit crops: deforestation, environment conservation, natural parks control, and indigenous territories protection (United Nations, 2006).

Classification of remote sensing imagery from Landsat, Aster, and Spot supports the methodologies established by the program to accomplish these goals. The Institute of Natural Resources and Applied Life Sciences of the University of Natural Resources and Applied Life Sciences in Vienna, Austria, concluded, after a technical evaluation, that the methodology is appropriate (United Nations, 2006). The methodology includes eight steps: identification and acquisition of satellite images, *spatial information database development*, image preprocessing, *digital land cover classification of land use and vegetation*, visual interpretation of coca fields (based on user experience), verification flights, *accuracy assessment*, and corrections (spraying, cloud cover and gaps, difference in acquisition dates).

The program developed an additional methodology to improve the accuracy of the estimation between areas, yields, and production. A multistage stratified area frame probability sampling divided the country into seven regions. The system automatically chooses primary sampling units based on a probability derived from the extent of coca cultivation areas in 2003 (United Nations, 2006).

2.4. Accuracy

In general, a correlation is the “identification of specific environmental conditions that are statistically significant precursors of a particular type of event” (Sarewitz et al., 2000). Correlated predictors are an important part of any prediction. However, correlations based on the past do not guarantee that any given model will represent the future because data contain errors, and natural systems are a constantly changing process.

The choice of relying on predictions as part of a decision-making process consumes resources and involves human oversight. That choice also requires accepting and measuring the irreducible prediction’s uncertainty. Accuracy is the best criterion to validate and justify predictions. It is defined as “the correlation between that which is predicted and that which actually occurs” (Sarewitz et al., 2000). Lack of accuracy will generate two kinds of errors: false alarms or surprises. Normally, any step taken to reduce one kind of error will increase the other (Stewart, 2000) but in any case, it is possible to measure it. Error measurement in parametric prediction commonly relies on mathematical operations; for example, the sum of squared errors can be used to measure a model’s accuracy or prediction closeness to reality. Sarewitz et al. (2000) state that a simple comparison of the prediction against the result is not enough to accept the prediction. Consequently, it is necessary to compare parametric results from other simpler or non-parametric predictions.

Gommes (2003) presented an extended list of errors affecting decisions based on parametric predictions. The list showed common errors in observation, processing (analysis, transmission and transcription), data bias, scale, parameter selection, simulation, lack of relevant factors, trend assumptions, decision-making assumptions, and conflicts between different forecasts. He proposed validating the use of non-parametric predictions, especially in under-developed countries.

3. Data

Data acquisition was a key task of the project. Similar projects, referenced in Chapter 2, pointed out data acquisition as a critical phase, as well as the importance of data quality. Data properties, such as precision, scale, or source trustworthiness relate to quality in GIS terms. One of the first risks identified for this project was data availability and assessing its associated quality. Although Procalculo Prosis and the Colombian Ministry of Agriculture supplied data for this project, on many occasions it was often necessary to validate, replace, or process the data. Those additional steps assured an acceptable degree of accuracy in the application.

This chapter presents the four classes of data used: remotely sensed imagery; crop yields statistics for Colombia and Cordoba; weather data incorporated to predict the future corn yields; and ancillary data from the area of study.

3.1. Remotely Sensed Imagery

Imagery was an important source of data for this project and replaced costly observation. Frequent cloud coverage over the Colombian territory increases the complexity of capturing, acquiring, and selecting appropriate imagery (United Nations, 2006). Only a small percentage of the commercially available images covering Cordoba had less than 50% cloud coverage. Additionally, during rainy seasons the chance to acquire useful images is severely limited.

Czajkowski & Grunwald (2000) considered that understanding the satellite factors is important before any classification effort. The factors included when selecting the appropriate satellite platforms are: fine spatial resolution (less than 30 meters pixel for classification), regional swath areas, monthly temporal resolution (15 to 30 days re-visit times), and wide multi-spectral resolutions. The selected satellite's sensors were: the Moderate Resolution Imaging Spectroradiometer (MODIS) in the Terra satellite, the Enhanced Thematic Mapper Plus (ETM+) in Landsat 7, and the High Resolution Geometric (HRG) in Spot 5 (see Figure 3-1). This subsection introduces the satellite platforms selected and the imagery datasets used in the project.

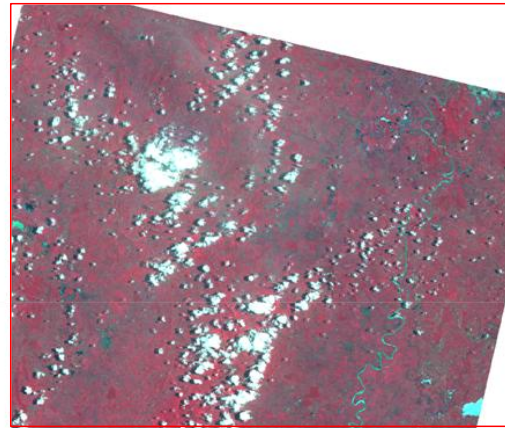
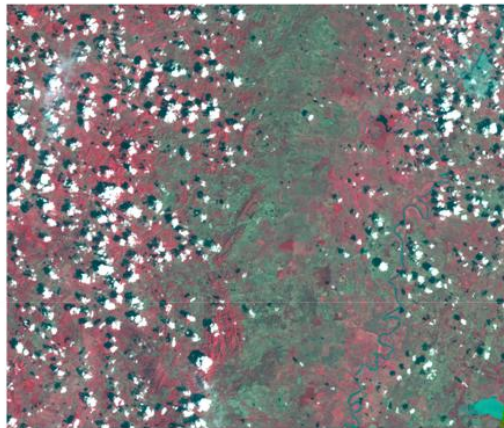
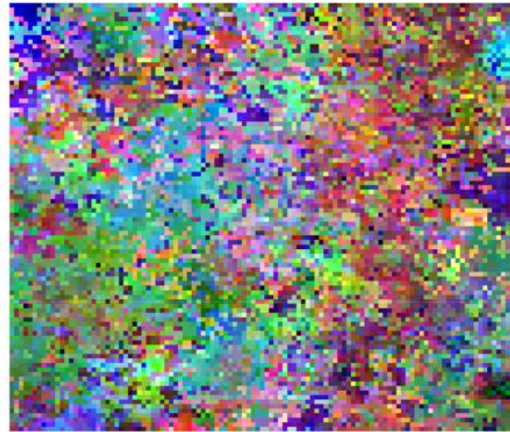


Figure 3-1 - Terra, Landsat, and Spot images over the study area

Although Procalculo Prosis, distributor of Landsat, Spot, Ikonos, Digital Globe, and Aster provided the imagery, the project required additional sources to increase temporal image availability, correct quality issues, and improve resolution. Temporal coverage is especially helpful to classify transitional crops such as corn and reduce the effect of cloudy conditions. The same images from different sources allowed comparisons to verify an image's correctness, and more detail helped the image's interpreter to mitigate the lack of field data.

3.1.1. Landsat

Landsat is a satellite program from the United States, which first launched Landsat 1 in 1972. Currently there are two operational Landsat satellites (Landsat 5 and 7). Both orbit the Earth at an altitude of 705 km in a sun-synchronous polar orbit of 16 days. ETM+ collects information in seven spectral bands (30m x 30m pixel) and one panchromatic band (15m x 15m pixel) covering a constant 185km x 185km area (Belgian Science

Policy - EO Edu, 2007). The Landsat program has an extensive database, both spatially and temporally. Different disciplines, including cartography, geology, geophysics, archeology, hydrology, agriculture, land use, environment, conservation, and oceanography have benefited from its multispectral images. Specifically, Bauer et al. (1979) and Badhwar (1984) used Landsat imagery to identify crops successfully.

Landsat imagery is the main source of historic imagery in the area of study. The path and row numbers of the images used were, respectively, between 9 and 10, and between 53 and 55 (see Figure 3-2A). Forty ortho-rectified Landsat images are available from 1985 to 2002. NASA (2007) provided twenty-one recent Landsat images through the University of Maryland Global Land Cover Facility. However, this imagery has degraded quality since May 2003, when the Scan Line Corrector failed. Gaps appear at the images' sides because of this "SLC-Off" malfunction in the Landsat 7 ETM+ instrument (University of Maryland, 2007b).



Figure 3-2 - Landsat (A) and Spot (B) scene grids over Cordoba, Colombia

3.1.2. Spot

Spot is the France-Belgium-Sweden partnership satellite program. Their satellite constellation currently has four operational satellites, allowing excellent daily coverage of almost the entire planet, and improved stereoscopic scenes. Launched in 2002, Spot 5 orbits sun-synchronously at an altitude of 832 kilometers with a temporal resolution of twenty-six days. Two High Resolution Geometric (HRG) sensors provide resolutions up to 2.5m x 2.5 m pixel (panchromatic) or 10m x 10 m (multispectral) in 60km swath images (Belgian Science Policy - EO Edu, 2007). Spot claims that because of the high number of sensors and bands, Spot's satellite imagery works in many applications including "mapping, cadastral surveying, security, defense, farming, forestry, fisheries, land management, planning, telecoms, geology, and risk mitigation" (Spot Image, 2007).

Figure 3-2B depicts the Spot scene grid. Some quadrants are not displayed because Spot identified them as having nearly 100% cloud coverage. Only two images with less than 50% cloud coverage were available. One image covers grid 641-331 and the other 640-

333; acquisition dates were December 18, 2006 and August 20, 2006 respectively. The remaining images indicated in Figure 3-2B have more than 50% cloud coverage.

3.1.3. Terra

Terra was launched on December 18, 1999 (NASA, 2006). The satellite has a built-in MODIS instrument, which captures 12-bit images in 36 spectral bands. The first two bands have a nominal resolution of 250 m, with the next 5 bands at 500 m, and the remaining 29 bands at 1 km. Terra orbits the Earth at 705 km, achieving a 2,330-km swath. Together with Aqua, its twin, they provide global coverage on a daily basis.

As observed in Subsection 2.2.1, the NDVI is a good indicator of vegetation health, one of the proposed predictor variables in the regression model. In theory, Landsat images had the required bands to generate the NDVI. However, in practice, they were extremely difficult to use because of the multiple differences in dates, the malfunction in the SLC, and the intense cloud coverage in the area of study. Different data were required. After evaluation, it was decided that the MODIS instrument in NASA's Terra satellite provides the necessary data to calculate NDVI values. NASA (2007) provided the MODIS product: 500m, 32-Day Global Composite images covering South America from years 2001 to 2005. The project required only red and near-infrared bands. This product uses multiple MODIS images from 32 days to remove cloud pixels. MODIS continental subsets are available in single band GeoTIFF files. Available bands are one to seven (red, NIR, green, blue, and three infrared bands), with an estimated average size per band of 800MB. This project required only red and near-infrared images of South America dated in the month of August.

3.1.4. Other Imagery Sources

Additional imagery sources that were considered included the Geoeye Foundation and NASA. Geoeye (2007) offered Ikonos imagery (1m panchromatic) and Orbview 3 imagery (1m panchromatic and 4m multispectral). The image search application required defining the AOS and the search parameters, such as cloud coverage less than 50%, acquisition dates since 2001, and angles from 0 to 90. Only Ikonos' panchromatic images were available and they were not useful for classification purposes. One image helped to identify land features; it was acquired directly by Procalculo Prosis from Geoeye.

NASA (2005) offered Aster imagery through its Terralook online application. Although the download process was simple and fast, the notes for the Colombian dataset showed that the Landsat images in the file were older than year 2001, and that the Aster files did not have the spectral resolution required to identify crops. The European Commission Joint Research Center (2007) offered a digital atlas, but it did not have any additional imagery.

3.1.5. Spectral Bands Selection

Two groups of researchers differed in opinion when recommending which bands to use. Turker & Arikan (2005) proposed to use all Landsat bands, including the panchromatic for additional detail, while Maxwell, Nuckols, Ward, & Hoffer (2004) used only Landsat bands 2 and 4. The project used all available bands in Landsat and Spot imagery and bands 1 and 2 in MODIS imagery. Every band detects different aspects on the Earth's

surface. Table 3-1 compares the bands' descriptions, number, and spectral ranges, among Landsat 7, Spot 5, and MODIS imagery.

Table 3-1 - Spectral band comparison among Landsat 7, Spot 5, and Terra

Detail/Uses	Satellite	Band	Spectral range (µm)
Blue-green light (scattered by atmosphere) Detects particles suspended in water Ground/ Plant differentiation	Landsat 7 Spot 5 Terra	1 3	0,45-0,515 0,45- 0,47
Green light Vegetation	Landsat 7 Spot 5 Terra	2 1 4	0,525-0,605 0,50-0,59 0,54-0,56
Red Light Plant species differentiation	Landsat 7 Spot 5 Terra	3 2 1	0,63-0,69 0,62-0,67 0,61-0,68
Near infrared Detects biomass Radiation from leafy vegetation in which chlorophyll reflects much of the NIR	Landsat 7 Spot 5 Terra	4 3 2	0,75-0,90 0,79-0,89 0,84-0,87
Middle Infrared Variations in moisture content (Plant/soil) Snow/Cloud differentiation	Landsat 7 Spot 5 Terra	5 SWIR 6	1,55-1,75 1,58-1,75 1,62-1,65
Thermal Measures temperature Soil/Plant/Mineral research	Landsat 7 Spot 5 Terra	6 5	10,4-12,5 1,23-1,25
Middle Infrared Lithology. Moisture in leafy vegetation Crop productivity and vegetation areas under stress	Landsat 7 Spot 5 Terra	7 7	2,09-2,35 2,1-2,15
Panchromatic Smaller objects identification Accurate land maps Increased band differentiation	Landsat 7 Spot 5 Terra	PAN PAN	0,50-0,90 0,51-0,73

Data sources: Belgian Science Policy - EOedu, 2007; Towson University, 2007; University of Maryland, 2007

Multispectral images offer better accuracy for crop identification (Turker & Arikan, 2005). Because a single multi-spectral image cannot differentiate between some crops, it might be necessary to include a multi-temporal image approach to identify crop types, as those can detect seasonal variation characteristics of each crop. A multi-temporal image consists of the different date images and bands stacked into one single image with more than 30 bands. In order to keep size and processing efforts manageable, it is necessary to eliminate bands when using multi-temporal images. The use of multi-temporal images requires a principal components analysis to identify which bands to use (Czajkowski & Grunwald, 2000). However, it was not the case with this prototype, given the lack of imagery within the same cropping cycle.

3.2. Crop Yield Statistics

The Colombian Ministry of Agriculture provided crop production data for the country. Agronet is the ministry portal for statistics, analysis, and a national agricultural survey.

(Ministerio de Agricultura, 2007). It provides electronic access to yield estimates by state, product, and year. Estimates of traditionally cropped corn yields and cultivated areas in Cordoba were available for the years from 1987 to 2005 (see Figure 3-3). Figure 3-4 shows a comparison between the two cropping methods. The dataset contains planted hectares, tons harvested, and yield per hectare. For comparison, in the United States the same information is available since 1866.

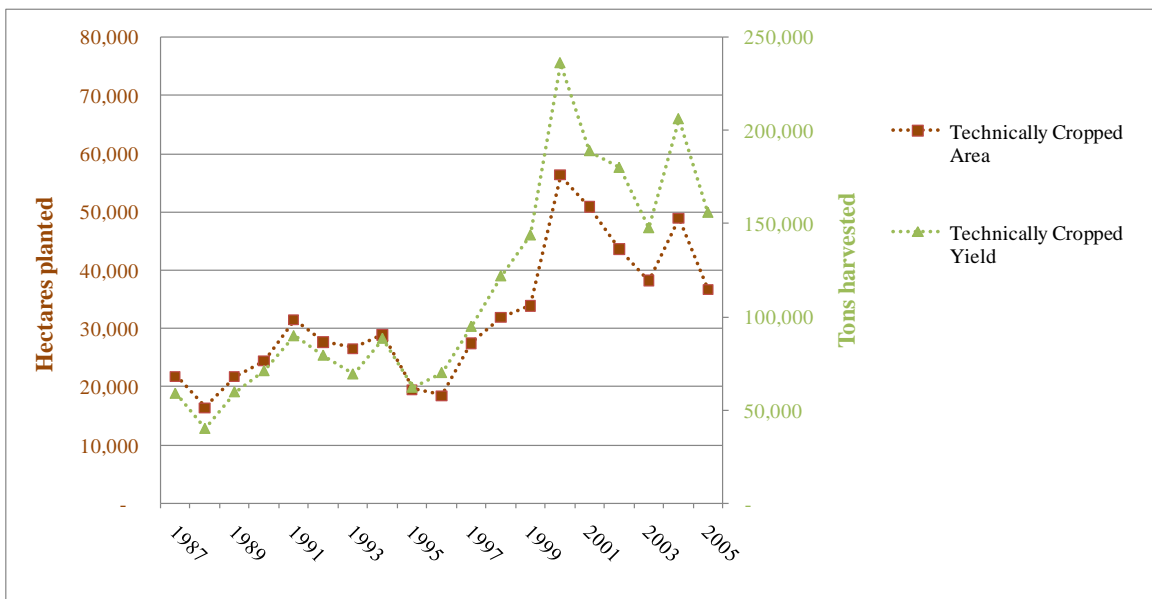
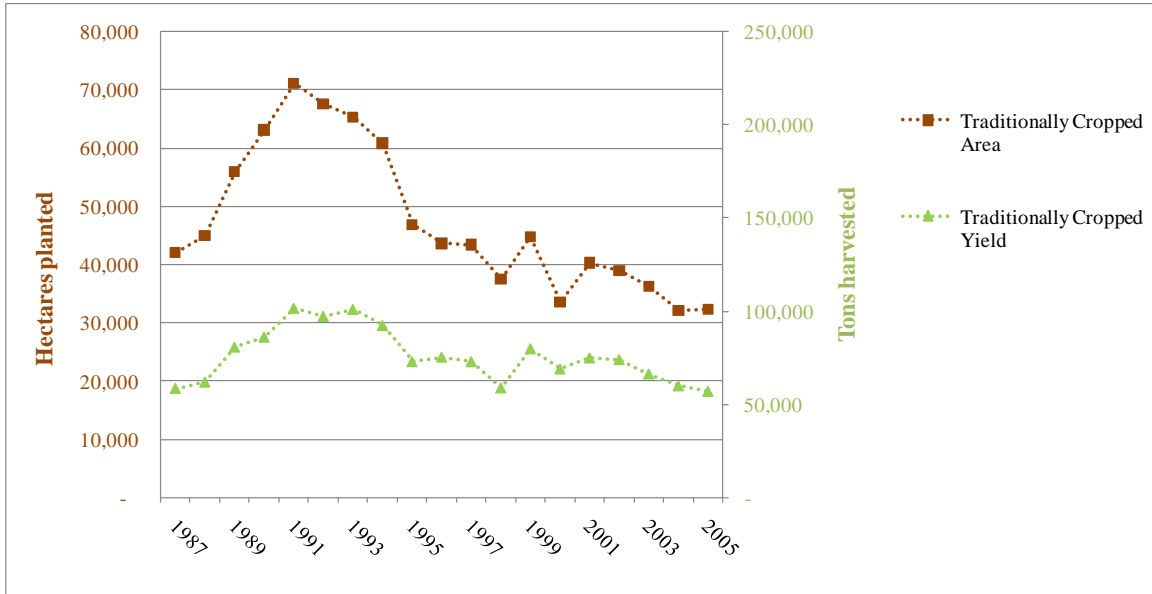


Figure 3-3 - Corn production in Cordoba (1987 -2005) Traditionally cropped (A) Technically cropped (B)

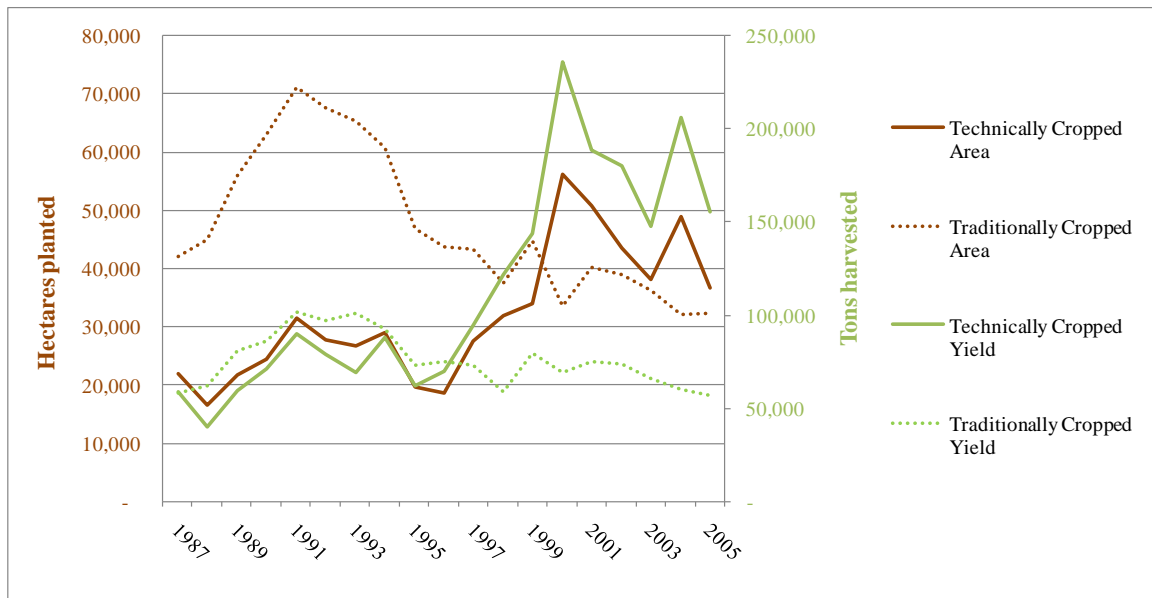


Figure 3-4 - Comparison between traditionally and technically cropped corn in Cordoba (1987-2005)

3.3. Weather Data

The Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM for the Spanish acronym of Colombian National Institute for Hydrology, Meteorology, and Environmental Sciences) offers a paid service which provides a historic dataset by weather station and weather variable (IDEAM, 2007b). Two timeframes were available at different costs: USD\$10 for daily readings per year, and USD\$15 for all the monthly average readings since 1988. Although daily information might improve accuracy, it was unnecessary because NDVI and production data were available only on a monthly basis. Inside the AOS, 241 weather stations were available (IDEAM, 2007a). To reduce the costs, Procalculo Prosis provided data for twenty-five stations, which extends beyond Cordoba boundaries to reduce edge effects.

The variables acquired from IDEAM were precipitation and temperature. The rainfall monthly average readings' units were millimeters, and temperature readings were in Celsius degrees. Each dataset included a year-by-month text matrices and the respective station's description (headings in Figure 3-5). In addition, the matrix contained metadata about each reading (Columns marked with * in Figure 3-5).

IDEAM - INSTITUTO DE HIDROLOGIA, METEOROLOGIA Y ESTUDIOS AMBIENTALES													SISTEMA DE INFORMACION NACIONAL AMBIENTAL				
VALORES TOTALES MENSUALES DE PRECIPITACION (mms)																	
FECHA DE PROCESO : 2007/09/04			TIPO EST			CO			DEPTO			CORDOBA			ESTACION : 1308503 SAN BERNARDO DEL VIENTO		
LATITUD 022 N			ENTIDAD 01 IDEAM			MUNICIPIO SAN BERNARDO DEL VIENTO			CORDOBA			FECHA-INSTALACION 1964-NOV					
LONGITUD 7558 W			REGIONAL 02 ATLANTICO			CORRIENTE CNO CICARA						FECHA-SUSPENSIÓN					
ELEVACION 0022 m. s. n. m																	
AÑO	EST	ENT	ENERO	FEBRE	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOST	SEPTI	OCTUB	NOVIE	DICIE	VR ANUAL		
1988	2	01	.0	.1	.0	104.4	136.4	198.9	258.2	193.9	205.6	186.7	77.4	8.7	1370.3		
1989	2	01	3.4	.3	35.7	.0	25.4	153.9	60.5	180.8	256.8	198.6	118.8	20.0	1054.2		
1990	2	01	29.8	3.8	.0	88.4	207.0	180.8	149.8	202.3	317.7	167.2	101.0	85.1	1532.9		
1991	2	01	.0	.0	1.7	46.9	235.0	169.2	114.3	206.6	248.0	110.0	97.3	6.5	1235.5		
1992	2	01	5.0	.0	.0	87.8	75.1	220.6	229.3	231.0	178.9	153.3	125.0	12.0	1318.0		
1993	2	01	32.6	.0	26.7	40.0	169.0	232.0	123.3	127.3	267.8	140.0	199.5	34.0	1392.2		
1994	2	01	2.0	.0	22.0	105.0	258.0	65.7	189.3	196.0	158.0	206.5	209.8	.0	1412.3		
1995	2	01	.0	.0	.0	50.0	220.8	123.1	173.4	139.3	145.1	254.1	106.9	144.5	1357.2		
1996	2	01	24.0	.4	82.0	63.0	185.0	109.0	272.0	219.0	398.0	93.0	186.0	15.0	1646.4		
1997	2	01	.2	.0	.0	85.5	81.4	178.5	176.9	200.6	210.2	183.7	19.5	10.0	1146.5		
1998	2	01	.0	.0	1.0	14.2	141.8	73.8	133.4	309.5	71.9	124.6	153.0	196.1	1219.3		
1999	2	01	6.8	22.2	14.1	19.4	119.8	214.1	327.1	270.7	324.0	126.0	170.1	51.8	1666.1		
2000	2	01	.4	.0	1.3	30.0	154.5	114.8	83.5	116.6	257.3	37.3	50.3		846.0		
2001	2	01	1.7	.0	5.0	2.1	68.6	61.1	142.0	308.7	331.7	211.3	263.4	4.9	1400.5		
2002	2	01	.0	.0	14.5	41.6	44.2	96.2	69.0	80.3	130.3	193.9	25.4	50.0	745.4		
2003	2	01	.0	2.0	.0	57.4	81.1	147.2	143.0	66.6	205.5	77.8			780.6		
2004	1	01	.0	.0	.0	9.3	233.8					276.2	210.1	32.2	761.6		
2005	1	01	.4	.0	13.5	57.6	144.9	206.6	196.1	235.0	321.9	235.3	379.9	43.2	1834.4		
2006	1	01	1.2	1.0											2.2		
MEDIOS			6.0	1.7	12.1	50.1	143.4	149.7	167.1	193.2	237.0	165.3	146.7	44.6	1316.9		
MAXIMOS			32.6	22.2	82.0	105.0	258.0	232.0	327.1	309.5	398.0	276.2	379.9	196.1	398.0		
MINIMOS			0.0	0.0	0.0	0.0	25.4	61.1	60.5	66.6	71.9	37.3	19.5	0.0	0.0		

Figure 3-5 - Original weather reading tables by station (IDEAM, 2007b)

Excel pivot tables organized the weather information into a table of stations by month-year (see Table 3-2). In addition, the process included saving the metadata, in a different Excel file for its future use in accuracy assessment.

Table 3-2 - Processed weather stations' readings

StationID	Name	Latitude	Longitude	Altitude (m)	Type	State	Date Installed	Rainfall (mm) 1988-01	Rainfall (mm) 1988-02	Rainfall (mm) 1988-03
1308503	San Bernardo del Viento	9.3667	-75.9500	22	CO	Cordoba	Nov-64	0	0.1	0
1111502	Canas Gordas	6.7333	-76.0333	1200	CO	Antioquia	Jul-73	23.1	67	69.4
1308502	Lorica	9.2667	-75.8167	30	CO	Cordoba	Jun-64	0	0	0
1307501	Chima	9.1500	-75.6167	20	CO	Cordoba	Oct-73	7.9	4.5	31.2
1204502	Cristo Rey	9.0667	-76.2333	15	CO	Cordoba	Oct-73	0	4.9	0
1307503	Turipana	8.8500	-75.8167	20	AM	Cordoba	May-60	0	1.5	0
1308504	Los Garzones	8.8167	-75.8500	20	SS	Cordoba	Jun-74	6.5	16.9	29.1
1307505	Univ Cordoba	8.8000	-75.8667	15	AM	Cordoba	Jun-79	0	19.3	6.3
2502517	Colomboy	8.7500	-75.5000	125	CO	Cordoba	Oct-73	0	1.2	52
1305503	Galan	8.6667	-75.9833	30	CO	Cordoba	Jun-79	0	8.3	0.2
1306502	Maracayo	8.4167	-75.8833	25	CO	Cordoba	May-79	0	15.4	0
2502515	Ayapel	8.3167	-75.1333	22	CO	Cordoba	Aug-67	0	7.8	0
2501501	Centro Alegre	8.1667	-75.6667	170	CO	Cordoba	Dec-73	0	6.8	93.8
1301504	Represa Urra	7.8833	-76.2500	300	CP	Cordoba	Nov-04			
2502531	Puerta roja	9.3300	-75.3833	160	ME	Sucre	Dec-84			
2502527	Univ Sucre	9.2000	-75.3833	160	AM	Sucre	Sep-83			
2502518	San Benito Abad	8.9333	-75.0333	20	CO	Sucre	Nov-73	0	5	0
2502524	Majagual	8.5333	-74.6167	20	CP	Sucre	Nov-74	0	9.2	0
1308501	Doctrina la	9.3000	-75.9000	20	CP	Cordoba	Jan-68			
2502516	Hacienda Cuba	8.0000	-75.4167	50	CO	Cordoba	Apr-73	2.6	5.8	0
1204501	Arboletes	8.8500	-76.4167	4	CP	Antioquia	Mar-72	0	4.7	14.5
1201509	La Palmera	7.5000	-76.6000	58	CO	Antioquia	Mar-91			
1201502	Uniban	7.8333	-76.6667	23	AM	Antioquia	Aug-77	9.1	77.1	1.5
2625504	Cacacoteras	7.9833	-75.1167	55	CP	Antioquia	Aug-68	1.2	21.8	0.9

Initially a weather readings table was joined to the weather stations point features. However, final organization changed because the “Add Join” tool did not work inside the ArcGIS Model Builder. The final weather station dataset integrated all rainfall and temperature readings by month in a single table.

3.4. Ancillary Data

Procalculo Prosis obtained additional features from the Colombian Geographic Entity (IGAC). ESRI Coverage Interchange files (e00) stored the features including: main roads;

roads; non-paved main roads; dirt roads; hydrography; towns; state capitals; and municipal divisions; individual raster files store the Digital Elevation Model (DEM) and soil use raster. These datasets served two purposes: they provided the end-user with known base map features, and their use improved accuracy by helping with crop identification (Kontoes, Wilkinson, Burrill, Goffredo, & Mégier, 1993).

4. Methodology

This chapter describes the processes required to transform the data. The project had three tasks. The first task was building a spatial information database. The second task was compiling, classifying, and interpolating the imagery. The last task was formulating and running a multiple regression analysis.

4.1. Spatial Information Database

Czajkowski & Grunwald (2000) considered assembling GIS data layers as the first task in their project. Kontoes et al. (1993) encountered problems with the time-consuming processes needed to integrate data from different sources. Fortunately, the modern versions of ESRI ArcGIS address the integration problem using geodatabases. In addition, the geodatabase format improves data processes such as storage, retrieval, display, and update (ESRI, 2007a, 2007b). Design was the first step to build the file geodatabase, which contains the data and models for this prototype. Editing the metadata was the next step, followed by defining an area of interest and populating the geodatabase.

4.1.1. Database Design

The project prototype required only a few data layers, helping to keep the design simple. The Mip.gdb, the project ArcGIS file geodatabase, stores raster images, vector data, and prediction tools. The Mipscratch.gdb, an ArcGIS file geodatabase, keeps the results of all the processes and analysis used during the project, without affecting the Mip.gdb design or the original data.

Imagery is at the core of the database and is stored as raster datasets: Landsat, Spot, DEM and SoilUse. A geodatabase Toolbox stores the models to process the data. Some entities, relationships, and fields in the ArcGIS model for agriculture (ESRI, 2003) served as a template for organizing ancillary data (see Figure 4-1).

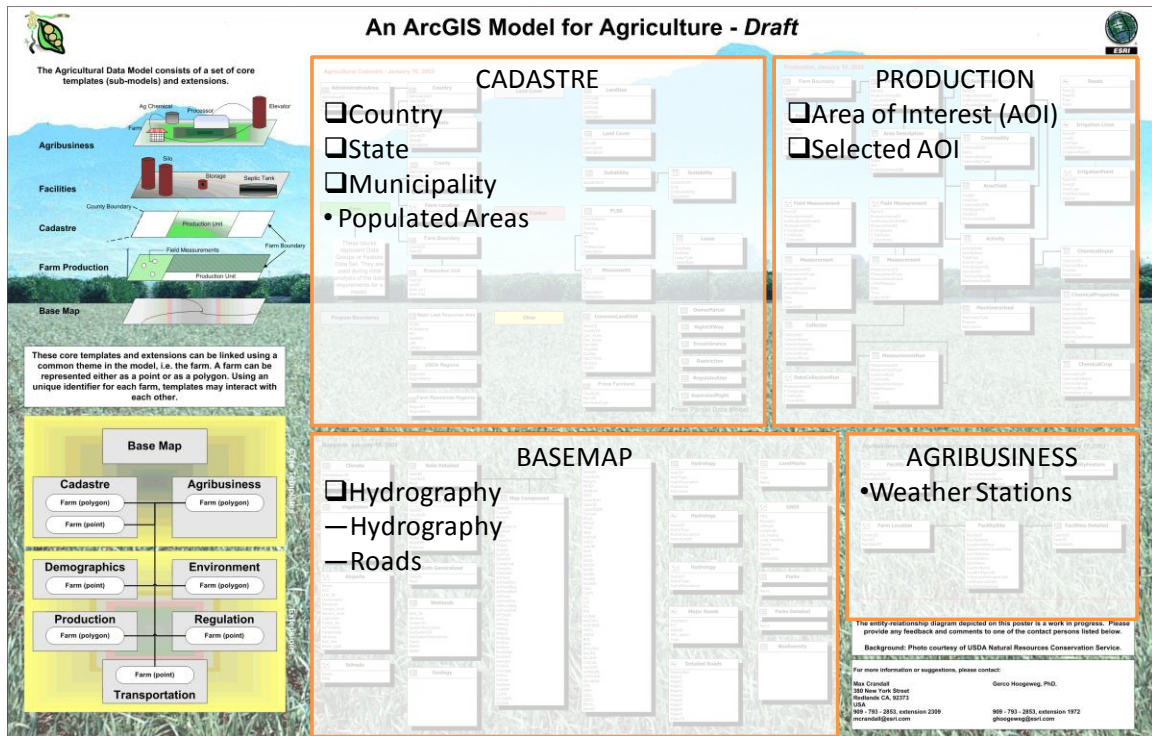


Figure 4-1 - An ArcGIS Model for Agriculture – Draft (Adapted from ESRI, 2003)

The main feature datasets, derived from the template, are Cadastre, Production, and Base Map. The Cadastre feature dataset stores the administrative divisions: country, state, municipality, and populated areas. The Production feature dataset stores the Production Area and Yield Area polygons. The Base Map feature dataset stores hydrography and roads features classes. Mip.gdb stores the Commodity and Area Yield tables, as well as the Weather Stations feature class, which contains the rainfall and temperature information. The roads and populated areas feature classes use subtypes for organization and display purposes.

4.1.2. Metadata Edition

ArcCatalog provided the tools for editing, loading, or copying the metadata. FGDC ESRI style sheet is the format used for presenting and updating the metadata. Every layer loaded into the Mip.gdb database has its respective metadata. For example, raster data have an accompanying xml metadata file and other data sources provided the additional metadata information on their webpage.

4.1.3. Spatial Frame Definition

Achieving a small geodatabase size and increased performance required reducing the dataset's spatial frame from the countrywide size to area of study size. Section 1.5.3 describes the process of selecting Cordoba as the AOS. A 50km buffer around the Cordoba boundary was used to eliminate the spatial inconsistencies that could appear near the AOS boundaries.

The Subset function in Erdas clipped all of the acquired images using the minimum bounding rectangle around the AOS. The Stack function not only stacked Landsat and

Terra imagery layers, but also had the option to subset the image at the same time (using AOS.aoi file into the AOI menu), reducing processing time; however, one more subsetting was required to delete the background. Executing a stacking process for Landsat SLR-off imagery created unusable images if the option to include the zero value in the stacking process was on.

The use of a smaller Area of Interest (AOI) avoids problems with multiple image spanning. Moreover, an AOI allows clipping other datasets and reducing their space requirements. Additional benefits of using the AOI are not only the improvement in the prototype stability and overall performance, but also that it allows the user to define or select different areas for the prediction.

Czajkowski & Grunwald (2000) recommended, as part of a preprocessing effort, to classify the clouds and remove images with more than 30% cloud coverage, in order to assure that images are geographically and atmospherically corrected, and that images have a small spatial error. Following this recommendation eliminated 47 Landsat images. Consequently, it was not possible to follow their last recommendation of including images from the complete growing season.

Although the prototype runs based on any polygon inside the AOS, it includes two predefined AOIs near the Cordoba capital, Monteria. Among the 40 Landsat images, the one in row 10 path 54 taken on February 2, 2001 (Test Image One) is the most recent and the least cloud-covered Landsat image (See Figure 4-2).



Figure 4-2 - Test Image One, Landsat

A multitemporal image was ideal for the classification process, but the cloud coverage of the available images from August 2000 to August 2001 is 70% or more (NASA, 2007). Figure 4-3 shows a Spot Image taken in grid 640-333 on August 20, 2006 (Test Image Two). The best time for prediction is before the main harvesting season, in August and September. Therefore, Test Image Two was the appropriate image to use; it was recent, had more resolution than a Landsat 7 image, and it overlapped with Test Image One. In addition, both images selected did not need additional corrections because they are geometrically and atmospherically corrected.

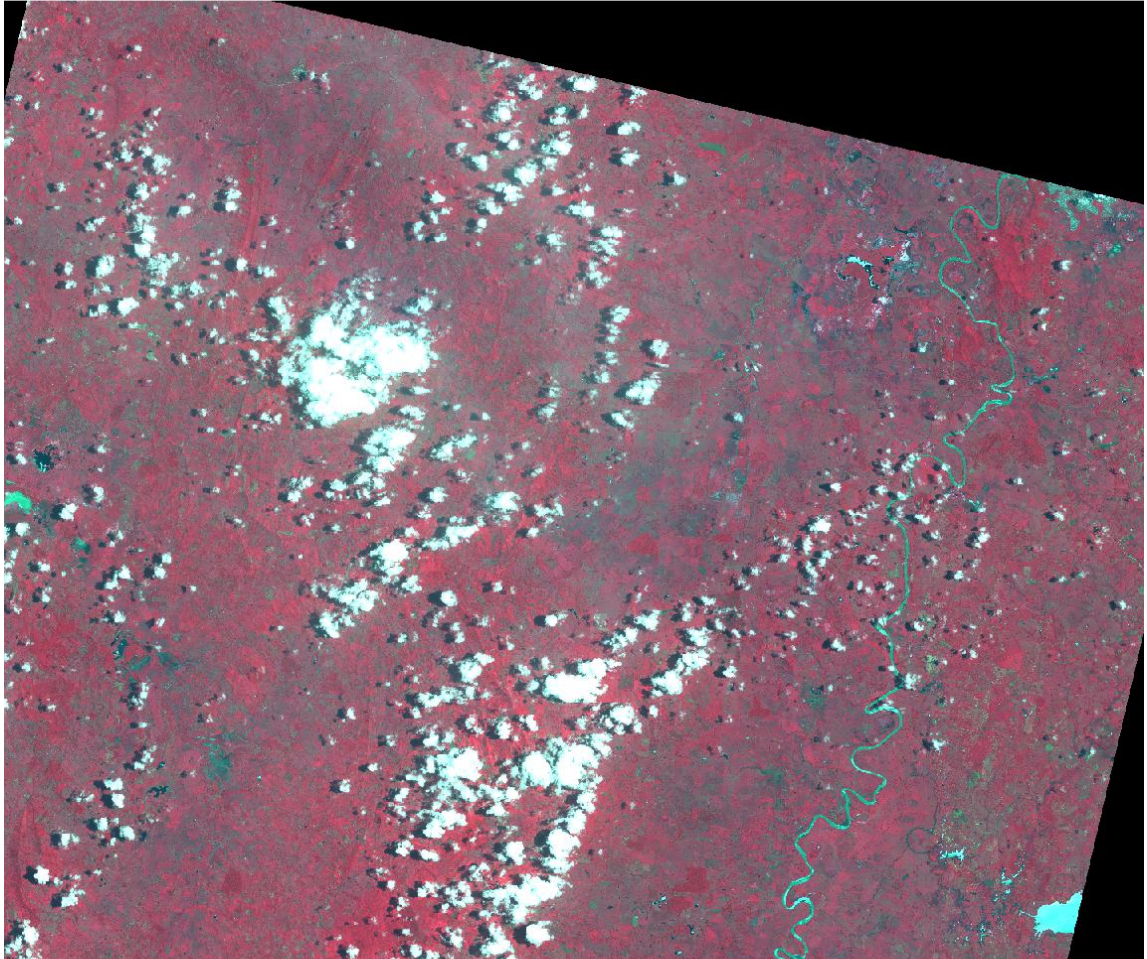


Figure 4-3 - Test Image Two, Spot

The Production Area feature class stores the AOI polygons, which match the extent of image to analyze. The AOI feature class contains two predefined polygons (AOIs). The fields of the AOI feature class are YieldID, Commodity, Yield Quantity and Yield Unit. The user can create new AOI polygons using ArcMap's editing functions.

4.1.4. Raster Data Population

Imagery.gdb was a temporal file geodatabase. It contained unprocessed original raster files and prevented the corruption of large files. Corruption occurred frequently at the beginning of the project when working with rasters of countrywide size. The first option proposed to organize the imagery was to use raster catalogs to store imagery by provider, image name, grid code, and date, and grid tables to identify the images. Although this solution could have helped to select the images by date and area thus improving performance, it was unnecessary for this prototype because there were not enough quality images in different path/row at the same month/year that could be mosaiced and clipped into grids. Landsat and MODIS images were available in files by layer, requiring a stacking process before or during population of Imagery.gdb.

The DEM was provided in an unfamiliar MapInfo GRID format; with help from Nathan Strout it was converted into a usable file geodatabase format. Consequently, initial

metadata files were not readable by ArcGIS. Although the DEM appeared to be in the UTM 18N reference system, the raster displayed was incorrectly referenced. After reprojecting using the soil use layer's reference system, the result remained unsatisfactory. Another alternative was to georeference the raster, aligning it with Colombia's boundary, but the registration accuracy was degraded significantly. The fact that ArcGIS displayed the map units in meters when they should be in degrees hinted at the solution: redefining the DEM layer's reference system to the geographic reference GCS_WGS_1984 from the projected UTM 18N system. It worked. The DEM then matched the country topography and other base map features such as rivers and lakes.

The Soil Use dataset format was GeoTIFF. The raster had 13 values described in Spanish (see Table 4-1). Once imported into the database, the raster had 8-bit pixel depth or 255 values. The value description was translated and typed as a new field in the raster file called Use Description. Although a color map could be especially useful for standardizing the display of this dataset, the functionalities to create or modify a color map were not available in ArcGIS. The alternative was using a Unique Value displaying method.

Table 4-1 - Soil use raster values and description

Value	Description in Spanish	Description in English
1	Mar	Sea
2	Vias	Roads
3	Agua	Water bodies
4	Urbano	Urban
5	Glaciar	Glacial
6	Paramo	Moor
7	Cultivos	Croplands
8	Bosque	Forest
9	Rastrojo	Stubble
10	Pastos	Pastures
11	ND	ND (Not defined)
12	Escasa vegetacion	Scarce vegetation
13	Abierto	Rock/Bare soil
Other values		ND (Not defined)

ArcCatalog functionalities assisted in the completion of the tasks. The New function allowed the creation of raster datasets, while the Load Data function incorporated an original raster file into the dataset. Load Data must have the Ignore Background Value filled to prevent black areas from displaying in the resulting images. Because the prototype geodatabase requires smaller files focused on the area of study, the Clip tool (Data Management) in ArcToolbox created two subsets from DEM and Soil Use datasets, using the AOS extent. The same function allowed the creation of two AOI rasters from the most cloud- free areas of Test Image One and Two. Mip.gdb stores the resulting smaller rasters.

4.1.5. Vector Data Population

As discussed in Section 3.4, the vector data used in this project was acquired in ESRI coverage format. ESRI (2006) ArcGIS Help files provided the instructions for

configuring ArcCatalog to view and export Interchange files (e00) into a geodatabase. After conversion, Vectors.gdb was the file geodatabase, which stored the converted feature classes, but not the annotation layers. The Clip tool (Analysis) created geographic subsets of all these features using a 50km buffer around Cordoba's boundary. Roads feature is the result of appending Main Roads, Unpaved Main Roads, Roads, and Dirt Roads feature classes, while Populated Areas joins Towns and State Capital feature classes. The Dissolve tool, using the Create Multiparts parameter unchecked, created the Countries and States feature classes with the Municipalities feature class serving as the input. After their creation, both Roads and Populated Areas required the creation of subtypes for better organization and display. The process had three steps: creating a new short integer Class field in each feature class; calculating the values from the Simbolo field; and, creating the subtypes in the properties menu. Table 4-2 details the created subtypes. Copying to the Mip.gdb geodatabase and translating the feature classes' names from Spanish were the final steps.

Table 4-2 - Subtypes for Roads and Populated Areas feature classes

Roads		Populated Areas	
0	Undefined	0	Undefined
1	Dirt Road	1	State Capital
2	Road	2	Town
3	Main Road		
4	Unpaved Main Road		

4.1.6. Weather Data Population

The objective of this process was to incorporate weather station location data and weather readings, described in Section 3.3, into the geodatabase. Initially, the weather variable (temperature or rainfall) was a field in the table, identifying the variable each station captures, but the table was organized by station including rainfall and temperature readings in independent columns by month and year. The Export to Feature Class tool required adding an initial letter to each reading field's label because if the field starts with numbers, the field name is changed. The letter R denotes rainfall readings by month/year, while the letter T means that the field is a temperature reading; for example, T200704 is the temperature reading from April 2007. The Month part is always indicated with two digits.

The exporting process repeatedly failed for several reasons: Excel 2007 files were not compatible with ArcGIS 9.2; changes in the Excel table did not update on ArcGIS consistently; and the weather station table had to have the same order as the weather readings table for linking. However, once the table was in the right organization, it transformed easily into a point feature class using the Add XY Data function in ArcGIS. This function displayed the points seamlessly inside the AOS, but the location provided was accurate only to the nearest minute. IDEAM provided improved accuracy in its webpage station's viewer (IDEAM, 2007a). A simple location comparison between the webpage and the file revealed that coordinates in the file were incorrect. After improving the precision of the coordinates, the station locations' accuracies improved as shown in Figure 4-4. With the more accurate locations, the Export Data to Geodatabase (single)

tool exported the geographically displayed station data into Weather Stations as a new feature class inside the Mip.gdb geodatabase.

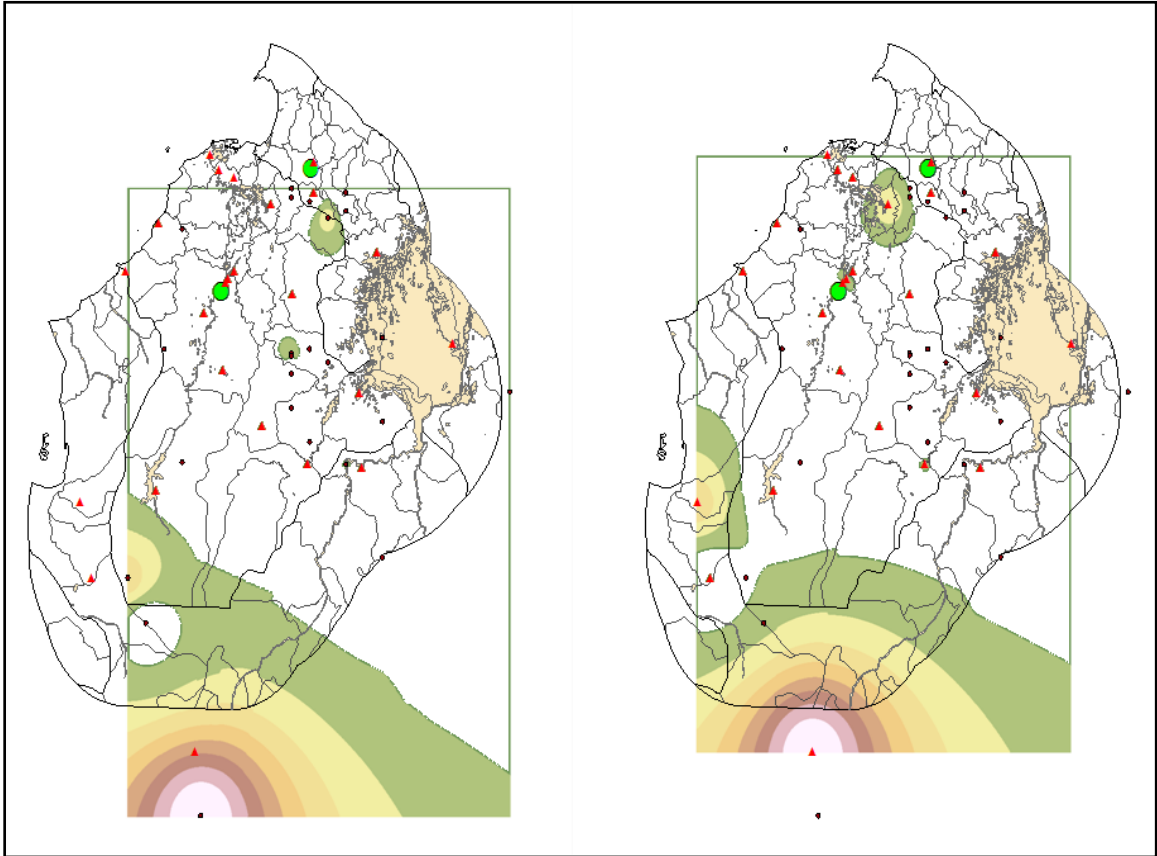


Figure 4-4 - January 1988 temperature interpolations using original locations in black (A) and corrected locations in red (B)

4.2. Imagery Processing

In this section, unsupervised classification and supervised classification, based on ancillary data, identified areas planted with corn. Arithmetic operations between the infrared and red bands obtained the NDVI. Interpolation between weather stations' readings then created temperature and rainfall surfaces.

4.2.1. Image Classification

Silapathong & Blasco (1992) considered remote sensing the most cost-effective method of gathering information about croplands. Image classification helps produce valuable information from remote sensed imagery. Three related applications are: crop identification, planted area measurement, and analysis of crop condition. For example, the United Nations (2006) did not use these classification methodologies to detect specific crop cultivated areas, but rather to implement a broad study of land cover.

For this purpose, United Nations (2006) developed a broad eight-step classification methodology. The first step is the identification and acquisition of satellite images; Section 3.1 explains the images acquired for this project. The second step is the creation of a spatial information database; subsections 4.1.4, 4.1.5 and 4.1.6 cover this topic. The

third step is image preprocessing, which includes Czajkowski & Grunwald's (2000) recommendations (see Subsection 4.1.3 and 4.2.1.1). Subsection 4.2.1.1 describes unsupervised classification. The next three steps are visual interpretation, verification flights, and accuracy assessment; a supervised classification replaced all of them (see Subsection 4.2.1.2). Section 6.1 addresses corrections as a possible final step.

4.2.1.1. Unsupervised Classification

In general, the two traditional methods used to classify imagery are unsupervised and supervised (Aronoff, 2005; Leica Geosystems, 2005a, 2005b; Towson University, 2007). In an unsupervised classification, the analyst defines only the number of classes expected, and the image processing software creates that number of classes based on the properties of the pixels. This method is reliable, repeatable, acceptably accurate, and is especially useful when no ground truth is available.

An initial task for unsupervised classification inside the project was to identify cloud covered areas on the images (see Subsection 4.1.3). Cloud coverage less than 30% was the threshold used to choose which images are suitable for the prototype. The Erdas Imagine unsupervised classification tool classified three classes using nine bands on Landsat images and bands one, two and three in the Spot images, iterating six times, with a convergence threshold of 0.95. The results showed class one as clouds, class two as surface, and class three as water and cloud shadows (see Figure 4-5). The image statistics indicated that class one comprised 22.90% of Test Image One and 8.4% of Test Image Two.

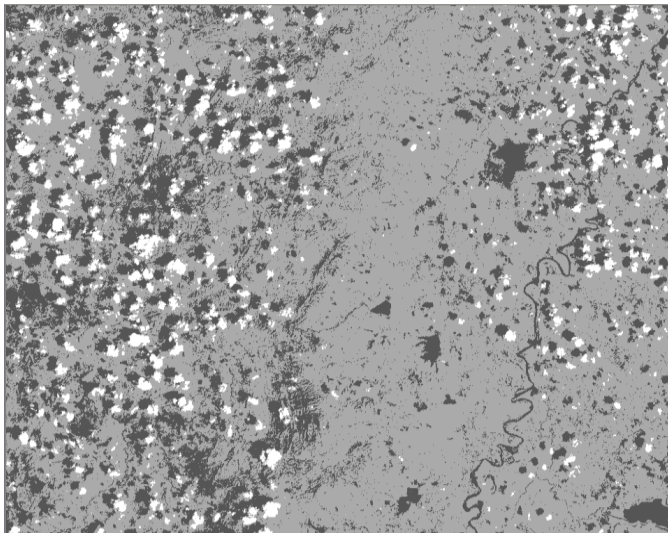


Figure 4-5 - Classification of Test Image One identifying clouds in the AOI

However, the main objective was deducing ground truth for the supervised classification, such as digital land cover classification, a vegetation index, or isolating areas used for agriculture from those not used. Increasing the number of classes allowed identification of land cover and testing for the appropriate number of classes. Only Test Image One and Test Image Two required this detailed unsupervised classification. Test Image One is a Landsat image with nine bands; while Test Image Two has three. At the beginning of the project, band one and band two in the Test Image One, provided by Procalculo Prosis,

were identical; the image had apparently been corrupted. Band one, downloaded from NASA (2007), replaced the original band one. After repeating the stacking, subsetting, and geodatabase loading process, this replacement had a small positive effect in the classification because the classes showed a better separability index.

Section 3.1.5 explains why all the bands were useful for classifying. Some classification tests in Test Image One, with different bands combinations led to similar results. Tests were with three bands (four, three and two); six bands (one thru five, and seven); seven bands (all except two thermal bands); and all nine bands. Different classifications of Test Image Two with 15, 20 and 25 classes searched for an appropriate level of distinctness and compactness. ESRI (2007c) recommended the use of the “Dendrogram” or “Class Probability” Tools to identify which classes should be merged or deleted to eliminate overlapping classes (see Figure 4-6). After this process, the final number of classes chosen was 16.

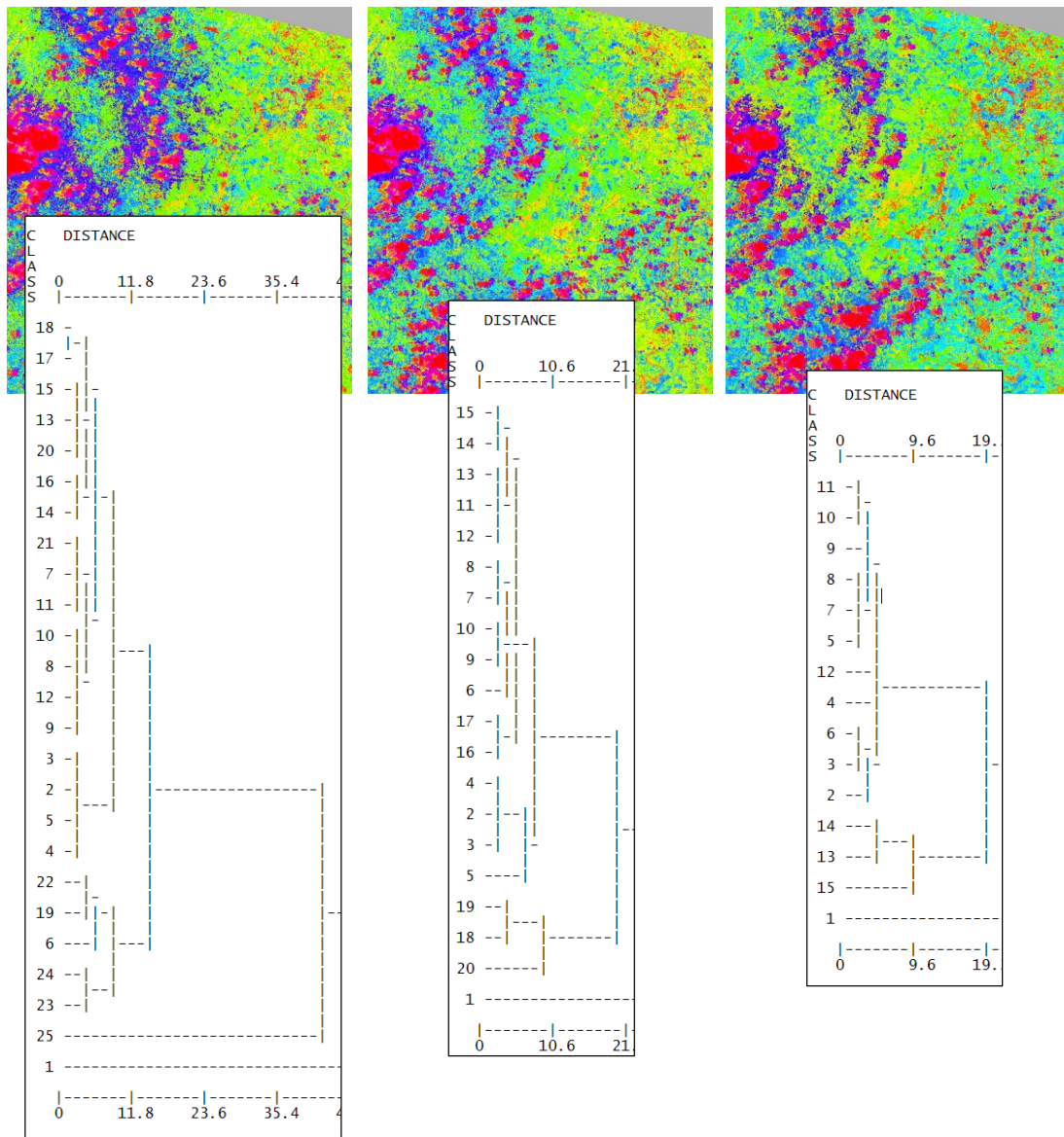


Figure 4-6 - Three different classifications of Test Image Two for 15, 20 and 25 classes

After the tests, the classification parameters for the Erdas Imagine unsupervised classification tool were: a maximum of 24 iterations, a Convergence Threshold of 0.995, and the Classify Zeros option unchecked. In ArcGIS, the ISO Cluster Tool, combined with the Maximum Likelihood Classification Tool, replaced the Unsupervised Classification Tool in Erdas, but it took three times longer. The ISO Cluster parameters were: 24 iterations, minimum class size of 2000, and sample interval of ten. The Maximum Likelihood Classification parameters were: reject fraction of 0.005, and EQUAL a priori probability weighting.

4.2.1.2. Supervised Classification

In a supervised classification, the analyst determines the classes and creates training areas. The image processing software then compares the spectral signature of the training areas against every other pixel in the image. A spectral signature refers to “a characteristic set of reflectances over the electromagnetic spectrum” (USGS, 2007). The methodology used is similar to that used by the Office for Drugs and Crime project that “performed a supervised classification, where training areas represent the features to be mapped in advance and class signatures are calculated. Each pixel is then assigned to a land cover class” using the maximum likelihood algorithm (United Nations, 2006). Czajkowski & Grunwald (2000) , Doraiswamy, (1997), Leica Geosystems (2005a, 2005b), and ESRI (2007c) provided classification technical procedures.

The maximum likelihood algorithm assumes that the distribution of a class sample is normal and assigns cells to a class based on the highest probability of being a member. Its wide use in similar applications and the proven functionality in both ESRI ArcGIS and Erdas Imagine are the two main reasons for using this algorithm. Although supervised classification is not a difficult process, it requires ground truth to be accurate. However, field and production data were not available for this project, as explained in Chapter 3. For this reason, this procedure required additional effort to produce signature files from ancillary data and analysis. Erdas Signature Editor was the tool used to create and analyze the spectral signature file for Test Image One (see Figure 4-7).

Class #	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1	> Shadow/water H		0.753	0.753	0.753	1	1	686849	1.000	X	X	X		
2	Shadow?water L		0.827	0.827	0.827	2	2	460282	1.000	X	X	X		
3	Primary forest and rainforest	Green	0.000	0.392	0.000	3	3	456530	1.000	X	X	X		
4	Shrubs	Yellow	0.961	0.961	0.863	4	4	619753	1.000	X	X	X		
5	Secondary forest	Light Green	0.000	1.000	0.000	5	5	768505	1.000	X	X	X		
6	Crop2	Blue	0.000	0.000	1.000	6	6	569716	1.000	X	X	X		
7	Crop3	Purple	0.627	0.125	0.941	7	7	1165112	1.000	X	X	X		
8	Rice	Red	1.000	0.000	0.000	8	8	1070242	1.000	X	X	X		
9	Corn	Magenta	0.690	0.188	0.376	9	9	1083111	1.000	X	X	X		
10	Stubble	Yellow	1.000	0.843	0.000	10	10	1407896	1.000	X	X	X		
11	Grassland2	Brown	0.824	0.702	0.549	11	11	740057	1.000	X	X	X		
12	Grassland3	Dark Brown	0.627	0.318	0.173	12	12	1298115	1.000	X	X	X		
13	Urban/High clouds	Cyan	0.251	0.878	0.816	13	13	1112225	1.000	X	X	X		
14	Bare soil/Grasslands	Light Yellow	1.000	0.843	0.000	14	14	684373	1.000	X	X	X		
15	Sand	Yellow	1.000	1.000	0.000	15	15	238339	1.000	X	X	X		
16	Cloud1	White	0.961	0.961	0.863	16	16	268449	1.000	X	X	X		
17	Cloud2	White	0.961	0.961	0.863	17	17	342320	1.000	X	X	X		
18	Cloud3	White	0.961	0.961	0.863	18	18	190140	1.000	X	X	X		
19	Cloud4	White	0.961	0.961	0.863	19	19	226198	1.000	X	X	X		
20	Cloud	White	1.000	1.000	1.000	20	20	267624	1.000	X	X	X		

Figure 4-7 - Erdas Imagine Signature Editor of Test Image One

After a lengthy review and interpretation process using ancillary data described in Section 1.5, it was possible to define or identify the classes, and to assign corn to one of them. High-resolution imagery (Ikonos) has enough detail to determine field differences and special features, such as cattle, small roads, or trees. When combined with the unsupervised classification results, the DEM and land use layers allowed the analyst to visualize different types of agricultural fields. Figure 4-8 shows an example of the five geo-linked Erdas viewer screens used for this step. The screens were Test Image Two, Test Image Two unsupervised classification (with class editor), Ikonos image, Test Image One unsupervised classification (with class editor), and Test Image One.

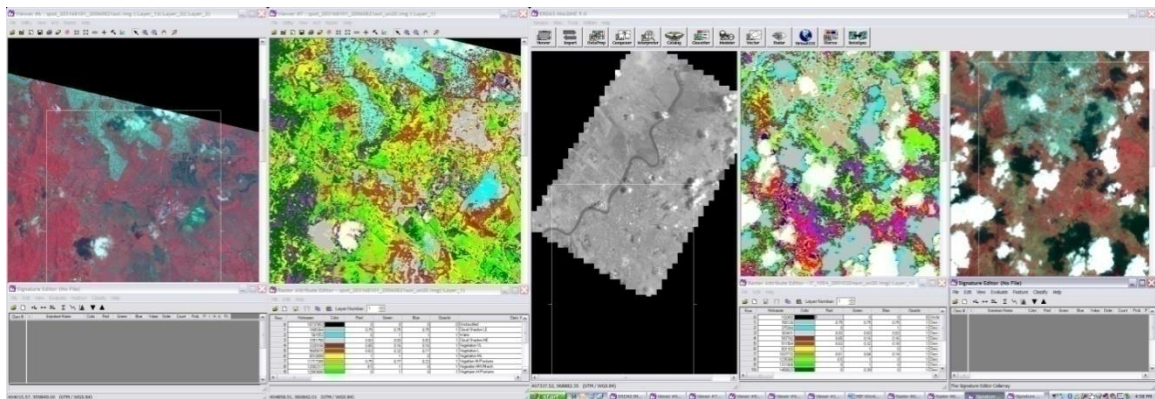


Figure 4-8 - Windows arrangement required for simulating ground truth from ancillary data

Following the ESRI (2007c) recommendations, the final number of training classes were 16 (see Subsection 4.2.1.1). The classes were: sea, clouds; water, primary forest, secondary forest, crops (corn and three more), grasslands (three types), bare soil, sand, clouds, and dense clouds. The Erdas Imagine Supervised Classification tool produced a classified raster based on the signature file and an input image with the same number of bands. A similar process took place in ArcGIS for Test Image Two: digitizing

representative polygons for each class and giving each polygon an integer class ID, using the Create Signatures tool, and using the Maximum Likelihood Classification tool. Although Erdas Imagine performance and functionality were better than ArcGIS, ArcGIS not only offered the same basic functionality but provided additional tools. For example, it has tools such as Filter and Aggregate for reducing the speckle (the latter one reduces the resolution); a tool for converting the raster into polygons (Raster to Polygon); and tools for merging the polygons like Eliminate and Dissolve.

4.2.2. NDVI Calculation

Subsection 2.2.1 explains the strong relationship between NDVI and yield estimation as well as the logic behind the index formula. NDVI may be derived from satellites that have sensors to measure the red and near-infrared spectral bands (see Table 4-3). Both Test Images have those two bands and consequently the model can use simple raster algebra to calculate NDVI on them. A different need was creating an historic series of NDVI values inside the AOS. The regression analysis required the NDVI series to create the mathematical equation that predicts the yield. A complete, ten-year series of images taken just before the August to September harvesting season was recommended for NDVI calculation.

Table 4-3 - Band's names for NDVI calculation from different sensors

Program	Sensor	Red Band	Near-Infrared Band
Landsat	ETM+, TM	Band 3	Band 4
Spot	HRG	Band 2	Band 3
NOAA	AVHRR	Channel 1	Channel 2
Terra - Aqua	MODIS	Band 2	Band 1

Although the project already had Landsat Imagery available, the imagery presented various problems: the dates of capture were not consistent, cloud coverage obscured the vegetation below, and quality was uneven. There were not enough Landsat images with light cloud coverage, during the ideal time of the year, and with the required quality (SLC-Off malfunction). A second option was to utilize the widely used AVHRR imagery from NOAA meteorological satellites. However, it was not freely available and the resolution is 1km, at best (Towson University, 2007). Finally, a GLCF product offered a cloud free option: the MODIS, 500m, 32 days composites (see Subsection 3.1.3). Different NDVI from different satellites are consistent (Buheasie, Tsuchiya, Kaneko, & Sung, 2003).

The ArcGIS Clip tool trimmed the bands and created significantly smaller files covering the AOS. The ArcCatalog Load Data function created a single raster dataset with ten bands (two bands per year for five years). Although the image had geographic references, it needed a georeference procedure to adjust the image to the AOS. The first procedure was to move the image using the Shift tool, but some features appeared displaced from their original location. The ArcGIS Georeferencing Toolbar provided the mechanism to adjust image to the state boundary.

The Math tools (Sum, Minus, Divide and Multiply) in the ArcGIS Spatial Analyst extension calculated the NDVI for the years 2001 to 2005. Initially it did not work because the temporary directory pathname was too long, but changing the file path in the

Spatial Analyst settings solved the problem. The math operations worked in part because the resulting NDVI had integer values of -1, 0, 1. Math tools created a raster with integer values because the original rasters have integer values; mathematical operations in ArcGIS should create raster with float values only if at least one of the rasters has float values. The solution was to convert the numerator (Near-infrared band + red band) into a floating-point value raster with the Float tool. The prototype used an ArcGIS Model to automate the execution of this operation (see Section 5.4). The regression analysis used the average value on the resulting image for each year (see Section 4.3).

4.2.3. Interpolation

A weather station captures temperature and rainfall readings on a given point over the Earth's surface. Consequently, the spatial representations of weather stations are points. However, it is not possible to have weather stations everywhere. Using Tobler's First Law of Geography and Spatial Autocorrelation, an interpolator creates the required surface representation based on sample points and their relative distances and weights (ESRI, 2007c). The reduced number of weather stations in the AOS and the \$500 budget limited this project's sample to 25 stations (see Section 3.3). This provided enough detail to capture important weather behaviors that extend beyond the AOS (see Figure 4-9). Gomme (2003) supported that a limited number of weather stations (25 per country, 245 over neighboring countries) provide enough information to create regional rasters and statistics.



Figure 4-9 - Weather stations location extending beyond Cordoba boundary

The next step was to determine which interpolator to use and its parameters. Because of the sparse sampling density, only the Inverse Distance Weighting (IDW) and Spline interpolators were considered. IDW averages values of samples near the cell, on the assumption that the data are spatially correlated. This interpolator estimates a value at a given point as the weighted average of the values of the known points, using the inverse of the distance as the weight factor. The spline interpolators create smooth surfaces that pass through the sample points. The spline method “is best suited for gently varying surfaces” (ESRI, 2007c). Temperature and rainfall surfaces vary locally and smoothly. Weather variables typically exhibit a significant spatial correlation. Both interpolators were deemed potentially suitable for the prediction model. Parameters selected for the IDW interpolation were Power = 2 (Controls the relative importance of distance) and a Search Radius Variable (the other option is a fixed distance around the point) for twelve

points (number of sample points used by the interpolator). Spline types tested were Tension (adjusts the sensibility of the interpolator to the modeled parameter) and Regularized (yields a smooth surface). Parameters for the Spline interpolators were Weight = .1 (determines how smooth the curvature is by determining the weight of the third derivatives in the curvature equation in Regularized type or the weight of the tension in the Tension type) and Number of points = 12 (number of sample points used by the interpolator) (ESRI, 2006).

By default, both interpolators transformed null data (<NULL>) to zero, creating erroneous anomalies in the resulting surface (see Figure 4-10). The possible solutions included checking that null data were effectively being imported as null data (presented in raster datasets as <nodata>), revising Spatial Analyst and ArcGIS Environment settings, comparing the results from the Geostatistical Wizard and installing ArcGIS Desktop service packs. The alternatives did not lead to determining the cause of the problem, but instead to the implementation of an alternative method that first selected by attributes (IS NOT NULL condition), and then exported the layer as a new feature preventing the interpolator from failing and showing the message “feature class does not exist”.

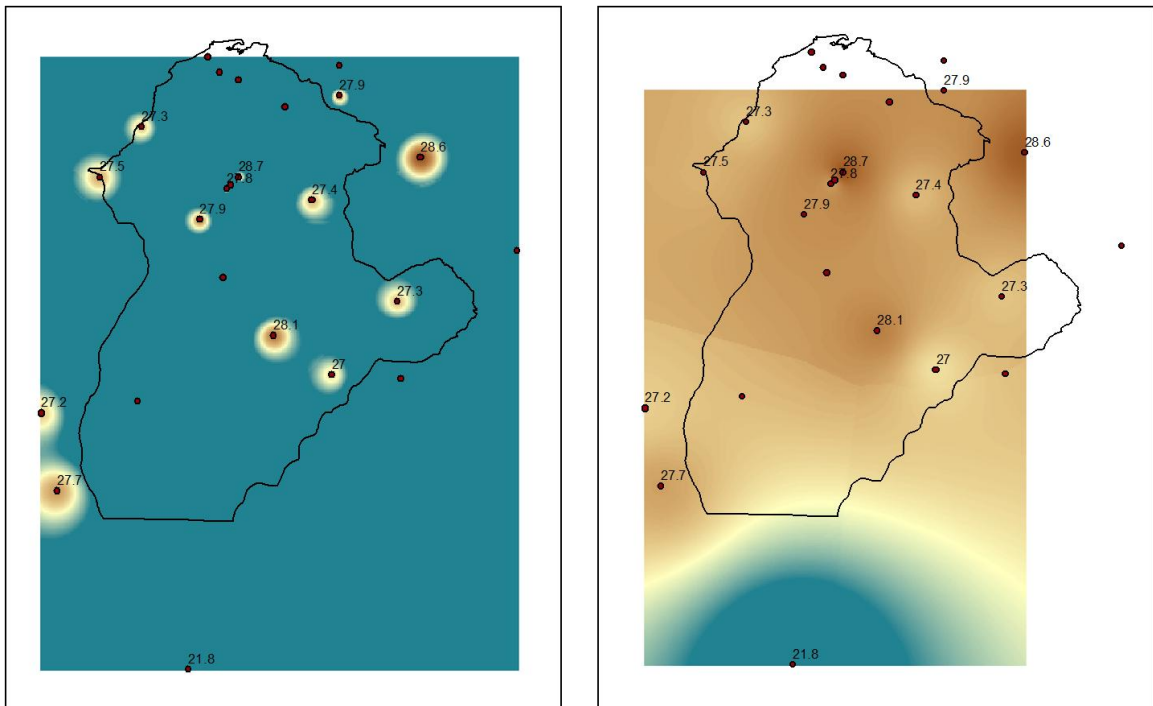


Figure 4-10 - IDW interpolation of temperature data including null values (A) and removing null values (B)

Viloria de la Hoz (2004) recommended a method to test the surface, first removing a sample point before creating the surface, then testing how the interpolator predicts the removed sample value. Then, this procedure, known as jackknifing, removes and tests a different station. The process repeats with a different station until all the stations, in both variables have been evaluated using the three interpolators. The Extract Values to Points tool read the raster for the exact values at the removed stations points. The jackknifing

test revealed that IDW estimated weather values more accurately than either of the Spline-based interpolators (see Table 4-4 and Table 4-5).

Table 4-4 - Summary of rainfall estimation errors by interpolant (best result highlighted)

Station ID	Actual Value (mm)	IDW		Spline Regularized		Spline Tension	
		Estimation with station removed	Squared error of estimation	Estimation with station removed	Squared error of estimation	Estimation with station removed	Squared error of estimation
1111502	119.90	314.88	38,017	1,077.22	916,462	690.09	325,117
1204502	220.80	283.18	3,891	263.06	1,786	331.96	12,357
1307503	211.20	270.32	3,495	287.11	5,762	186.16	627
1307505	269.60	219.92	2,468	209.02	3,670	208.93	3,681
2502517	389.40	250.09	19,407	71.63	100,978	157.88	53,602
1305503	223.60	262.18	1,488	373.92	22,596	305.08	6,639
2502515	376.60	423.39	2,189	648.04	73,680	536.61	25,603
2501501	252.20	390.91	19,240	389.88	18,956	401.12	22,177
2502531	170.20	170.50	0	90.99	6,274	96.70	5,402
2502527	144.40	198.16	2,890	112.30	1,030	141.85	7
2502518	105.10	280.12	30,632	301.85	38,711	242.23	18,805
2502516	500.50	401.58	9,785	438.60	3,832	473.28	741
1204501	358.60	230.59	16,387	232.54	15,891	188.00	29,104
1201509	260.00	184.60	5,685	52.78	42,940	78.77	32,844
1201502	103.60	282.99	32,181	257.52	23,691	232.01	16,489
2625504	687.00	373.34	98,383	541.82	21,077	508.08	32,012
Total squared error			286,140		1,297,336		585,206

Table 4-5 - Summary of temperature estimation errors by interpolant (best result highlighted)

Station ID	Actual Value (C)	IDW		Spline Regularized		Spline Tension	
		Estimation with station removed	Squared error of estimation	Estimation with station removed	Squared error of estimation	Estimation with station removed	Squared error of estimation
1111502	21.80	25.15	11	(5.84)	764	10.48	128
1204502	27.30	26.76	0	29.05	3	28.06	1
1307503	28.70	27.37	2	27.56	1	27.62	1
1307505	27.80	28.22	0	28.65	1	28.58	1
2502517	27.40	25.23	5	37.69	106	33.32	35
1305503	27.90	27.93	0	26.21	3	27.15	1
2502515	27.30	18.83	72	9.23	327	11.73	242
2501501	28.10	27.46	0	27.10	1	27.03	1
2502527	27.90	17.77	103	31.16	11	28.92	1
2502518	28.60	22.31	40	27.47	1	27.76	1
2502516	27.00	19.20	61	14.23	163	14.03	168
1204501	27.50	26.93	0	26.47	1	27.10	0
1201509	27.70	26.19	2	28.83	1	27.63	0
1201502	27.20	26.74	0	28.04	1	27.81	0
Total squared error			296		1,383		580

4.3. Multiple Linear Regression (MLR) Analysis

Priya & Shibasaki (2001) considered the selection of the model as an important step that should consider the objectives of the project, data requirements, broad usage, and realistic accuracy levels. The project required a moderately accurate prediction model to fulfill its objectives. Broad use, discretionary data requirements, and accuracy dependent of the number of samples are the characteristics of a MLR model (see Section 2.1). Section 2.2

describes the regression analysis model based on rainfall, temperature, soil moisture and NDVI used by Prasad et al. (2006).

Using Excel's Statistical tools, the regression analysis validated and estimated the relationships between the three historic independent variables: NDVI, rainfall, and temperature; and the dependent variable: Cordoba's corn yield rate. Section 4.2.2 describes the NDVI series creation. Averaging the Cordoba weather station readings for each cropping season created the historic series year average of temperature and rainfall. Traditionally cropped corn yield rate (tons/hectare) was the response variable. Table 4-6 shows the original data used in the linear regression. Table 4-7 shows the results of a MLR of corn yield rate on rainfall, temperature, and NDVI between 2001 and 2005.

Table 4-6 - Data used in multiple regression analysis

Year	Traditional cropped corn			Predecessors		
	Area (hectares)	Yield (tons)	Yield Rate (tons / hectare)	Rainfall (mm)	Temp (°C)	NDVI
1988	44,800	61,600	1.3750	155.56	27.57	
1989	55,900	80,500	1.4401	132.51	27.47	
1990	63,000	85,900	1.3635	107.23	27.78	
1991	71,000	101,500	1.4296	127.51	27.82	
1992	67,470	97,105	1.4392	137.27	27.78	
1993	65,190	100,829	1.5467	120.96	27.86	
1994	60,721	92,143	1.5175	119.23	27.73	
1995	46,747	72,622	1.5535	95.01	27.66	
1996	43,521	75,112	1.7259	96.23	27.34	
1997	43,268	72,594	1.6778	93.93	28.24	
1998	37,370	58,449	1.5641	160.33	28.28	
1999	44,678	79,606	1.7818	115.37	27.90	
2000	33,473	68,713	2.0528	122.25	27.59	
2001	40,178	74,660	1.8582	101.85	27.98	0.7328
2002	38,784	73,739	1.9013	79.14	28.32	0.7532
2003	36,060	66,009	1.8306	155.61	27.85	0.7472
2004	32,000	59,553	1.8610	127.10	27.77	0.7222
2005	32,125	56,611	1.7622	197.22	27.83	0.6580
2006				117.42	27.73	

Note: Area and Yield obtained from Ministerio de Agricultura (2007).

Table 4-7 - Results of a MLR of yield rate on rainfall, temperature and NDVI (2001-2005)

<i>Regression Statistics</i>	
Multiple R	0.9821
R ²	0.9645
Adjusted R ²	0.8582
Standard Error	0.0194
Observations	5

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig F</i>
Regression	3	0.01026	0.00342	9.06709	0.23833
Residual	1	0.00038	0.00038		
Total	4	0.01064			

	<i>Coeff</i>	<i>SE</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	15.9767	27.5284	0.5804	0.6652
Rainfall	-0.0036	0.0062	-0.5761	0.6673
Temperature	-0.3513	0.9274	-0.3788	0.7695
NDVI	-3.6554	5.9039	-0.6191	0.6471

Table 4-7 shows the Excel MLR output tables, the Analysis of Variance table (ANOVA) and the coefficients table. The ANOVA table has five columns: degrees of freedom, sums of squares, mean square, the f-test and its significance. Degrees of freedom (df) represent the number of observations available for estimating other variable. Sums of squares (SS) represent how disperse are the observations. The mean square (MS) is a standardized measure of dispersion based on the sum of squares and the degrees of freedom. The f-test (F) and its significance (Sig F) indicate whether the variance of the observations behaves as a normal distribution.

The coefficients table shows the results of the MLR in terms of the numeric relationships between the predecessor variables and the phenomenon being predicted. In this case, the table shows that yield rate starts at 15.9767 tons per hectare and decreases 0.0036 for each mm of rainfall, decreases 0.3513 for each degree Celsius and decreases 3.6554 for NDVI values of one. The negative coefficients were the result of the data values used on the MLR, and were not interpreted as erroneous. For example, Prasad et al. (2006) found negative rainfall coefficients for corn yields; and negative rainfall, temperature and NDVI coefficients for soybean yields. The coherence of these coefficients can be tested using the standardized error (SE) and the t-test (t Stat), but in this case the small number of observations generates a meaningless value.

A recognized accuracy indicator of the MLR is R Square (also written as R²). The basic MLR procedures are running the MLR and evaluating R². Although a MLR based on only five data points tends to show the R² value near one, it is less significant. Rogerson (2006) presents two alternatives to this issue: removing the NDVI series (see Table 4-8), or estimating NDVI values for the other years based on rainfall and temperature (see Table 4-9).

Table 4-8 - Results of a MLR of yield rate on rainfall and temperature (1988 -2005)

<i>Regression Statistics</i>	
Multiple R	0.2454
R ²	0.0602
Adjusted R ²	-0.0651
Standard Error	0.2127
Observations	18

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig F</i>
Regression	2	0.04347	0.02173	0.48044	0.62771
Residual	15	0.67859	0.04524		
Total	17	0.72206			

	<i>Coeff</i>	<i>SE</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	-2.5915	5.5145	-0.4699	0.6452
Rainfall	-0.0009	0.0018	-0.4884	0.6324
Temperature	0.1563	0.1972	0.7928	0.4403

Table 4-9 - Results of a MLR of NDVI on rainfall and temperature (2001-2005)

<i>Regression Statistics</i>	
Multiple R	0.7941
R ²	0.6306
Adjusted R ²	0.2613
Standard Error	0.0328
Observations	5

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig F</i>
Regression	2	0.00367	0.00183	1.70733	0.36937
Residual	2	0.00215	0.00107		
Total	4	0.00581			

	<i>Coeff</i>	<i>SE</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	1.5281	3.1150	0.4906	0.6723
Rainfall	-0.0007	0.0005	-1.4118	0.2935
Temperature	-0.0253	0.1096	-0.2310	0.8388

The R² was significantly better in the second alternative. The resulting coefficients of the MLR of NDVI on Rainfall and Temperature allowed the modeler to estimate NDVI values for additional years (see Table 4-10). The equation used for estimating NDVI is:

$$NDVI = 1.5281 - .0007 \text{ Rainfall} - 0.0225 \text{ Temperature}$$

This equation may be used to estimate values since 1988. However, a MLR using estimated NDVI since 1988 produced a R² of 0.06. Other MLRs suggested that it is extremely inaccurate using the equation for estimating NDVI before 1997.

Table 4-10 - Data including estimated NDVI for years 1997 to 2000

Year	Yield Rate (tons / hectare)	Rainfall (mm)	Temp (°C)	NDVI
1997	1.6778	93.93	28.24	0.7435
1998	1.5641	160.33	28.28	0.6936
1999	1.7818	115.37	27.90	0.7363
2000	2.0528	122.25	27.59	0.7391
2001	1.8582	101.85	27.98	0.7328
2002	1.9013	79.14	28.32	0.7532
2003	1.8306	155.61	27.85	0.7472
2004	1.8610	127.10	27.77	0.7222
2005	1.7622	197.22	27.83	0.6580

The final step was running a MLR of yield rate on rainfall, temperature, and estimated NDVI (see Table 4-11). The result showed an acceptable R² of 0.7196. The prediction model equation is:

$$\text{Yield \%} = 14.4283 - .0020 \text{ Rainfall} - .4516 \text{ Temperature} + .3747 \text{ NDVI}$$

Table 4-11 - Results of a MLR of yield rate on rainfall, temperature and estimated NDVI (1997 - 2005)

<i>Regression Statistics</i>	
Multiple R	0.8483
R ²	0.7196
Adjusted R ²	0.5514
Standard Error	0.0928
Observations	9

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Sig F</i>
Regression	3	0.11059	0.03686	4.27746	0.07579
Residual	5	0.04309	0.00862		
Total	8	0.15368			

	<i>Coeff</i>	<i>SE</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	14.4283	4.9812	2.8965	0.0339
Rainfall	-0.0020	0.0018	-1.1507	0.3019
Temperature	-0.4516	0.1479	-3.0541	0.0283
NDVI	0.3747	2.0032	0.1871	0.8590

5. Model-GIS Integration and Results

One of the client's requirements is to encourage the use of ESRI software. The prototype uses only the functionality of ESRI ArcGIS software. ArcInfo license level is the most complete ESRI desktop product and provides most of the tools for storing, retrieving, visualizing, and processing the information. The software provides its own user interface. Figure 5-1 shows how ESRI ArcGIS provides a unique way for the user to interact with all the elements of the system.

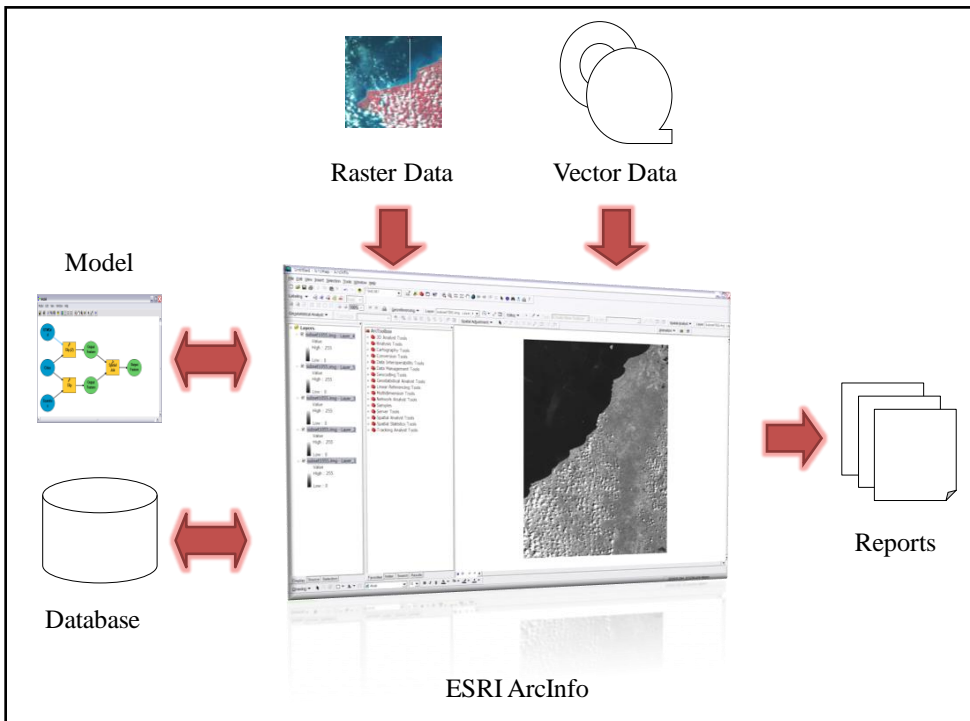


Figure 5-1 - ArcGIS provides the software and the correspondent user interface

ArcGIS not only has the functionality to store the information, but also the option to create and run multiple automated functions (Models). The modeling tool used to combine the geodatabase, the raster imagery processing, and the regression coefficients is called Model Builder. A model, or a sequence of parameterized tools, reduces the user time required to present the demonstration, the interface complexity, and some user errors. Furthermore, models improve the impact of a commercial demonstration using the prototype and permitting a relatively easy incorporation and use of future data.

This chapter discusses how to integrate the tasks described in Chapter 4, such as classifying a raster image, interpolating weather readings, calculating NDVI and running the MLR model. A Geodatabase toolbox called Prediction stores the project models inside the geodatabase Mip.gdb.

5.1. General Considerations

All models, unless specified, have pre-defined environment settings. In the general settings, the current workspace is Mip.gdb; the scratch workspace is Mipscratch.gdb; the output coordinate system is WGS_UTM_18N; and the extent is set to default. In the

raster analysis settings, cell size depends on the output coordinate system. By modifying it, it is possible to manage model performance and detail. Smaller cell sizes offer more detail but increase raster operations time. Mask is the SelectedAOI dataset. These setting transform pixels outside the AOI in Nodata.

The Import Symbology option, in the output dataset properties, assigns default symbology for each output layer. Output layers also have the Intermediate Data option unchecked, so that all results are stored within the geodatabase. If the model is open while running, Intermediate Data is not deleted. Model parameters have predetermined default values. A Spatial Analyst license is required.

5.2. AOI Selection Model

A professional presentation was one of the required deliverables of the project. The AOI Selection Model allows the user to select a predefined AOI and clip the analysis image using the AOI shape. Furthermore, this model reduces the size of the images used and assures a clean results display in the interface. The first part uses a Select SQL expression to pick the AOI by feature ID from the AOI feature class. The second part extracts the pixels from the input image inside the Selected AOI feature created in the first part. Figure 5-2 shows the model diagram in Model Builder and Figure 5-3 shows the model's result using Test Image Two and AOI = 3.

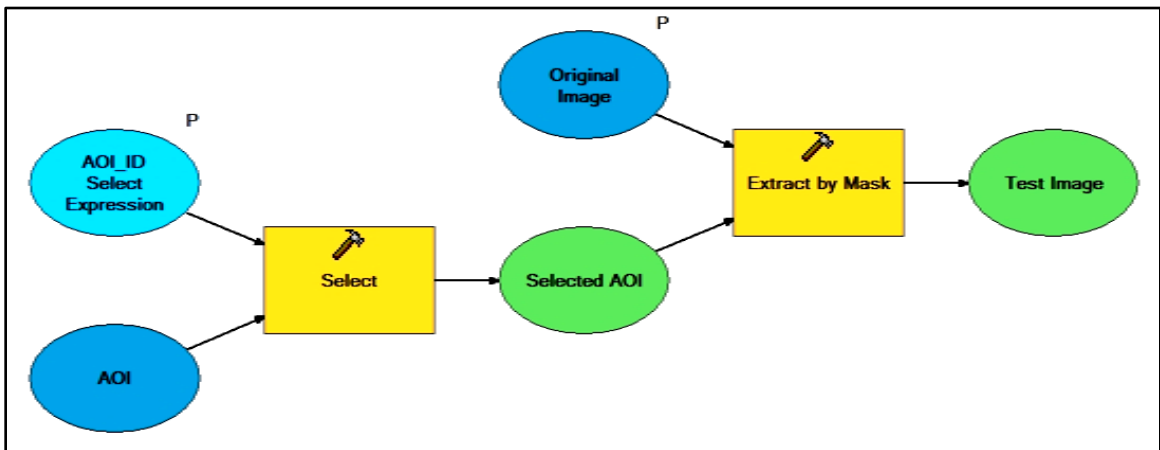


Figure 5-2 - AOI Selection model diagram

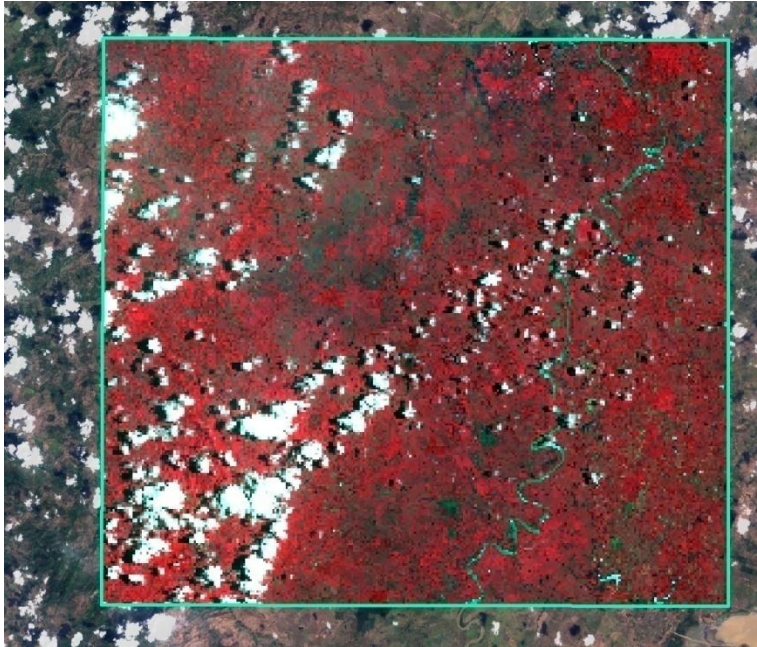


Figure 5-3 - Result from the AOI Selection Model using Test Image Two

Initially, the Clip tool in ArcGIS did not create the raster based on the features. The problem apparently occurred because of the different reference system between the AOI layer and the raster. Tests such as saving the AOI layer in the same reference system as the raster to clip, changing the file paths to reduce the incidence of long pathnames or special characters, or modifying spatial analyst settings did not give a satisfactory result. The ESRI Support Center recommended reinstalling and patching ArcGIS; after reinstallation, the problem continued but the installation of ArcGIS Desktop Service Pack 3 partially fixed the problem. The fact that the Clip tool worked when the user manually typed the coordinates proved that the tool works, but also proved that the SelectedAOI feature class must have the same coordinate system as the image. At the end, Extract by Mask produced better results than the Clip tool because it handled the user's AOI preference dynamically. In addition, Extract by Mask was slower than Clip, but it handled different AOI shapes, while Clip handled only rectangles. For cell size as the minimum of inputs, average time was three minutes and twenty seconds, and seventeen seconds for cell size equal to 125 meters.

5.3. Interpolation Model

The Interpolation model automates the process described in Subsection 4.2.3. The model requires special organization of the Weather Station dataset as described in Section 3.3. The model's extent is the same of States dataset. After this change, IDW operations covered completely the AOS and prevented the re-sampling of the raster pixels. The Interpolation Model has two parameters that select variables, such as temperature (T) or rainfall (R), and month and year. Field determines the field that contains the weather readings. SQL Expression selects the weather stations with data in the Field parameter. Both should use the same field value. Figure 5-4 shows the model diagram in Model Builder and Figure 5-5 shows the model's result using rainfall readings in August 2008.

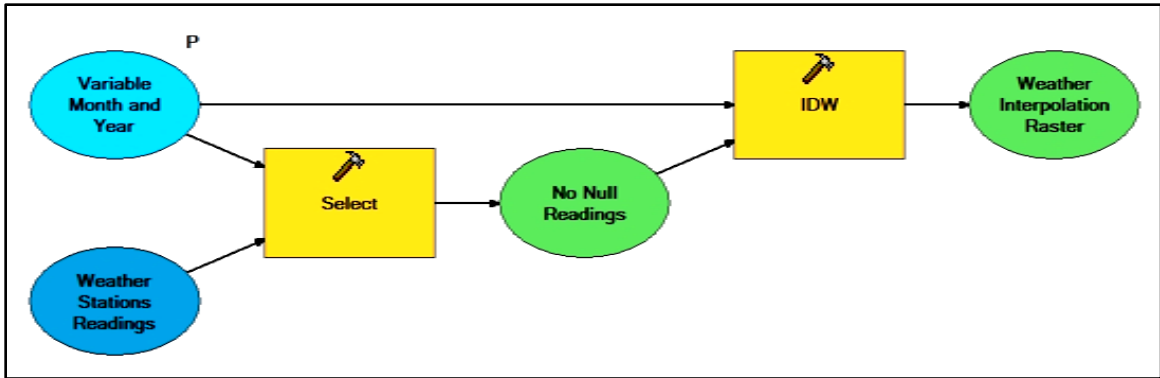


Figure 5-4 - Interpolation model diagram

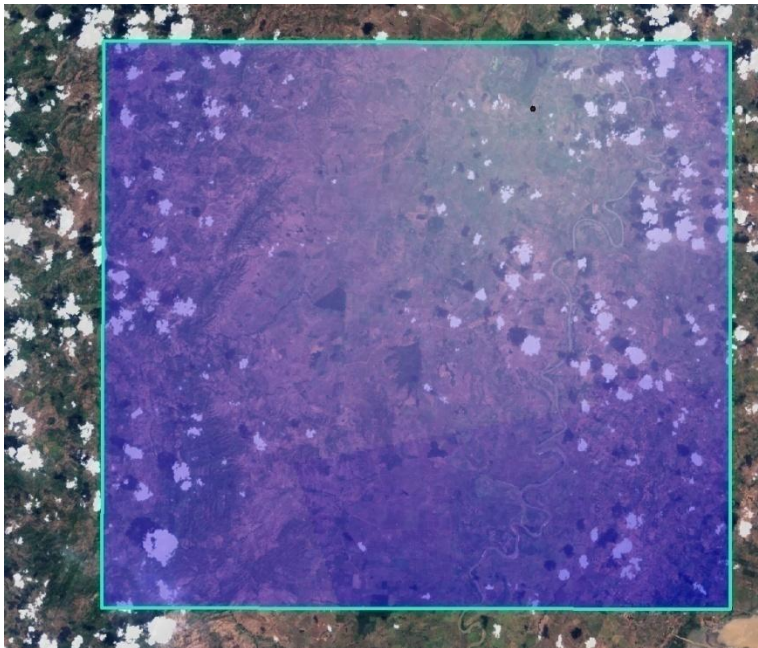


Figure 5-5 - Result from the Interpolation Model using rainfall readings in August 2008

5.4. NDVI Model

The NDVI model replicates the NDVI formula described in Subsection 4.2.2. This model has two parameters: Red Band and Near-infrared Band. The parameters add the flexibility for the use of both Landsat and Spot Imagery bands in a unique NDVI model. The Float tool allows the Divide tool to produce floating point NDVI values as required in Subsection 4.2.3. Figure 5-6 shows the model diagram in Model Builder and Figure 5-7 show the model's result using red and near-infrared bands on Test Image Two.

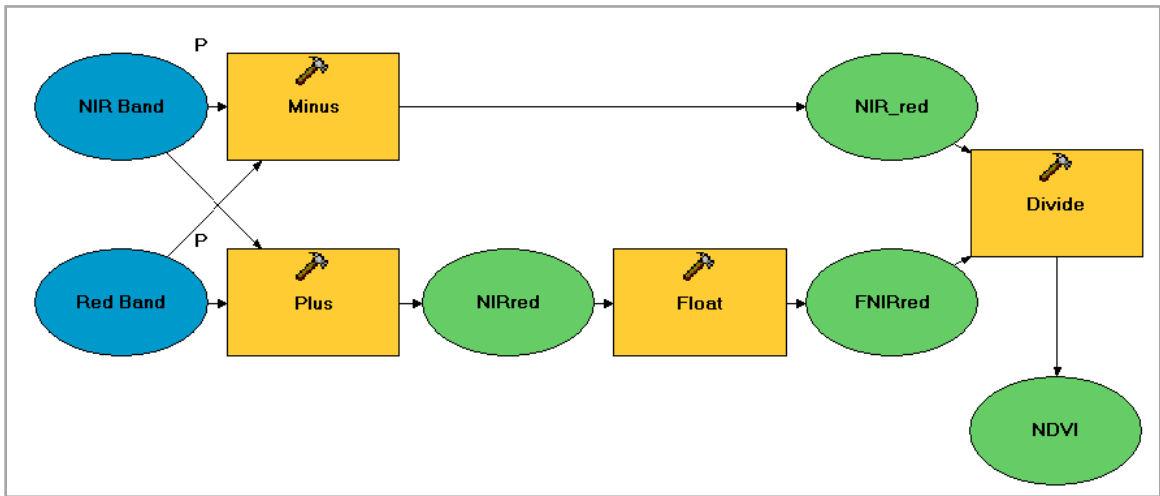


Figure 5-6 - NDVI model diagram

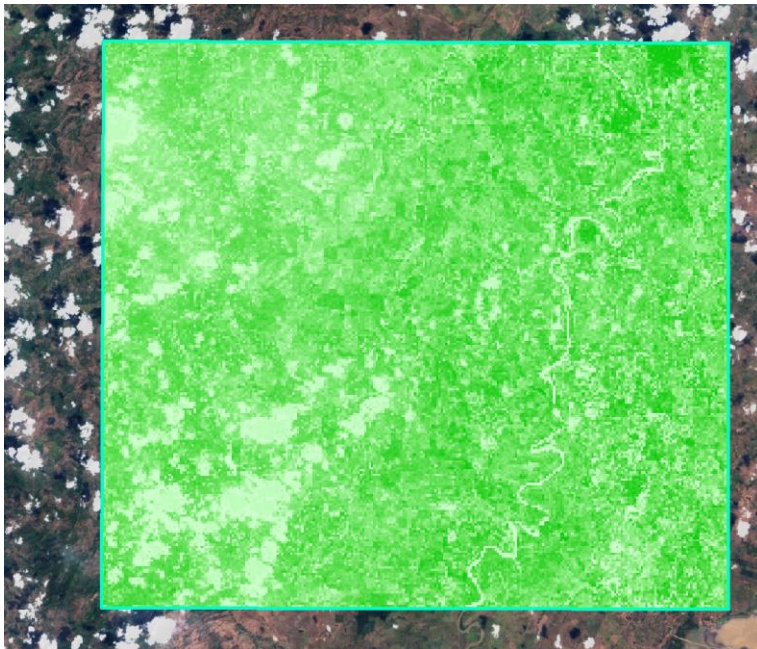


Figure 5-7 - Result from the NDVI Model using Test Image Two

5.5. Classification Model

Although ESRI does not provide robust image processing software, this project used ESRI ArcGIS classification functionality because it integrated seamlessly with other ESRI tools that smooth the raster, such as Filter or Majority Filter.

Classification Model is the result of attempting to develop a model that automatically identifies corn from a multispectral image, and automates the process described in Subsection 4.2.1.1 and 4.2.1.2. The parameter is the Signature File. Different pre-made signature files are available for unsupervised and supervised classification in both Landsat and Spot Images (see Table 5-1).

Table 5-1 - Available signature files

Signature File Name	Satellite	Classes	Method
Spot_3.gsg	Spot	3	Unsupervised
Spot_4.gsg	Spot	4	Unsupervised
Spot_15.gsg	Spot	15	Unsupervised
Spot_18.gsg	Spot	18	Unsupervised
Spot_sup18.gsg	Spot	18	Supervised
Spot_20.gsg	Spot	20	Unsupervised
Spot_25.gsg	Spot	25	Unsupervised
Landsat_18.gsg	Landsat	18	Unsupervised
Landsat_18sup.gsg	Landsat	18	Supervised

Again, the environment's cell size had a direct impact on processing time. For example, when the model ran using the default value of 125, it took only 19 seconds; with 25 it took 42 seconds; and when it ran with minimum of input options it took five minutes. Immediately after the Maximum Likelihood Classification process, a Majority filter reduces the raster speckle. Figure 5-8 shows the model diagram in Model Builder.

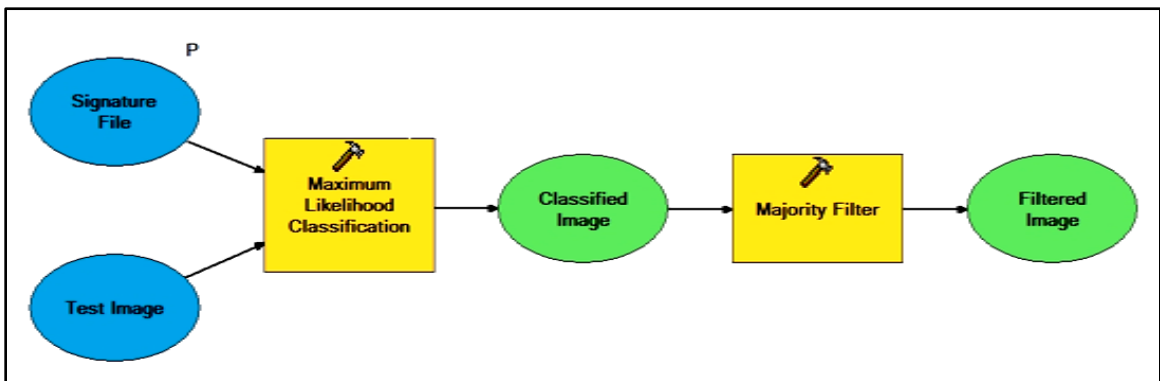


Figure 5-8 - Classification model diagram

Figure 5-9 shows the model's result for the selected AOI in the Test Image Two. For this example, the signature file used was Spot-sup18.gsg. This result is comparable to other supervised classifications on a Spot Image. Yellow pixels are classified as corn.

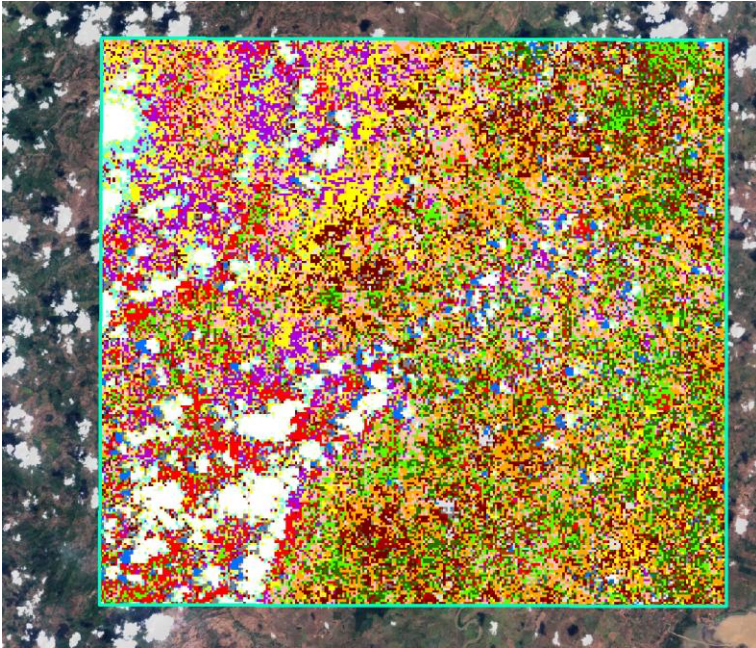


Figure 5-9 - Result from the Classification Model using Test Image Two

5.6. Integration Model

The Integration Model is the ultimate tool of the prototype, using the regression model, the NDVI calculation, the weather readings interpolation, and the image classification to obtain estimated crop yields. The Linear Regression model contains two ArcGIS tools: a Weighted Sum tool that multiplies the regression coefficients (obtained in Section 4.3) by the values in the NDVI raster, and temperature and rainfall interpolation rasters; and a Plus tool that adds the MLR intercept value into the equation. The initial result is a raster that contains corn yield rate (see Figure 5-10).

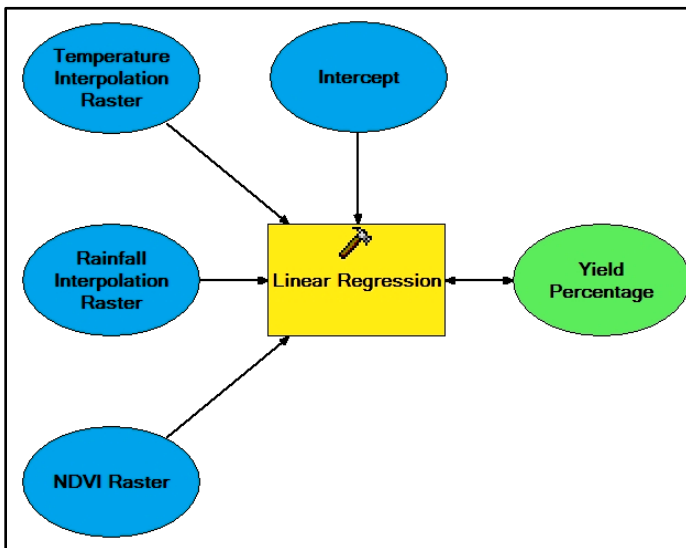


Figure 5-10 - Regression model diagram

This model has only the parameter Hectare Equals to normalize corn yield rate (tons per hectare) accordingly to the cell size. Yield Per Cell Normalization divides Yield Rate by hectare equivalence. The reclassify tool selects Corn Cells from the image previously classified and filtered. The result is a raster of the corn yield rate. It is equivalent to assigning a Yield per Cell to Corn Cells. Figure 5-11 shows the second part of the model diagram in Model Builder.

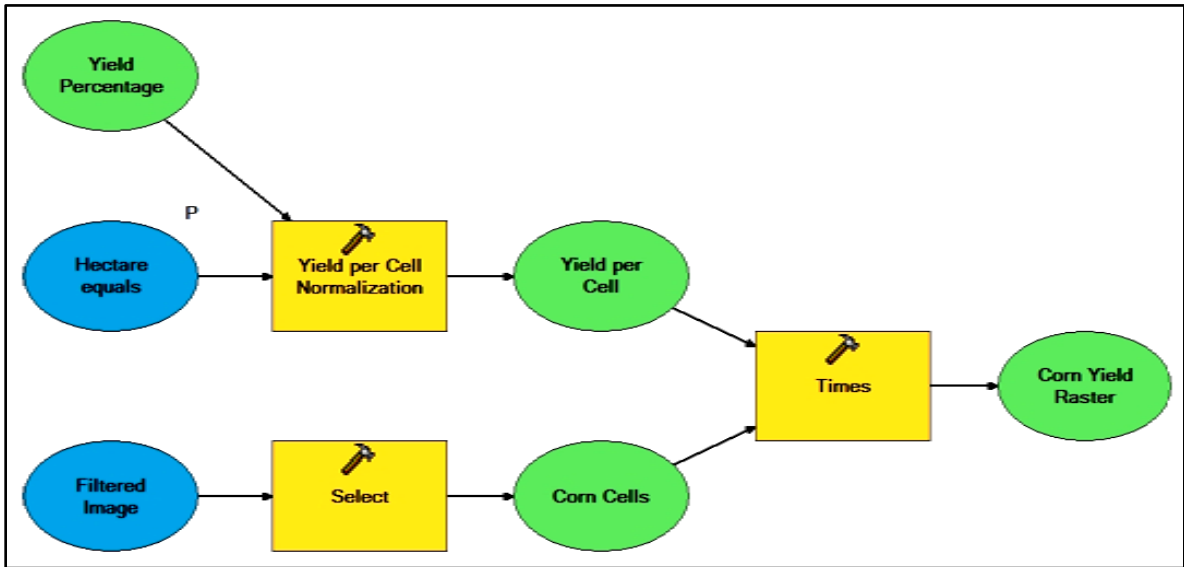


Figure 5-11 - Integration model diagram

Figure 5-12 shows the final raster obtained from the integration model based on the resulting raster from Sections 5.3, 5.4, and 5.5. The example shows the cells planted with corn inside AOI 3, in August 2006. The cells are displayed in different colors equivalent to five Quantiles of estimated corn yield rate per hectare.

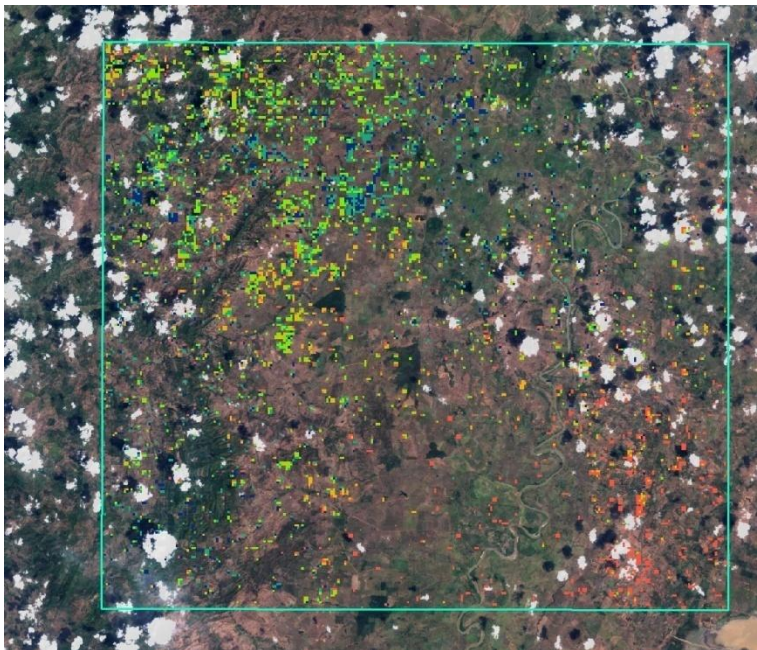


Figure 5-12 - Areas planted with corn displayed by the estimated yield of each cell

5.7. Result

The prototype produced a raster in which each pixel value denoted the corn yield rate per hectare within the pixel. The raster presents to the user corn areas and yields inside a given AOI. This final raster is presented to the user in a Yield Report which include the estimation statistics (see Figure 5-13).

The Yield Report is an ArcMap file (Report.mxd) with its layout pre-defined. The report communicates the basic statistics from the resulting raster. The raster itself provides all the data, such as cell size, number of cells, and the pixel values, required to estimate corn yield on a given AOI. Total AOI area is obtained by multiplying cell area per the total number of cells in the raster. Total corn planted area is obtained by multiplying cell area per the number of cells with values different from zero. An estimated yield in the AOI is calculated by adding all cell values and dividing the result by the cell to hectare equivalent. Estimated average yield per hectare is the result of dividing forecasted yield by the number of hectares planted.

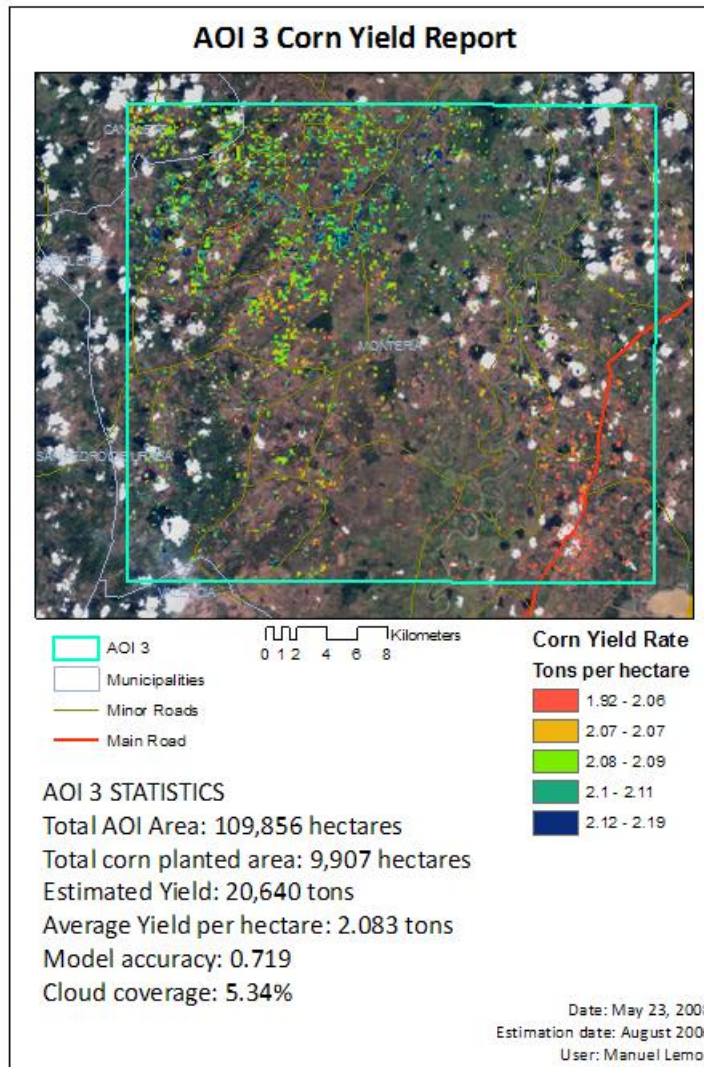


Figure 5-13 - Corn Yield Report inside AOI in August, 2006

In the example shown in Figure 5-13 the raster size is 279 cells by 252 cells equivalent to 70,308 cells, which multiplied per the cell size of 125m by 125m, equals 1,098,562,500 squared meters, or 109,856 hectares. The area planted with corn is 6,341 cells or 9,907 hectares. The sum of pixel values is 13,209.6 and when divided by .64 (cell size equivalent to hectare) obtains a yield of 20,640 tons. Therefore, the average yield is 2.083 tons per hectare

Model accuracy is also shown in the report, as well as cloud coverage and gaps from SLC-off Landsat malfunctions. Although the prototype does not specifically include images with SLC-Off characteristics, cloud presence is an unwanted constant over the Colombian territory. The report shows the model accuracy and cloud coverage data, both of them affecting the overall accuracy.

6. Conclusions and Recommendations

The project demonstrated that GIS technologies could effectively help to estimate crop yields in Colombia. The project organized and brought together different data types, and integrated them into an interesting GIS application for agriculture clients. The prototype demonstrates the basic functionalities of using GIS and remote sensing imagery technologies as a decision-making tool in terms of analysis, data storage, source integration, process productivity, and accuracy, and uses historic weather data and satellite imagery to predict corn yields in the area of study, based on a linear regression model.

Data acquisition was a key issue in this project, especially quality historic satellite imagery. Remotely sensed imagery proved to be an effective method to identify clouds and their shadows. The cloud coverage played an important role in assessing model accuracy, defining the AOIs, and pre-selecting acceptable imagery. Although better accuracy is desirable and can be improved, its improvement was not an objective of this project. However, additional project achievements included defining the required and available data, supporting the used procedures, designing the models, and defining key parameters.

The resulting prototype is portable and it works on a single workstation without using network resources or interfaces with other systems. However, the ArcGIS interface permits the use of many compatible data formats. The implementation of the system had the continuous support of Procalculo Prosis and the University of Redlands staff. The resulting prototype is meaningful and operational as a demonstration of a GIS. It does not pretend to be the final system that the customers can use.

The prototype will support the decisions about GIS acquisition inside the BNA and Ministry of Agriculture. These key customers are currently interested in developing a full-scale system in the country. Procalculo Prosis could use the prototype as a proof-of-concept project for these customers. The prototype budget was \$500.00; small if compared with the resources needed to implement an operational countrywide system. Although the initial investment in technology in similar Latin American countries' projects started around two million dollars, many financial institutions have the resources to promote these initiatives. An example is the \$52.3 million fund for the Productive Agribusiness/Farmer Partnerships Initiative of the World Bank (World Bank Group, 2007)

6.1. Recommendations

Based on the results and the data used, this section presents some recommendations for the improvement of accuracy in this prototype and the further development of the system. Fiset et al. (2006) discussed different methodologies related to the image classification process. The most important were linking and understanding the phenology and remote sensing readings, and recollecting and storing ground truth information. A real yield estimation application, based on remote sensing and GIS technologies, requires including recent field observations. In addition, several studies have shown that multi-temporal images improve the classification accuracy by utilizing different spectral responses of the vegetation over a period. In relation to the weather data, including available weather

stations and using each reading metadata for accuracy assessment can improve the interpolation accuracy or at least determine the lack thereof.

Radarsat-2, the new generation of cloud-penetrating radar satellites from MDA, has a spatial resolution of three meters and multi-polarization beams. Brown & Ellis (2000) demonstrated that the use of multi-polarized imagery (3 bands) provides enough information for fairly effective crop identification and measurement. The satellite was finally launched in 2007 and has been operative since May, 2008. It offers a new alternative to multispectral data in places with high levels of cloud coverage.

In the prediction model realm, Gomma (2003) states that developing countries can benefit from the use of non-parametric models. "Simulation models have also reached a degree of sophistication where only marginal improvements can be expected. Non-parametric methods, on the other hand, are still young. Significant progress for crop yield forecasting can be expected" (Gomma, 2003). It is the opinion of the author that GIS can also help in this initiative.

From the technology perspective, the author considers that the use of ESRI Server products will increase performance and functionality. For example, Image Server can organize the raster files, manage larger amount of images and storage requirements, work with multiple spanning images, and facilitate the retrieval and mosaicing of the system imager. The use of ArcGIS Server enables GIS as decision support system, allowing an organization not only to analyze the information, but also to compare and share the results.

7. References

- Allen, R., Hanuschak, G., & Craig, M. (2002). History of remote sensing for crop acreage. *National Agricultural Statistics Service - United States Department of Agriculture* Retrieved October 10th, 2007, from <http://www.usda.gov/nass/nassinfo/remotehistory.htm>
- Andersen, P., Pandya-Lorch, R., & Rosegrant, M. W. (1999). *World Food Prospects: Critical Issues for the EarlyTwenty-First Century*. Washington, D.C.
- Aronoff, S. (2005). *Remote Sensing for GIS Managers*. Redlands, CA: ESRI Press.
- Badhwar, G. D. (1984). Automatic corn-soybean classification using Landsat MSS data, near harvest crop proportion estimation. *Remote Sensing of Environment* (14), 15-19. Abstract obtained from <http://www.sciencedirect.com/>
- Bauer, M. E., Cipra, J. E., Anuta, P. E., & Etheridge, J. B. (1979). Identification and area estimation of agricultural crops by computer classification of Landsat MSS data. *Remote Sensing of Environment* (8), 77-92. Abstract obtained from <http://md1.csa.com/>
- Belgian Science Policy - EO Edu. (2007). Observing our planet. Retrieved May 14th, 2007, from <http://telsat.belspo.be/bEO/en/satellites>
- Bleasdale, J. K. A., & Nelder, J. A. (1960). Plant population and crop yield. *Nature*, 188 (4747), 342. Abstract obtained from <http://www.nature.com/nature/journal/v188/n4747/abs/188342a0.html>
- Bolsa Nacional Agropecuaria. (2007). Nuestra Empresa [Our company]. Retrieved February 17th, 2007, from <http://www.bna.com.co/LaBNA/NuestraEmpresa/tabid/99/Default.aspx>
- Bouman, B. A. M. (1995). Crop modelling and remote sensing for yield prediction. *Netherlands Journal of Agricultural Science* (43), 143-161. Wageningen, Netherlands.
- Brown, R.J. & Ellis, J. (2000) *The Potential of RADARSAT-2 for Crop Mapping and Assessing Crop Condition*. Paper presented at the Second International Conference on Geospatial Information in Agriculture & Forestry, Lake Buena Vista, Florida.
- Buheasie, R., Tsuchiya, K., Kaneko, M., & Sung, S. J. (2003). Comparison of image data acquired with AVHRR, MODIS, ETM super+ and ASTER over Hokkaido, Japan. *Advances in Space Research*, 32(11), 2211-2216. Abstract obtained from <http://md1.csa.com/>
- Companhia Nacional de Abastecimento CONAB. (2007). ConabWeb - Geosafra. Retrieved September 9th, 2007, from <http://www.conab.gov.br/conabweb/index.php?PAG=81>
- Czajkowski, K. P., & Grunwald, S. (2000). *Final report: Water quality simulation modeling using SWAT in the Sandusky Watershed*. Retrieved in August 10th, 2007

from: <http://www.epa.state.oh.us/oleo/Grant/freports/largefinals/2000/lepf00-10a.pdf>

- Doraiswamy, P. (1997). Land Cover Classification in the SGP97 Region. Retrieved February 17th, 2007, from <http://disc.gsfc.nasa.gov/fieldexp/SGP97/oklarev.html>
- ESRI. (2003). Agriculture Data Model (Vol. 2007). Retrieved March 14th, 2007 from <http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&mid=35>
- ESRI. (2006). ArcGIS Desktop Help (Version 9.2). Redlands, CA.
- ESRI. (2007a). *Building Geodatabases. Course Lectures and Exercises* (Vol. 2007). Redlands, CA: ESRI.
- ESRI. (2007b). *Geodatabase design concepts. Course Lectures and Exercises* (Vol. 2007). Redlands, CA: ESRI.
- ESRI. (2007c). *Working with ArcGIS Spatial Analyst* (Vol. 2007). Redlands, CA: ESRI.
- European Commission Joint Research Centre. (2007). Digital Atlas. Retrieved September 24th, 2007 from <http://dma.jrc.it/services/dmaexplorer/>
- Fisette, T., Chenier, R., Maloley, M., Gasser, P., Huffman, T., White, L., et al. (2006). *Methodology for a Canadian agricultural land cover classification*. Paper presented at the 1st International Conference on Object-based Image Analysis, Ottawa.
- GeoEye. (2007). ImageSearch Application. Retrieved September 24th, 2007 from <http://imagesearch.geoeye.com/>
- Gommes, R. (2003). *Non-parametric yield forecasting*. Paper presented at the Seminar on Tropical Agroclimatology, Bogota, Colombia.
- Grilli, M. P., Ravelo, A., Santa, A., & Zanvetto, R. (2006). *Estimation of the cultivated areas of soybean, maize and wheat using remote sensing techniques*. Paper presented at the 2nd International workshop on crop monitoring and forecasting in South America, Montevideo, Uruguay.
- Hanuschak, G., & Delince, J. (2004). Utilization of remotely sensed data and Geographic Information Systems (GIS) for agricultural statistics in the United States and the European Union [Electronic Version]. *UNECE: United Nations Economic Commission for Europe*, from <http://www.unece.org/stats/documents/ces/ac.61/2004/wp.38.e.pdf>
- Hayes, J. T., O'Rourke, P. A., Terjung, W. E., & Todhunter, P. E. (1982). Yield: A numerical crop yield model of irrigated and rainfed agriculture. *Publications on Climatology*, 35.
- IDEAM (2007a). Estaciones meteorológicas de Colombia. Retrieved April 23th, 2007, from <http://comercio.ideam.gov.co/website/estaciones/viewer.htm>
- IDEAM. (2007b). Venta de información meteorológica. Retrieved April 23th, 2007, from <http://intranet.ideam.gov.co/vinfmteto/Mventaig.asp>

- Kontoes, C., Wilkinson, G. G., Burrill, A., Goffredo, S., & Mégier, J. (1993). An experimental system for the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture. *International Journal of Geographical Information Science*, 7(3), 247-262.
- Leica Geosystems. (2005a). *ERDAS Field Guide*.
- Leica Geosystems. (2005b). *ERDAS IMAGINE Tour Guides*.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic Information Systems and Science*. Barcelona, Spain: John Wiley & Sons Ltd.
- MDA. (2007). MDA Information products. Retrieved May 27th, 2008 from <http://gs.mdacorporation.com/rsic/agriculture/croptype.asp>
- Maas, S. J. (1988). Using satellite data to improve model estimates of crop yield. *Agronomy Journal*, 80(4), 655-662.
- Maxwell, S. K., Nuckols, J. R., Ward, M. H., & Hoffer, R. M. (2004). An automated approach to mapping corn from Landsat Imagery. *Computers and Electronics in Agriculture* (43), 43-54.
- Ministerio de Agricultura. (2004). *El agro colombiano frente al TLC con los Estados Unidos [Colombian Agriculture and the United States Foreign Trade Agreement]*. Bogotá, Colombia.
- Ministerio de Agricultura. (2007). Agronet. Retrieved March 22nd, 2007 from <http://www.agronet.gov.co/>
- Ministerio de Agricultura, Departamento Administrativo Nacional de Estadística, Federación Nacional de Arroceros Fedearroz, & Distrito de Riego Usocoello. (2001). *Aspectos Agrícolas y pecuarios*. Bogotá, Colombia.
- Ministerio de Agricultura, & IICA. (2006). *Cuarto informe de Coyuntura: Maíz*. Bogota, Colombia.
- NASA. (2005). Terralook from Jet Propulsion Laboratory. Retrieved November 18th, 2007, from (http://asterweb.jpl.nasa.gov/TerraLook_download.asp)
- NASA. (2006). MODIS Website. Retrieved November 18th, 2007, from <http://modis.gsfc.nasa.gov/about/components.php>
- NASA. (2007). University of Maryland - Earth Science Data Interface (ESDI) - Global Land Cover Facility (GLCF). Retrieved May 28th, 2007, from <http://glcfapp.umiacs.umd.edu/index.shtml>
- Pinter, P. J., Jr., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S. T., et al. (2003). Remote sensing for crop management. *Photogrammetric Engineering & Remote Sensing*, 69(6), 647-664.
- Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26-33.
- Priya, S., & Shibasaki, R. (2001). National spatial crop yield simulation using GIS-based crop production model. *Ecological Modeling* 136(2-3), 113-129.

- Procalculo Prosis S.A. (2006). *Reunión de Planeación Estratégica enero 2006*.
- Procalculo Prosis S.A. (2007). Nosotros [About us]. Retrieved February 17th, 2007, from <http://www.procalculoproisis.com>
- Ravelo, A., Casa, A. D. I., Ovando, G., & Planchuelo, A. M. (2006). *Modeling soybean and corn yields using ground and remote sensing information in central Argentina*. Paper presented at the 2nd International workshop on crop monitoring and forecasting in South America, Montevideo, Uruguay.
- Ravelo, A., Rojas, O., Negre, T., Cherlet, M., & Planchuelo, A. M. (2002). *Taller en monitoreo de cultivo y pronóstico de rendimientos*. Paper presented at the Workshop on crop monitoring and forecasting Cordoba, Argentina.
- Reichert, G. C., & Caissy, D. (2002). *A reliable crop-condition assessment (CCAP) incorporating NOAA AVHRR data, a geographic information system, and the Internet*. Paper presented at the Proceedings of the 2002 ESRI User Conference, Redlands, CA.
- Rogerson, P. (2006). *Statistical methods for geography*. (Second ed.). Sage.
- Rojas, O., Delince, J., & Leo, O. (2006). *Crop yield monitoring in eastern Africa: Bulletin for rain-fed maize and sorghum yield prospects in 2006*. Ispra, Italy: Joint Research Centre - European Commission.
- Sadler, E. J., & Russell, G. (1995). Modeling Crop Yield for site-specific management. In F. J. Pierce & E. J. Sadler (Eds.), *The state of site specific management for agriculture*. Madison, WI: American Society of Agronomy: Crop Science Society of America: Soil Science Society of America.
- Sarewitz, D., Pielke, R. A. J., & Byerly, R. J. (2000). Prediction in Science and Nature; Decision Making and the future of Nature: Understanding and Using Predictions. In D. Sarewitz, R. A. J. Pielke & R. J. Byerly (Eds.), *Prediction: Science, decision making and the future of nature*. Washington, D.C.: Island Press.
- Servilla, M., & Towner, M. (2000). *Integrating high-resolution satellite imagery and weather data for improved agricultural management decisions*. Paper presented at the ESRI User Conference, San Diego, CA.
- Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., et al. (2001). Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, 93(3), 583-589.
- Silapathong, C., & Blasco, F. (1992). The application of geographic information system to mangrove forest management: Khlung, Thailand. *Asian Pacific Remote Sensing Journal*, 5(1), 97-104.
- Spot Image. (2007). Solutions. Retrieved May 23, 2007, 2007, from <http://www.spotimage.fr/web/en/194-solutions.php>
- Steffen, K. (2007). Applications to Biosphere [Electronic Version]. *Cooperative Institute for Research in Environmental Sciences*, Retrieved July 25th, 2007 from <http://cires.colorado.edu/steffen/classes/geog5093/Biosphere.pdf>

- Stewart, T. R. (2000). Uncertainty, Judgment, and Error in Prediction. In D. Sarewitz, R. A. J. Pielke & R. J. Byerly (Eds.), *Prediction, science, decision making and the future of Nature*. Washington, D.C.: Island Press.
- Syam, T., & Jusoff, K. (1999). *Remote Sensing (RS) and Geographic Information system (GIS) Technology for Field Implementation in Malaysian Agriculture*. Paper presented at the Seminar on Repositioning Agriculture Industry in the Next Millennium, Serdang, Selangor Malaysia. Retrieved March 20th, 2007 from <http://www.econ.upm.edu.my/~peta/book1/chapter13.pdf>.
- Towson University. (2007). Center for GIS. Chesapeake Bay & Mid-Atlantic from Space. Retrieved November 14th, 2007, from <http://chesapeake.towson.edu/data/>
- Turker, M., & Arikan, M. (2005). Sequential masking classification of multi-temporal Landsat7 ETM+ images for field-based crop mapping in Karacabey, Turkey. *International Journal of Remote Sensing*, 26(17), 3813 - 3830.
- USGS. (2007). Satellite Images of Environmental Change - EarthShots. Retrieved May 19th, 2007, from <http://earthshots.usgs.gov/Help-GardenCity/beyondimages>
- United Nations. (2006). *Colombia Coca Cultivation Survey 2005*. Bogota, Colombia: Office for drugs and crime.
- University of Maryland. (2007b). Global Land Cover Facility (GLCF) - Data Formats. Retrieved November 2nd, 2007, from <http://glcf.umiacs.umd.edu/esdi2-help/dataFormats.html>
- Viloria de la Hoz, J. (2004). *La economía del departamento de Córdoba "Cordoba Department's Economy"*. Cartagena, Colombia: Centro de Estudios Regionales (CEER) - Banco de la República.
- Williams, J. R. (1990). The Erosion-Productivity Impact Calculator (EPIC) Model: A Case History. *Philosophical Transactions: Biological Sciences*, 329, 421-428.
- World Bank Group. (2007). Agriculture Investment Sourcebook. Retrieved February 14th, 2007, from <http://www-esd.worldbank.org/ais/index.cfm?Page=mdisp&m=06&p=8>
- World Meteorological Organization (WMO). (1982). The effect of meteorological factors and crop yields and methods of forecasting the yield. In *Technical Note no. 174* (pp. 54). Geneva: World Meteorological Organization.
- Yang, C., & Anderson, G. L. (2000). Mapping grain sorghum yield variability using airborne digital videography. *Precision Agriculture* (2), 7-23.