DISSERTATION

SPATIAL MODELING OF SOIL SALINITY USING REMOTE SENSING, GIS, AND FIELD DATA

Submitted by

Ahmed Aly Mohamed Eldeiry

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

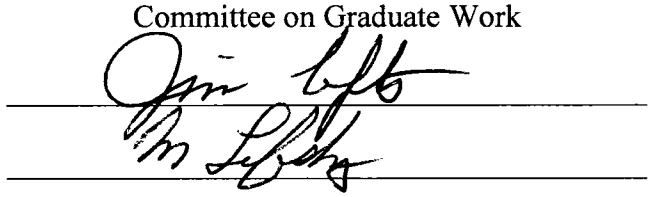
Fall 2006

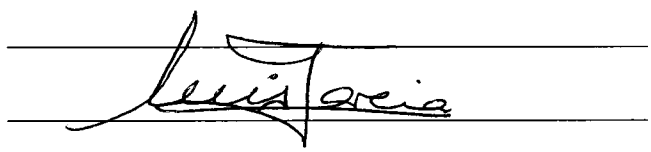
COLORADO STATE UNIVERSITY

November 7, 2006

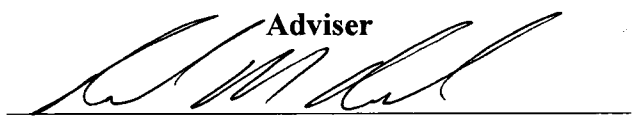
WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY AHMED ALY MOHAMED ELDEIRY ENTITLED SPATIAL MODELING OF SOIL SALINITY USING REMOTE SENSING, GIS, AND FIELD DATA BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

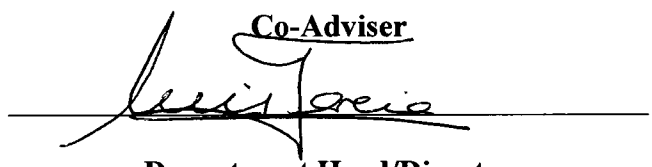




Adviser



Co-Adviser



Department Head/Director

ABSTRACT

SPATIAL MODELING OF SOIL SALINITY USING REMOTE SENSING, GIS AND FIELD DATA

In this study a new methodology was developed to generate accurate predicted soil salinity maps using remote sensing data. The techniques used include integrating field data, geographic information systems (GIS), remote sensing, and spatial modelling techniques. Corn and alfalfa crops were selected as indicators of soil salinity during 2001 and 2004 respectively. Five images were acquired from Aster, Ikonos, and Landsat to check the correlation between measured soil salinity and remote sensing data. Observed data from four corn fields during 2001 and four alfalfa fields during 2004 were used in conjunction with the Aster, Ikonos, and Landsat images. Three subsets of 75%, 50%, and 25% were randomly selected from each main set of observed data to be used in conjunction with the Ikonos and Landsat images.

Three models were applied to predict soil salinity from remote sensing: the ordinary least squares model (OLS), spatial autoregressive model (SAR), and modified kriging model. The combination of satellite imagery bands that had the best correlation with measured soil salinity was used to predict soil salinity. A number of criteria were used to select the best model. The results show that the modified kriging model provides the best results over the OLS and the SAR models. The OLS model meets the model selection criteria, but, in most cases, it involves some autocorrelation among the residuals. The SAR model was able to remove some of the autocorrelation

among the residuals, but the R^2 was reduced. The R^2 values of the OLS model were 0.34, 0.47, 0.52, 0.26, and 0.37 for the 2001 Aster, Landsat, Ikonos images for corn, the 2004 Landsat and Ikonos image for alfalfa respectively. The R^2 values of the SAR model were 0.05, 0.18, 0.25, 0.03, and 0.15 for the same images. The R^2 values of the modified kriging model were 0.81, 0.83, 0.91, 0.60 and 0.68 for the same images. Also, the mean absolute error (MAE) improved significantly when using modified kriging over the OLS and SAR models for all data sets. When the modified kriging model was applied to the subsets of data it showed encouraging results as well.

Ahmed Aly Mohamed Eldeiry
Civil and Environmental Engineering Department
Colorado State University
Fort Collins, CO 80523
Fall 2006

ACKNOWLEDGEMENTS

First, I would like to thank my academic advisor, Dr. Luis Garcia; my co-advisor, Dr. Robin Reich; committee member, Dr. Jim Loftis; and outside committee member, Dr. Michael Lefsky. Special thanks to Dr. Garcia for his contributions and efforts throughout the entire research. His involvement on a personal, scholarly, financial and professional basis has been a blessing. His dedication, motivational ideas, words of encouragement, commitment to excellence, constructive critiques, and tolerance made him both an excellent advisor and a wonderful friend. I am forever indebted to Dr. Robin Reich for his frequent and candid input, and his unselfish support.

I would like to express my deep gratitude and appreciation to the members of my graduate committee, Dr. Jim Loftis and Dr. Michael Lefsky, for their valuable comments and suggestions during the course of the research and for all the time they devoted to helping me complete this research.

Thanks to Leslie Patterson at the Integrated Decision Support (IDS) group for her help in editing this manuscript and to David Patterson for his assistance on the technical and computer aspects of the research. I am also very lucky to have had the support of my son Mohamed in making the automatic indices, sorting the figures, and finalizing some of the unique flowcharts and maps.

I would like to thank all the members of the IDS group for providing a healthy and pleasant environment throughout the years of this research.

Many thanks to the Drainage Research Institute staff in Egypt: Dr. Safwat Abdel Dayem, Dr. Shaden Tawfeeq, Dr. Hesham Kandeel, Dr. Hosam Fahmy, and Jim Krider; they have planted and provided care for the seed from which my Ph.D. blossomed.

Last but not least, my unparalleled gratefulness and appreciation goes towards the one person who was always by my side, my wife Manal, and also to my wonderful kids: Mohamed, Asmaa, Israa, Alaa, Ali; and My brother Mahmoud.

To my blessed and wonderful parents and brother Mohamed who are no longer with us, may their souls rest in peace.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1 Diagnosis of soil salinity problem

Soil salinity is a severe environmental hazard which increasingly impacts crop yields and agricultural production. Soil salinity refers to the presence in soil and water of various electrolytic mineral solutes in concentrations that are harmful to many agricultural crops (Hillel 2000). Natural salinization or primary salinization results from the long-term influence of natural processes. In contrast, human-induced salinization or secondary salinization is the result of salt stored in the soil profile being mobilized by extra water provided by human activities such as irrigation (Szabolcs 1989).

Statistics relating to the extent of salt-affected areas vary but estimates are that close to 1 billion hectares or about 7% of the earth's landscape, are affected by primary salinity, while about 77 million hectares have been salinized as a consequence of human activities, with 58% of these concentrated in irrigated areas (Ghassemi et al. 1995). On average, 20% of the world's irrigated lands are affected by salts, but this figure increases to more than 30% in countries such as Egypt, Iran and Argentina (Ghassemi et al. 1995). The development of saline soils is a dynamic phenomenon, which needs to be monitored regularly in order to secure up-to-date knowledge of their extent, spatial distribution, nature and magnitude.

1.2 Potential of remote sensing in assessing soil salinity

Remote sensing data has great potential for monitoring dynamic processes, including salinization. Remote sensing of surface features using aerial photography,

videography, infrared thermometry and multispectral scanners has been used intensively to identify and map salt-affected areas (Robbins and Wiegand 1990). Multispectral data acquired from platforms such as Landsat, SPOT, and the Indian Remote Sensing (IRS) series of satellites have been found to be useful in detecting, mapping and monitoring salt affected soils (Dwivedi and Rao 1992).

However, the digital analysis of multispectral data using the spectral response pattern of salt-affected soils is plagued by misclassification. In order to improve the detectability of these soils and other natural features using remote sensing data, various image transforms have been developed. These transforms not only enhance the detectability of these features, but also aid data compression resulting in substantially reduced computational time and cost. Salinity hazard mapping has included salt load studies, trend-based methodologies, strongly inverse methods, composite index methods and integrated geoscience approaches (Lawrie et al., 2000, 2003; Spies and Woodgate, 2004). Composite Index approaches have been further developed for salinity risk assessments (Clifton and Heislers, 2004).

Principal component (PC) is one such transformation. Essentially, PC undertakes a linear transformation of a set of image bands to create a new band set with images that are uncorrelated and ordered in terms of the amount of variance explained in the original data (Johnston 1980, Mather 1987). This technique has been most commonly used as a procedure for data compression by discarding minor components with little explanatory value. It has been shown that when the data image set consists of a single variable time series of environmental data, the first standardized component indicates the characteristic value of that variable, while the second and all remaining standardized components represent change elements of successively decreasing magnitude (Eastman 1992).

Ratioing, on the other hand, has been used to remove redundancy in remote sensing data while retaining the information desired. In addition, ratioing tends to reduce the effects of atmospheric degradation and shadowing in hilly and rugged terrain. Ratioing techniques are popular because they can be performed digitally as well as photographically. Various transformations, such as perpendicular vegetation index (Richardson and Weigand 1977), transformed vegetation index (Rouse et al. 1973), and normalized difference vegetation index (NDVI), have been used to exploit information on vegetation. The NDVI generated from satellite data has been used to differentiate various types of salt-affected soils, including potential saline soils (Toth et al. 1991). Fakuhara et al. (1979) introduced a soil index (SI) to eliminate the effect of vegetation canopy reflectance in a partially vegetated terrain. Generally, the transformed values of this index are very large because arctangent is used to compress the ratio value.

For mapping surface land salinity, color, and thermal infrared aerial photography and spectral image interpretation techniques such as satellite (Landsat TM, SPOT), and other airborne remote sensing techniques are used (George et al., 2003; Spies and Woodgate, 2004). Other techniques, such as gamma radiometrics (Wilford et al., 2001), are useful for mapping soils and shallow sub-soil materials that can assist with interpretation of likely recharge and discharge areas. Digital elevation models (DEMs) that use current topography to predict salt accumulation and salinity discharge zones (McFarlane et al., 2004) are only sometimes useful by themselves in providing reliable salinity predictions (Lawrie et al., 2004).

1.3 Integration of remote sensing, soil salinity, GIS and spatial modeling

Farifteh et al. (2005) suggested that an integrated approach using techniques and facilities the use of remote sensing, in particular image spectrometry, solute transport modeling and advanced geophysical survey offer to timely detect salinization must be introduced. The integration of remotely sensed data, Geographic Information Systems (GIS), and spatial statistics provides useful tools for modeling large-scale variability to predict the distribution, presence, and pattern of exotic and native plant species as well as soil characteristics (Kalkhan et al. 2000). This integration also provides tools for assessing the landscape-scale structure of forest and rangelands (Kalkhan et al. 2000, 2002; Chong et al. 2001). Researchers using this integrated method are concerned with reducing the cost of field sampling by considering sampling design, sample size, and plot size. There is an obvious trade-off between plot size and sample size. Researchers can select smaller plots with more replicates, or larger plots with fewer replicates for the same cost. Large sample size can provide better estimates of the variable of interest; however, collecting field data can be time consuming and self-limiting (Scheaffer et al. 1990; Olsen and Schreuder 1997).

Triantafilis et al. (2006) mentioned that the results of spatial prediction suggest that regardless of what method was used, reasonable estimates of soil salinity were achieved. Strong statistical tools for measuring autocorrelation available are Moran's I (Moran 1948), and the spatial cross-correlation statistic (Bonham et al. 1995; Reich et al. 1994). These statistical tools are especially useful in studies of patchily distributed resources in time and space as determinants of vegetation patterns (Reich et al., 1994). Spatial auto-correlation and cross-correlation statistics can be used to evaluate spatially explicit information on vegetation characteristics and structures and

associated environmental characteristics. Earlier research in the area of spatial statistics led to the development of a multivariate spatial correlation statistic by Wartenberg (1985) based on a mantle type coefficient to quantify the spatial relationships among univariate data (Reich et al., 1994). The diagonal elements of Wartenberg's (1985) multivariate spatial correlation matrix are themselves Moran's I statistic, while the off-diagonal elements are cross-correlation coefficients. Czaplewski and Reich (1993) showed that the off-diagonal elements can be interpreted as a weighted coefficient between any two variables. This creates a way to assess the null hypothesis that no spatial cross-correlation exists between the variables in space.

The ability to model the small-scale variability in landscape characteristics requires the generation of full-coverage maps depicting characteristics measured in the field (Goward et al., 1994). Goward et al. (1994) point out that while many spatial data sets describing land characteristics have proven reliable for macro-scale ecological monitoring; these relatively coarse-scale data fall short in providing the precision required by more refined ecosystem resource models.

1.4 The problem of soil salinity in the Arkansas River

The Arkansas River is one of the most saline rivers of its size in the United States. Salinity levels, measured as dissolved solid concentrations, increase from 300 mg/L near Pueblo to over 4,000 mg/L at the Colorado-Kansas border (Ghassemi et al. 1995). According to data observed between April 1990 and March 1993, concentrations increased 183%, or about 30 mg/L per mile, from Las Animas on the western edge of John Martin Reservoir to Coolidge, Kansas. The water of the Arkansas River is used to irrigate crops in the river valley, and farmers are facing

decreasing crop yields due in part to these high levels of salinity. In some areas, land is being taken out of production due to unsustainable crop yields.

Based on data observed over a 53,100 ha sub region of the Lower Arkansas Valley, Gates et al. (2002) stated that a shallow water table with average salinity concentration of 3,100 mg/L spreads out under the land at an average depth of 2.1 m below the ground surface. In addition, Gates et al. (2002) mentioned that the crop yield reduction due to salinization is estimated to be 0 to 75% on fields spread over the study area with a total revenue loss ranging from $0-$750/ha based on 1999 crop prices. Burkhalter et al. (2005) mentioned that over three irrigation seasons, average seasonal aquifer recharge from irrigated fields in a 50,600 ha study area ranges from 0.59 to 0.99 m, including contribution from precipitation. The salinity of irrigation water varies from 618 to 1,090 mg/L. The water table is shallow, with 16 to 33% of irrigated land underlaid by an average water table less than 2 m deep. Average water table salinity ranges from 2,680 to 3,015 mg/L, and average soil salinity from 2,490 to 3,860 mg/L. Crop yield reductions from salinity and waterlogging range from 0 to 89% on fields, with regional averages ranging from 11 to 19%. Annual salt loading to the river from subsurface return flows, generated in large part by dissolution from irrigation recharge, averages about 533 kg/ irrigated ha per km.

1.5 The approach of this research

The approach presented in this research involves integrating remotely sensed, GIS, and field observed data of soil salinity to evaluate what are the most appropriate spatial techniques is to develop high quality soil salinity maps. The approach is being tested on soil salinity data observed in the Lower Arkansas River Valley near the Kansas border in Colorado. Most satellite imagery bands do not have the ability to

penetrate through the soil profile. Detecting soil salinity from satellite images with bare soil is difficult. In winter, snow can skew results as can crop residue left on the ground after the growing season is over. However, during the growing season, most cultivated crops reflect the salinity status of the soil profile through their yields. Crop biomass can provide an indication of the salinity in the plants root zone, and the results of this evaluation appear on the ground through reduced crop biomass and vegetation. Therefore, the correlation between soil salinity data and the satellite images will be based on the reflection of the crop cover type as an indicator of soil salinity. For this study, alfalfa and corn were selected as soil salinity indicators since they were the dominant crops in the study area and the monitored fields.

Four satellite images were acquired, two Ikonos and two Landsat to develop the correlation between soil salinity and remote sensing data. For both 2001 and 2004 a pair of Ikonos and Landsat images was brought as close in time as possible. In addition to the main sets of observed data, two subsets 20 and 12 observed points were randomly selected form the main sets to test for correlation with images. The two subsets were used in conjunction with satellite images as a way to rely on remote sensing and reduce field observations.

Three statistical regression techniques were applied to test for cross correlation between soil salinity data and satellite images. First, the ordinary least squares (OLS) model was applied to the soil salinity data and the satellite images. If the OLS model involves autocorrelation among the residuals, the other techniques are used. Spatial autocorrelation causes a false indication of the cross correlation between the soil salinity data and the satellite images and makes this cross correlation seem stronger than it actually is. The spatial autoregressive (SAR) model and the modified kriging model are applied to the images to remove the autocorrelation among the residuals.

The best model for each satellite image is selected based on some pre-specified criteria. The selection criteria is the following: 1) high multiple correlation coefficient R^2; 2) p-value should be < 0.05 to guarantee that there is a strong cross correlation between soil salinity and the image; 3) Moran's I p-value should be > 0.05 to guarantee that there is no autocorrelation among the residuals; 4) Lagrange multiplier *p*-value should be < 0.05 as another test to guarantee that there is no autocorrelation among the residuals; and 5) the AICC of the model selected should have the smallest value among the tested models. Having selected the best model, the following are used to validate the model: 1) cross validation; 2) the model selected is applied to the other sets of data; 3) finally, the predicted and observed soil salinity data is compared to check the accuracy of the model.

1.6 The contribution of this research

This research presents a new methodology by using statistical modelling techniques to predict soil salinity from remote sensing data. It utilizes the integration of GIS, GPS, remote sensing data and geospatial models along with field data to develop precise and accurate soil salinity maps. In addition to applying the methodology to the main set of observed data, the methodology was also applied to two subsets where 20 observed points and 12 observed points were selected randomly. This was a way to reduce the field work and take advantage of remote sensing data. In this research, a lot of effort was devoted to investigating the autocorrelation in the residuals. If autocorrelation exists among the residuals, the results might give a false indication of a stronger cross correlation than actually is present and causes the R^2 value to be higher than it should be. Residuals were neglected to some degree in most of the previous research. Previous studies mainly

focused on using individual techniques and variables to detect soil salinity such as best band combinations, vegetation indices, and principal component analysis. Also, in most previous studies there were not enough statistical measures to guarantee high accuracy in the predicted soil salinity maps. In this study, the observed soil salinity was integrated with the best combination of bands using a spatial modelling technique to predict soil salinity.

CHAPTER 2. LITERATURE REVIEW

2.1 Soil salinity as a worldwide problem

Saline landscapes differ from other environments in that they are subject to the stresses of salinity, waterlogging and/or inundation (Barrett-Lennard et al., 2003). Salt decreases the availability of water to plants ("osmotic effects") and have direct adverse effects on metabolism (Greenway and Munns, 1980). Waterlogging decreases the availability of oxygen to roots and makes roots energy deficient. It also interacts with salinity to increase the transport of salt to shoots, impairing plant growth and survival (Barrett-Lennard, 2003). Inundation covers shoots with water, decreasing gaseous exchange between leaves and the atmosphere. Each of these stresses has the capacity to compromise plant growth and survival. Their interaction therefore creates a range of niches for highly specialised plants (Galloway and Davidson 1993; Short and Colmer, 1999; Garthwaite et al., 2003).

Previous research has established that a number of key spatially explicit biophysical datasets are required to map the distribution of salt in the landscape and place it in a landscape context, to understand the processes leading to salinization, to produce maps of salinity hazard, and predict the risk of salt mobility to land, river and groundwater (George and Woodgate 2002; George et al., 2003). Chamberlain and Wilkinson (2004) identified areas in danger of salinization if current practices are maintained. Appropriate responses to salinity are likely to vary in response to biological, physical, economic and social circumstances (Ridley and Pannell 2005). Restructuring within the agricultural industry is leading to business consolidation and

more profitable farm businesses, a major benefit sought from science is likely to be improved income (Pannell 2001). Barr (2003) has demonstrated that this situation is most commonly encountered in the traditional farming regions well removed from regional cities and services. Farmers in this situation are most unlikely to adopt new farming systems that are not commercially attractive. Nearer to regional centers and particularly in the higher rainfall regions, off-farm income is an increasingly significant part of total family income and in many instances the major component (Gleeson et al., 2002). Landholders are by no means a homogeneous group (Vanclay 1997). Aside from obvious differences due to industry, location and size of enterprise, annual surveys by the Australian Bureau of Agricultural and Resource Economics (Nelson et al., 2004) reveal that farmers have a range of major priorities not unlike the range found in any other social grouping. Salinization is a growing problem throughout the world and poses a global threat to crops and to land use. Mineral salts naturally occur dissolved in rain, rivers, and groundwater and are bound up in soil particles as well. These mineral salts include sodium, calcium, magnesium and potassium chlorides, sulfates and carbonates. When farmers irrigate their crops, salts in the irrigation water get deposited in the soil. Even good quality water typically has salt concentrations of 200 – 500 parts per million, the higher figure being the U.S. government's recommended limit for drinking water. If a farmer annually applies 1 meter of irrigation water to a hectare of land (10,000 tons of irrigated water), between 2 and 5 tons of salt will be added to that land every year (Postel 1999). Worldwide, one in five hectares of irrigated land suffers from a build-up of salts in the soil. Soil salinization costs the world's farmers an estimated $11 billion a year and this figure is growing. Salinization problems continue to spread at a rate of up to 2 million hectares

a year. Soil salinity is offsetting a good portion of the increased productivity achieved by expanding irrigation (Postel 1999).

The environmental damage that salinity can cause is massive. In its early stages, soil salinization reduces soil productivity, but in advanced stages, salinization kills all vegetation and transforms fertile and productive land to barren waste. This leads to loss of habitat and a reduction of biodiversity (Ghassemi et al., 1995). Environmental damage also occurs when saline water from drainage schemes or groundwater is accumulated in off-stream floodplain areas or wetlands and either is left to evaporate naturally or discharged when high flow rates in the main stream prevail (Williams 1987). When saline water is accumulated in floodplain areas or wetlands and left to evaporate, damage occurs directly in the areas inundated with saline water. This is the situation with the Kesterson Reservoir and National Wildlife Refuge in California, where high levels of potentially toxic trace elements, particularly selenium, were in drainage waters and caused deformities in waterfowl (Tanji et al. 1986). When saline water is discharged, damage may occur as pulses of highly saline water are allowed to flush downstream. As salinity degrades vegetation communities, such as wetlands and grasslands, it adds to the threats facing wildlife. Animals rely on those plant communities for their survival. Also, all native freshwater fish are susceptible to change in salt concentration (Ghassemi et al., 1995).

Ghassemi et al. (1995) mentioned that the economic damage that salinity can cause is severe. He mentioned that in Pakistan, a study conducted by the Water Resources and Planning Division on the entire irrigated area of the Upper Indus Plain looked at the economic impact of soil salinity. The area comprises the Punjab and Northwest Frontier. The study concluded that the economy of the two provinces suffer a loss of approximately 4.3 billion rupees (US$300 million) annually from

decreased farm production on soils which are slightly to moderately affected by salinity. In Australia, it has been estimated that annual agricultural losses from salinization in the Murray-Darling Basin amount to A$260 million (US$208 million). In the Colorado River Basin of the United States, the heavy annual salt load of about 8.2 megatons is costing water users more than US$750 million per year. According to Reisner (1986), it is estimated that each additional mg L^{-1} of salt in the water supply systems of the cities using Colorado River water causes US $300,000 worth of damage per year to the objects coming in contact with the water (pipes, fixtures, machinery, etc). In the Republic of South Africa, Heynike (1981) estimated an annual economic loss of US $83 million due to salinity. In Egypt, the government is spending more than LE70 million (about US$30 million) annually on drainage to overcome the problem of waterlogging and salinization in the Nile Valley and its delta (Amer and de Ridder 1989). On a global scale, in 1990 it was estimated that the loss in production capacity due to all processes of land degradation was about US $42.2 billion (Dregne et al. 1991).

The social cost of salinization is not easy to quantify. Salinization causes occupational or geographic shifting of the farm population and reduction in aggregate income and expenditure. These events have social and economic repercussions on the region as a whole, including dependent country towns. The impacts are most apparent in rural areas and small towns because the opportunities for adjustment of the local economic base are more limited (Peck et al., 1983). It is perhaps not surprising that a decade of research coordinated by Australia's National Dryland Salinity Program (NDSP) has revealed that the more we know about salinity the more we realise we don't know (van Bueren and Price 2004). As just one example, until the early 1990s trees were widely seen to be the almost universal solution to dryland salinity, but as

Passioura (2004) notes "it did not take long for trees not to be the answer", while research has provided many answers, there is still much to be done if we are to confidently identify the best management strategies for salinity in diverse situations. But while more research will no doubt lead to better land management options, an equally pressing issue is how we use what we already know.

2.2 Measuring yield affected by salinity

Wiegand et al. (1994) carried out a procedure to assess the extent and severity of soil salinity in fields in terms of economic impact on crop production and effectiveness of reclamation efforts. They developed their procedure based on soil salinity, plant height and boll counts, utilizing digitized color infrared aerial photography and videography acquired during midboll set development stage for four salt-affected cotton fields in the San Joaquin Valley of California. Unsupervised classification procedures were used to produce a seven-category spectral map for each field. Regression equations were developed from salinity measurements in the surface 30 cm at 100-200 sample sites per field along with the photography and videography digital counts at those same sites. The equations were used to estimate the salinity of each of the approximately 100,000 pixels per field, and the salinity categories corresponding to the spectral ones were mapped. The spectral classification maps and the estimated salinity maps correlated well. Boll counts, made at about 20 sites per field, were converted to lint yield and regressed on NDVI from both the photography and videography; the correlation coefficient (r) was 0.72 for video and 0.73 for the photographic data. Their results illustrate practical ways to combine image analysis capability, spectral observations, and ground truth to map and quantify the severity of soil salinity and its effects on crops.

Golovina et al. (1992) made an effort to automate methods of air photo interpretation in order to speed up and make the interpretation process more objective when compiling maps of soil salinization. Their basic approach involved the determination of spectral brightness coefficients and image photodensities for areas known to have specific characteristics with respect to relative areas of healthy and dead cotton plants and soil salinity. They used these relationships in algorithms to discriminate saline, nonsaline, and mixed (saline/nonsaline) areas. They were able to identify spots where cotton plants were dead by means of the threshold of image phototone (tone density). They stated that automated analysis methods are capable of performing a more complex evaluation of the quantitative degree of salinization, which is difficult to achieve in visual interpretation.

2.3 Estimating groundwater depth and groundwater salinity

Srivastava et al. (1997) studied the accuracy of mapping shallow groundwater depth and salinity using remote sensing data. At their study area, groundwater quality had been mapped on the basis of field sampling and chemical analysis. Then, they made a scheme of image processing and GIS techniques using false color, vegetation indices, density slicing, overlaying, and supervised classification and applied it to IRS-1B LISSS II data. They found that the groundwater salinity zones which they mapped in their remote sensing data research are in close agreement with that which was done in the earlier research. They established a regular variation between Green Vegetation Index (GVI) on one hand and water table depth and chloride concentration on the other hand. Thus, in the study, groundwater depth and salinity maps were based on reflectance variations of vegetation above the ground surface. The study

asserts that the species of vegetation found in an area and their densities can provide evidence of shallow groundwater conditions.

2.4 Remote sensing and soil salinity

Howai (2003) stated that remote sensing has a potential application for rapid and large scale mapping of salt-affected lands. They presented some examples of the capability of remote sensing data such as Landsat ETM+, airborne visible/infrared imaging spectrometer (AVIRIS), color infrared aerial photos (CIR), and high-resolution field spectroradiometer (CER 3700) to extract surface information about soil salinity. They used image processing techniques such as supervised classification, spectral extraction, and matching techniques to investigate types and occurrences of salts in the Rio Grande Valley on the United States-Mexico border. Soil salinity groups were established using soil physico-chemical properties and image elements (absorption-reflectivity profiles, band combinations, grey tones of the investigated images, and textures of soil and vegetation covers as they appear in images). The lack of vegetation or scattered vegetation on salt-affected soil surfaces makes it possible to detect salt in several locations of the investigated area.

Fernnandez-Buces et al. (2006) made a study aimed to correlate soil characteristics (electric conductivity in saturation extract (EC) and sodium absorption ratio with the spectral response of plant species and bare soils, integrating an algorithm to allow multi-scale mapping using remote sensors. Ground radiance was measured on different plant species and bare soils. A Combined Spectral Response Index (COSRI) was calculated for bare soils and vegetation by adjusting the normalized difference vegetation index (NDVI). EC, and SAR were determined in

surface soil samples. Correlation coefficients between COSRI and soil salinity were obtained and a model was adjusted to predict soil salinity.

2.4.1 Selection of best band combination

Dwivedi and Rao (1992) conducted research to try to determine the most appropriate three-band combination of Landsat Thematic Mapper (TM) reflective band data for delineating salt-affected soils of the Indo-Gangetic alluvial plain. They used the standard deviation and correlation coefficients value of the TM data for computing a statistical perimeter called the "Optimum Index Factor" (OIF) that is indicative of the information (variance) content of the data. Amongst all the 20 three-band combinations considered, the band combination 1, 3, and 5 was found to be the best in terms of information content. The frequently-used band combination 2, 3, and 4 ranked relatively low. The validation of this conclusion with accuracy estimates of the delineation of salt-affected soils using the same data revealed a mixed relation between the ranking obtained from the OIF values and accuracy estimates, pointing to the need for further investigation in other areas having similar terrain.

Metternicht and Zinck (1997) provided an approach for mapping salt and sodium affected surfaces by combining digital image classification with field observation of soil degradation features and laboratory determinations to map salt and sodium-affected areas in the semiarid valleys of Cochabamba, Bolivia. Using the electrical conductivity and pH values, they established salinity-alkalinity classes. A neighborhood operator, with user-defined spatial and spectral constraints, determined the spectral objects constituting the training set. They found that a combination of six Landsat TM bands (1, 2, 4, 5, 6, and 7) provided the highest separability among salt and sodium-affected soil classes. Although their overall accuracy rate was rather low

at 64 percent, accuracies of 100 percent were obtained for some classes. The main causes of spectral confusion of different salinity-alkalinity degrees were the type and abundance of salt-tolerant vegetation cover, the topsoil textures and the mixture of topsoil properties under field conditions.

2.4.2 Band ratios and principal components analysis

Csillag et al. (1993) conducted research in which they measured the reflectance spectra of salt-affected surface soil samples at 10 nm spectral resolution between 495 nm and 2395 nm. They observed 41 samples in California and 49 in Hungary to compare the effects of different salinization and alkanization processes on reflectance properties. They analysed the reflectance data set statistically using a modified stepwise principal component analysis (MSPCA) approach to select the best bands for classification of salinity.

The modified stepwise principal component analysis (MSPCA) procedure can be efficiently used to select a particular subset of original bands (without combination) to decrease the dimensionality requirements of high-resolution reflectance measurements at the data collection phase. Then, they tested the classification using discriminant function analysis (DFA). Discriminant function analysis (DFA) is more reliable in determining the bands most sensitive to the salinity status of soils than analysis of individual variables. They used principal component analysis (PCA) successfully to identify a linear combination of spectral measurements leading to a smaller number of uncorrelated axes in the measurement space. Csillag et al. (1993) stated that salinity status is a complex phenomenon because variation in the reflectance spectra of soils cannot be attributed to a single soil property such as pH or salt content. Also there is no known narrow absorption band linked to salinity status.

However, using the stepwise regression analysis they found regression coefficients (R^2) of 0.85, 0.78, 0.60, 0.75, 0.73, and 0.32 for clay content, organic matter, organic carbon, pH, salt content, EC, and ESP, respectively with 6-16 10-nm wide bands. Their analysis has identified six intervals between 500 nm and 2400 nm, where the majority of significant narrow bands, that is, 10 nm, 20 nm, and 40 nm, were located. Using 35-42 of these narrow bands led to 80-90% overall recognition accuracy for the entire data set and 100% accuracy for the 41 U.S. samples.

Band ratios of visible to near-infrared and between infrared bands have proven to be better for identifying salts in soils and salt-stressed crops than individual bands (Craig et al., 1998; Hick and Russell, 1990; and Hick et al., 1984). Theoretically, principal components analysis offers an interesting approach to dealing with salt identification and change detection simultaneously. The stable brightness of the first principal component (PC1) and the stable greenness of the second principal component (PC2) allows for the separation of saline from nonsaline soils, while the differential brightness in the thirds principal component (PC3) and the differential greenness in the fourth principal component (PC4) show changes occurring in surface salinity. A case study in the Indo-Gangetic Plain showed that PC1 was very effective in identifying salt-affected areas, but changes taking place between two reference dates (1975 and 1992) could not be detected for PC3 and PC4 (Dwivedi and Sreenivas, 1998). This shortcoming is attributed to the fact that PCA uses the spectral response pattern of the entire scene. It can be removed using the Kauth–Thomas transform, which deals separately with soil and vegetation reflectances to generate individual brightness and greenness images. The higher the brightness, the lower the humidity and the higher the salinity (Peng 1998).

2.4.3 Intensity–hue–saturation transformation

Dwivedi et al. (2001) made an attempt to evaluate the potential of utilizing high spatial resolution (5.8 m) Panchromatic (PAN) sensor data and Linear Imaging Self-scanning Sensor (LISS-III) data from the Indian Remote Sensing satellite (IRS) for detection and delineation of salt-affected soils. Their approach involved the merging of LISS-III and PAN data through an Intensity, Hue, and Saturation (IHS) transformation and a subsequent supervised classification using a per-pixel Gaussian maximum-likelihood classification algorithm. Their results indicated deterioration in detecting the overall accuracy of salt-affected soils derived from LISS-III data as compared to IRS-1B. Their overall accuracy for LISS-II, LISS-III, and PAN and LISS-III hybrid data with the HIS transformation were 89.6, 85.9, and 81.5 percent respectively.

2.4.4 Unmixing of surface features

One major drawback to the use of remote sensing data for salinity and alkalinity mapping is that the intricate distribution of salts on the terrain surface causes heterogeneous pixels to occur frequently. The mixed surface components can be separated using linear mixture modelling from the reflectance of pure pixels. A variety of techniques, including minimum noise fraction, principal components analysis, and identification of average pure pixels representing a particular surface component from false color composites, can be applied for identifying endmembers directly from the imagery (Metternicht and Fermont, 1998; Taylor and Dehaan, 2000). The output of the unmixing analysis is represented in a proportion map for each of the considered surface components, one image representing the sum of all the abundances and a root–mean–square (RMS) image. The RMS image represents the error between

the original mixed spectrum and the best-fit spectrum computed from the resulting endmember abundances (Smith et al., 1990). The error ranges generated by the spectral unmixing allows researchers to reclassify the image and improve its interpretability (Metternicht and Zinck, 1996 and Metternicht and Zinck, 1997). According to Taylor and Dehaan (2000) who derived salinity maps from hyperspectral data, image endmembers can efficiently illustrate the distribution of salinity, though they are likely to have different compositions within multitemporal data. They should be combined with maps of field-derived endmembers to be useful for monitoring changes in salinity and implementing remedial measures.

2.4.5 Fuzzy classification

Successful discrimination of saline and alkaline areas from remotely sensed data requires correct determination of information classes. The traditional approach to soil salinity mapping is based on rigid information classes with crisp boundaries, using soil reaction (pH) and electrical conductivity ranges as established, for instance, by the US Salinity Laboratory Staff (Richards, 1954). However, in nature, salt contents vary in a gradual manner, horizontally as well as vertically, even when conspicuous white patches of salt crust appear at the soil surface. As a consequence, broad zones of gradual transition may be misrepresented because of the arbitrary assignment of sharp class boundaries. Fuzzy classification, based on the fuzzy set theory proposed by Zadeh (1965), where a given class has a continuum of membership grades, can cope with the diffuse spatial distribution of salts and sodium better than crisp classification. Expanding the concepts highlighted by Burrough (1989) for soil classification to the issue of soil salinity, Metternicht (1998) proposes the determination of transitional fuzzy class boundaries, derived from continuous

salinity classes, to better represent real-world situations. This method provided a reliable detection of salt-affected areas, with an overall accuracy of 81% (Metternicht 1998). In all cases, fuzzy classification improved, or at least equalled, the crisp classification. One main drawback was that surface roughness produced by the cultivation of slightly alkaline land affected the radar backscattering, causing erroneous allocations of alkaline and saline–alkaline soils to non-affected areas.

2.4.6 Decision trees and neural networks

The performance of maximum likelihood classifiers, nonparametric decision trees, neural networks, and conditional probabilitistic networks for discriminating saline areas from non-affected ones has been tested using Landsat TM data, landform data derived from a digital elevation model, and conditional statements about salinity–landscape relationships (Evans 1998). Conditional probabilitistic networks, enabling the inclusion of prior knowledge about the relationships between input attributes and salinity in the classification of Landsat TM data, showed to be the best technique for salinity mapping in southwest Australia (Evans 1998).

2.4.7 Using airborne geophysics to assess soil salinity

In the last few years a number of projects have demonstrated the potential of airborne geophysics to dramatically improve the understanding of groundwater and salinity systems at a range of scales (Lawrie et al., 2000; Munday, 2004; Spies and Woodgate, 2004). In particular, airborne electromagnetics (AEM) has been shown to provide unparalleled insights into the distribution of salt accumulation, groundwater systems, and the materials that influence and/or control groundwater and salt movement within the regolith (Lawrie et al., 2000; Fitzpatrick et al., 2004; Munday,

2004). A national review of salinity mapping methods in the Australian context recently concluded that only airborne electromagnetics (AEM) can provides a proven approach to broad area mapping of salinity below the root zone (Spies and Woodgate, 2004).

A new method has been developed for producing reliable constrained inversions of time-domain AEM data in areas of electrically conductive basement (Lane et al., 2004). This enabled integration of the AEM products with other biophysical data in an integrated geoscience approach that has revealed a compartmentalized sub-surface salt accumulation and bedrock architecture that is not evident from the surface landscape (Clarke and Riesz, 2004). The study has demonstrated that airborne geophysics, particularly AEM, significantly improves the understanding of salinity risk and water security in flat alluvial landscapes that are relatively data-poor. 'Post-mortem' analysis of the airborne geophysics survey design suggests that significant reductions in data acquisition costs could be achieved for future surveys in similar landscapes by utilizing lessons learned on the scale of salt accumulation landscape elements (Lawrie et al., 2003).

However, while past studies involving airborne geophysics have been effective at knowledge generation, most have been ineffective in demonstrating that these datasets can deliver outcomes for salinity management (George et al., 2003). This is despite the insights into salt accumulation, groundwater, and salinity that they bring (George and Woodgate, 2002). In part, this is because gaining a better understanding of salinity distribution and processes is only a first step in delivering salinity management outcomes (George et al., 2003). Also, many of these early surveys had very general exploratory objectives designed to test the technology or assess methodologies (Lawrie et al., 2000). In the last few years several projects have

demonstrated the importance of gaining an understanding of salt accumulation architecture using AEM techniques to dramatically improve the understanding of groundwater and salinity systems at a range of scales (Lawrie et al., 2000; George and Woodgate, 2002; Munday et al., 2004; Fitzpatrick et al., 2004).

2.5 Remote sensing for yield assessment

2.5.1 Measuring yield by using vegetation indices

Wiegand et al. (1991) made a study on Vegetation Indices (VI), such as greenness (GVI), perpendicular (PVI), transformed soil adjusted (TSAVI), and normalized difference (NDVI) to assess yield. In one study they made boll counts and percent plant cover measurements at 60 m grid intervals in a salt-affected cotton field and calculated GVI, PVI, TSAVI, and NDVI at the grid intersections for SPOT-1 HRV and videography scenes. They found that the four VI from SPOT accounted for more of the variations in the lint yield. Shanahan et al. (2001) used remotely sensed imagery to compare different vegetation indices as a means of assessing canopy variation and its resultant impact on corn grain yield. Imagery data with 0.5-m spatial resolution were observed from aircraft on several dates during the 1997 and 1998 seasons using a multispectral, four-band (blue, green, red, and near-infrared reflectance) digital camera system. Grain yield for each plot was determined at maturity. Their results showed that green normalized difference vegetation index (GNDVI) values derived from images acquired during midgrain filling were the most highly correlated with grain yield; maximum correlations were 0.7 and 0.92 in 1997 and 1998, respectively. Normalizing GNDVI and grain yield variability within hybrids improved the correlations in both years, but more dramatic increases were observed in 1997 (0.7 to 0.82) than in 1998 (0.92 to 0.95). This suggested GNDVI

acquired during midgrain filling could be used to produce relative yield maps depicting spatial variability in fields.

Hill and Donald (2003) examined the patterns observed for a time series of 8 years over the southwest of Australia. TINDVI showed a relationship to agricultural productivity that is dependent on the extent to which the crop or total agricultural production was directly reduced by deficient rainfall. They found that TINDVI proved most sensitive to crop productivity for areas having less than 600 mm of rain and in years when rainfall and crop production were highly correlated. Wilhelm et al. (2000) stated that measurement of leaf area index (LAI) is critical to understanding many aspects of crop development, growth, and management. They also mentioned that the availability of portable meters to estimate LAI non-destructively has greatly increased their ability to determine this parameter during the cropping season. Their objective was to compare the LAI estimates by three means (AccuPAR, LAI-2000, and SunScan) to LAI measured by destructive sampling. They found that all meters underestimated LAI compared with destructive sampling. When all data from all rings of the LAI-2000 meter were included in the calculations, LAI-2000 estimates of the LAI differed from those of the other two meters. However, when data from Ring 5 was removed from the calculations, estimates LAI for the LAI-2000 improved and were indistinguishable from the other meters. The relationship between LAI estimated destructively and by each of the other meters was described by a unique linear equation for each hybrid. They suggested that users should consider protocols for operating each meter before deciding which device best suits their application.

2.5.2 Measuring yield by using biomass

Purcell et al. (2002) hypothesized that if radiation use efficiency (RUE) was constant across a range of population densities in a non-stressful environment, then increasing population density (POP) would increase biomass at the end of the season. To test this hypothesis, they evaluated the response of total biomass produced during the season to cumulative intercepted photosynthetically active radiation (PAR). They conducted their experiments using fields in Fayetteville and Keiser, AR, planted with soybean cultivars. They assessed the response of RUE to POP. At both locations they found a late sowing date shortened the life cycle of the crop by 13 to 25 days compared with an early sowing date, resulting in less PAR accumulated. Similarly, early maturing cultivars had less time for PAR and biomass accumulation relative to later maturing cultivars. At Keiser, in three of the four environments, RUE decreased linearly by 26 to 30% as the POP increased from 7 to 135 plants m^2. Final biomass at the end of the season, as a function of PAR accumulated from emergence to the full-seed-size stage of development, responded linearly to intercept PAR up to $\approx$ 400 MJ/m^2. Above 400 MJ/m^2, the response was curvilinear with little increases in biomass > 700 MJ/m^2. Their data clearly indicate that RUE decreased as POP increased and that maximum biomass production in these environments was not limited by intercepted PAR.

Serrano et al. (2000) studied the relationships between reflectance-based Vegetation Index (VI) and canopy variables (aboveground biomass, LAI canopy chlorophyll A content [LAI x Chl A], and fIPAR) for a wheat crop growing under different N supplies. They found that Nitrogen fertilization promoted significant increases in radiation interception (plant growth) and, to a lesser extent, in radiation use efficiency (RUE). The VI versus LAI relationships varied significantly among

treatments, rendering the VI-based equations unreliable to estimate LAI under contrasting N conditions. They found that the aboveground biomass was poorly correlated with grain yield, whereas cumulative VI simple ratio (SR) was a good predictor of grain yield, probably because cumulative SR closely tracked the duration and intensity of the canopy's photosynthetic capacity.

2.6 Spatial modeling

2.6.1 Geostatistical and spatial data

Geostatistics (Cressie, 1991) emerged in the early 1980s as a hybrid discipline of mining engineering, geology, mathematics, and statistics. Its strength over more classical approaches is that it recognizes spatial variability at both the large scale and the small scale. Trend-surface methods (Whitten, 1970) include only large-scale variation, assuming independent errors. Watson (1972) compares the two approaches and points out that: most geological problems have a small-scale variation, typically exhibiting strong positive correlation between data at nearby spatial locations.

Fisher (1935) was clearly aware of spatial dependence in agricultural field experiments, because he went to such great lengths to remove it. In the 1920s and 1930s, at Rothamsted Experimental Station in England, he established the principles of randomization, blocking, and replication. As well as controlling for unwanted bias, randomization also neutralizes (but does not remove) the effect of spatial correlation (Yates, 1938). There are two popular indices for measuring spatial autocorrelation in a point distribution: *Geary's Ratio* and *Moran's I.* Both indices measure spatial autocorrelation for interval or ratio attribute data.

2.6.2 Review of the terms used in the methodology

2.6.2.1 Neighborhood and spatial weights matrices

The neighborhood corresponds to the *adjacency measure.* The concept of *neighborhood* has to be quantified so that it can be applied in the calculation of spatial autocorrelation statistics. Given different criteria to define the neighborhood relationship, different matrices can be derived (Griffith, 1996). The proximity values of the **W** matrix have so far been considered as measures of *spatial* proximity. It is possible, however, to allow the **W** matrix exactly the same flexibility as the **Y** matrix and so to contain values reflecting proximity in *any* dimension. Haggett (1976) illustrates this implicitly in his analysis of the intensity of a measles epidemic in which he postulates as one possible alternative hypothesis that the disease spreads by a spatial diffusion through an urban hierarchy.

2.6.2.2 Stepwise regression analysis

Upton and Fingleton (1985) mentioned that in most of the situations when attempting to explain the variation in **Y** by means of a linear model, one would be unsure as to weather i) all the relevant **X** variables have been included in the model, and ii) all the **X** variables that have been included in the model are actually relevant. If the test of the conditional importance of an **X** variable suggests that the variable provides little extra information about the variation in **Y**, then clearly we can omit the variable from the model. There will often be many possible **X** variables that could be included in the model. The model that included all these might provide an excellent quantitative explanation of the variation in **Y**, while being virtually impossible to interpret.

The analyst would wish to find some subset of these **X** variables which provided nearly as good a numerical explanation while having a relatively simple interpretation. To select such a model from the myriad of possible models is no easy task, but various computational algorithms have been produced. Many algorithms rely on stepwise selection in which **X** variables are added to the existing model, one at a time, until there are no variables remaining which make any significant improvement to the fit. This is known as *forward selection.* An alternative is to repeatedly delete **X** variable from a complex model while retaining an adequate explanation of the variation in **Y**. This is called *backward elimination.*

2.6.2.3 Akaike information corrected criteria (AICC)

The AICC statistic allows introducing a penalty for increasing the number of model parameters. The order of the model is then selected by choosing the subset of regressors (or (k+1)) that minimizes this statistic. This statistic is given by the following equation:

$$AICC = -2L(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\sigma}}^2) + \frac{2(k+1)n}{n-k} \tag{2.1}$$

Where:
L : log likelihood
$\hat{\boldsymbol{\beta}}$: is the estimate of $\boldsymbol{\beta}$
$\hat{\boldsymbol{\sigma}}^2$: is the estimate of $\boldsymbol{\sigma}^2$
The equation of log likelihood is:

$$\begin{aligned} L(\boldsymbol{\beta},\boldsymbol{\theta}) &= (n/2)\log(2\pi) + (1/2)\log\left|\Sigma(\boldsymbol{\theta})\right| \\ &\quad + (1/2)(\mathbf{Z}-X\boldsymbol{\beta})'\Sigma(\boldsymbol{\theta})^{-1}(\mathbf{Z}-X\boldsymbol{\beta}), \qquad \boldsymbol{\beta}\in\mathbb{R}^q, \boldsymbol{\theta}\in\Theta \end{aligned} \tag{2.2}$$

2.6.2.4 Lagrange multiplier tests

Lagrange multipliers are a method for dealing with constraints. One approach suggested by Cliff and Ord (1981, p. 134) is based on estimating partial

autocorrelations. More recently there has been interest in applying Lagrange Multiplier (LM) tests (Burridge, 1981; Anselin, 1988) for nested spatial hypothesis testing, treating estimation under the null hypothesis as a constrained maximization problem.

2.6.2.5 Multiple correlation coefficient R^2

Multiple correlation coefficient R^2, which is a measure of linear fit of the data, should be reasonably high. R^2 is defined by the following equation:

$$R^2 = \frac{\hat{\boldsymbol{\beta}}\mathbf{X}\mathbf{X}'\hat{\boldsymbol{\beta}} - \bar{\mathbf{Y}}^2}{\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{Y}}^2}, \quad 0 \le R^2 \le 1 \tag{2.3}$$

2.6.2.6 P-value

P-value is the probability, if the test statistic really were distributed as it would be under the null hypothesis, of observing a test statistic as extreme as, or more extreme than the one actually observed. The smaller the *p*-value, the more strongly the test confirms the null hypothesis. *p*-value of 0.05 or less confirms the null hypothesis at the 5% level, that is the statistical assumptions used imply that only 5% of the time would the supposed statistical process produce a finding this extreme if the null hypothesis were false, where *p*-value.

2.6.2.7 Moran's I

Moran's I is a statistic to test the autocorrelation in the residuals and it should be > 0.05 to guarantee that there is no autocorrelation among the residuals. Moran's I is expressed by the following equation Moran (1950):

$$I = \frac{n}{2A} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}(Z_i - \overline{Z})(Z_j - \overline{Z})}{\sum_{i=1}^{n}(Z_i - \overline{Z})^2} \quad (2.4)$$

Where:
n: is the sample size.
Z_i: equals $x_i - \overline{x}$
$\overline{Z}$: the mean of variables
A: is area of the sample plot.
I is a correlation coefficient between "neighboring" Z's. A large value of I implies positive correlation and s small (negative) value of I implies negative correlation

2.6.2.8 Kriging technique

Kriging is a minimum-mean-squared-error method of spatial prediction that (usually) depends on the second-order properties of the process Z(•). Matheron (1963) named this method of optimal spatial linear prediction after D. G. Krige, a South African mining engineer who, in the 1950s, developed empirical methods for determining true ore-grade distributions from distributions based on sampled ore grades (Krige, 1951). The contributions of Wold (1938) and Wiener (1949) all contain optimal linear prediction equations that reflect the notion that observations closer to the prediction, point (for them, closer in time) should be given more weight in the predictor. Isaaks (1989) mentioned that ordinary kriging, a method that is often associated with the acronym B.L.U.E. for "best linear unbiased estimator." Ordinary kriging is "linear" because its estimates are weighted linear combinations of the available data; it is "unbiased" since it tries to have the mean residual or error, equal to 0; it is "best" because it aims at minimizing $\sigma^{\wedge}$, the variance of the errors. All of the other estimation methods measured so far are also linear and, are also theoretically unbiased. The distinguishing feature of ordinary kriging, therefore, is its aim of minimizing the error variance.

2.6.2.9 Variogram

Suppose,

$$\mathrm{var}\left(Z(s_1) - Z(s_2)\right) = 2\gamma(s_1 - s_2), \text{ for all } s_1, s_2 \in D \tag{2.5}$$

The quantity $2\gamma(\cdot)$, which is a function only of the increment $s_1 — s_2$, has been called a *variogram* [and $\gamma(\cdot)$ has been called a *semivariogram*] by Matheron (1962), although earlier appearances of it can be found in the scientific literature. Cressie (1988) can be consulted for some of the details, but briefly it has been called a *structure function* by Yaglom (1957) in probability and by Gandin (1963) in meteorology, and a *mean-squared difference* by Jowett (1952) in time series. Kolmogorov et. al. (1941) introduced it to study the local structure of turbulence in a fluid and Matern (1960, p. 51) makes incidental note of a Swedish forester, A. Langsaeter, who used this way of expressing variation when dealing with systematic sampling in forest surveys. Nevertheless, it has been Matheron's mining terminology that has persisted.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(ij)|h_{ij}=h} (z_i - z_j)^2 \tag{2.6}$$

Where:

h: is the distance separating sample locations *i* and *j*, z_i is the variable of interest at location *i* and *j*.

N(*h*): is the number of data pairs separated by distance *h*.

As the distance between pairs of points increases the corresponding variogram will generally increase. As the distance increases the variogram will reach a plateau. The distance at which the variogram reaches this plateau is called the range. The plateau the variogram reaches at the range is called the sill. When $h = 0$, $\gamma(h) = 0$, however because of sampling error and small scale variability $\gamma(h)$ may not be zero at

small distances. The vertical jump from 0 at the origin to the small value of the variogram at the extremely small separation distance is called the nugget effect.

2.7 Integration of remote sensing and spatial statistics

Remotely sensed data suffer from instrument induced spatial dependency effects. The problem arises because sensors measure light reflectance from the earth's surface, but this light is scattered so that reflectance from one small area of ground can be distributed over several contiguous pixels (picture elements) on the image (Forster, 1980). Craig and Labovitz (1980) discuss the effects of the type and age of the hardware employed and natural conditions (sun angle, cloud cover, geographic location and season) on the measurement process and induced spatial correlation. This type of error appears to be most serious along scan lines so that measurement induced dependency may take the form of a serial structure (Craig, 1979; Labovilz and Masuoka, 1984). The errors associated with sets of contiguous Landsat pixels are often modelled as normally distributed random variables with low order spatial correlation.

Triantafilis et al. (2006) mentioned that the results of spatial prediction suggest that regardless of what method was used, reasonable estimates of soil salinity, were achieved. They used various approaches to estimate soil salinity. Of these, geostatistical methods such as ordinary kriging, regression kriging, three-dimensional kriging, and cokriging, provide best linear unbiased estimates. These methods were tested with an electromagnetic induction instrument (EM38) in irrigated cotton *(Gossypium hirsutum L.)* in a growing area in the Edgeroi district of the lower Namoi valley, northern New South Wales, Australia. They compared these methods, on the basis of precision and bias of estimation, and found that regression kriging was the

best performer. This is because of the incorporation of regression residuals within the kriging system. Mean and standard deviation of ranks showed that cokriging performed best against these criteria.

Kalkhan et al., 2000 made a study on an area of 9,252 ha within the eastern region of Rocky Mountain National Park, Colorado, to predict the probability of exotic species richness. They used a new model incorporating trend surface analysis and stepwise regression. This process is based on the Ordinary Least Square (OLS) estimate. Landsat TM bands 1, 2, 5, 6, and 7; elevation; slope; and aspect were found to be significant predictors with $R^2 = 0.27$, residual standard error = 0.41, F-statistics = 43.64, and p-value < 0. The predictive method for the probability of exotic species richness was selected for the lowest values of AIC and AICC statistics. To model the spatial continuity of small-scale variability based on cokriging, a Gaussian semi-variogram model was selected for the lowest values of AIC and AICC statistics. The residuals from the trend surface model were modelled using ordinary kriging to account for the small-scale variability to the data. They obtained the final surfaces by combining the trend surface model based and the kriged surface of residuals. They selected all models based on the lowest value of standard errors, modified Akaike's Information Criterion (AICC) statistics, and high R^2. The trend surfaces models had R^2 values ranging from 10.04% to 58.6%. When the kriged residuals were added to the trend surface models, the R^2 values ranged from 60% to 84%.

Kalkhan et al., 1999 used the cross-correlation statistic I_{YZ}, to test for the presence of spatial cross-correlation between pair-wise combinations of soil characteristics, topographic variables, plant species richness, and cover of vascular plants in a 754 ha study site in Rocky Mountain National Park, Colorado. Using 25 large plots (1,000 m^2) with five vegetation types, they found that 8 of 12 variables

showed significant spatial cross-correlation, while 6 of 12 variables showed significant spatial auto-correlation. Elevation and slope showed significant spatial cross-correlation with all variables except percent cover of native and exotic species. Percent cover of native species had significant spatial cross-correlations with soil variables, but not with exotic species. The researchers believe that this is probably because of the patchy distributions of vegetation types in the study area. At a finer resolution, using data from 1 m^2 subplots within each of the 1000 m^2 plots, all variables showed significant spatial auto- and cross-correlation. Large-plot sampling was more affected by topographic factors than species distribution patterns, while with finer resolution sampling, the opposite was true. However, the statistically and biologically significant spatial correlation of native and exotic species could only be detected with finer resolution sampling. They found exotic plant species invading areas with high native plant richness and cover and in fertile soils high in nitrogen, silt, and clay.

Stohlgren et al. (1997) presented a rapid, cost-efficient methodology to link plant diversity surveys from plots to landscapes using: (1) unbiased site selection based on remotely sensed information; (2) multi-scale field techniques to assess plant diversity; (3) mathematical models (species-area curves) to estimate the number of species in larger areas corrected for within-type heterogeneity; and (4) mathematical techniques to estimate total species richness and patterns of plant diversity in a landscape. They demonstrated the methodology in a 754 ha study area in Rocky Mountain National Park, Colorado using four 0.025 ha and twenty-one 0.1 ha multi-scale vegetation plots. They recorded 330 plant species (~ 1/3 the number of plants recorded in the 1,074 km^2 Park) in the 2.2 ha area within the plots which represents a sampling intensity of 0.29% of the 754 ha study site. They estimated 552 plant

species, about half the plant species recorded in the Park, in just 0.7% of the Park's area. They demonstrated how this rapid, cost-effective methodology: (1) produces a rich information base on the patterns of native plant diversity and the distribution of non-native plant species and keystone ecosystems; and (2) can be easily adapted for the other national and state parks, national forests, wildfire refuges, and nature reserves.

Utset et al. (1998) used a calibrated Four-Electrode Probe (FEP) for inexpensive and indirect determinations of salinity-sensor Electrical Conductivity (EC) in a plot in Cauto Valley, Cuba. They made two transects in the North-South (N-S) and East West (E-W) directions. They made laboratory measurements of soil EC from samples taken on a 50-m square grid. They obtained a linear semivariogram from the salinity-sensor EC measurements of the E-W transect, which agreed with the topographical slope of the plot and with the expected soil salinity variation. It also coincided with the spatial structure of laboratory-measured soil EC. A cross-validation analysis showed that EC semivariograms obtained from FEP measurements can characterize the soil EC spatial variation in a similar way to semivariograms of laboratory-measured soil EC.

CHAPTER 3. SITE DESCRIPTION AND DATA COLLECTION

3.1 Site description

Colorado State University's Arkansas River Basin Salinity Mapping Project where this research has been conducted began in 1999. The study area (figure 3.1) is located in southeastern Colorado, near the town of La Junta. Fields in this area are cultivated with alfalfa, corn, wheat, onions, cantaloupe and other vegetables and irrigated by a variety of systems including a mixture of border and basin, center pivot, and furrow. Salinity levels in the canal systems along the river increase from 300 ppm total dissolved solids (TDS) near Pueblo to over 4,000 ppm at the Colorado-Kansas border (Gates et al., 2002). The study area consists of a number of fields covering a total area of about 20 miles in length and 10 miles in width. Sixty eight observation wells have been installed in these fields. In the wells, water table fluctuations and groundwater salinity are measured on a weekly basis throughout the irrigation season and on a monthly basis after the irrigation season. Soil salinity is also measured in the fields at the beginning and at the end of the irrigation season. The description of the alfalfa and corn monitored fields is given in the following table.

Table 3.1: The description of the alfalfa and the corn monitored fields.

Field #	Area (acre)	Crop type (2001)	Crop type (2004)	Soil salinity range (dS/m)	Irrigation system
US14	29	Alfalfa	Alfalfa	3.07 – 6.77	Furrow (gated pipes)
US04	261	Alfalfa	Alfalfa	2.70 – 20.71	Center pivot sprinkler
US09	52	Corn	Alfalfa	2.50 – 3.70	Furrow (gated pipes)
US10	10.8	Corn	Alfalfa	3.14 – 13.22	Furrow (gated pipes)
US40	20.2	Corn	Corn	3.00 – 12.20	Furrow (gated pipes)
US80	33.6	Corn	Corn	2.70 – 11.70	Furrow (siphons)

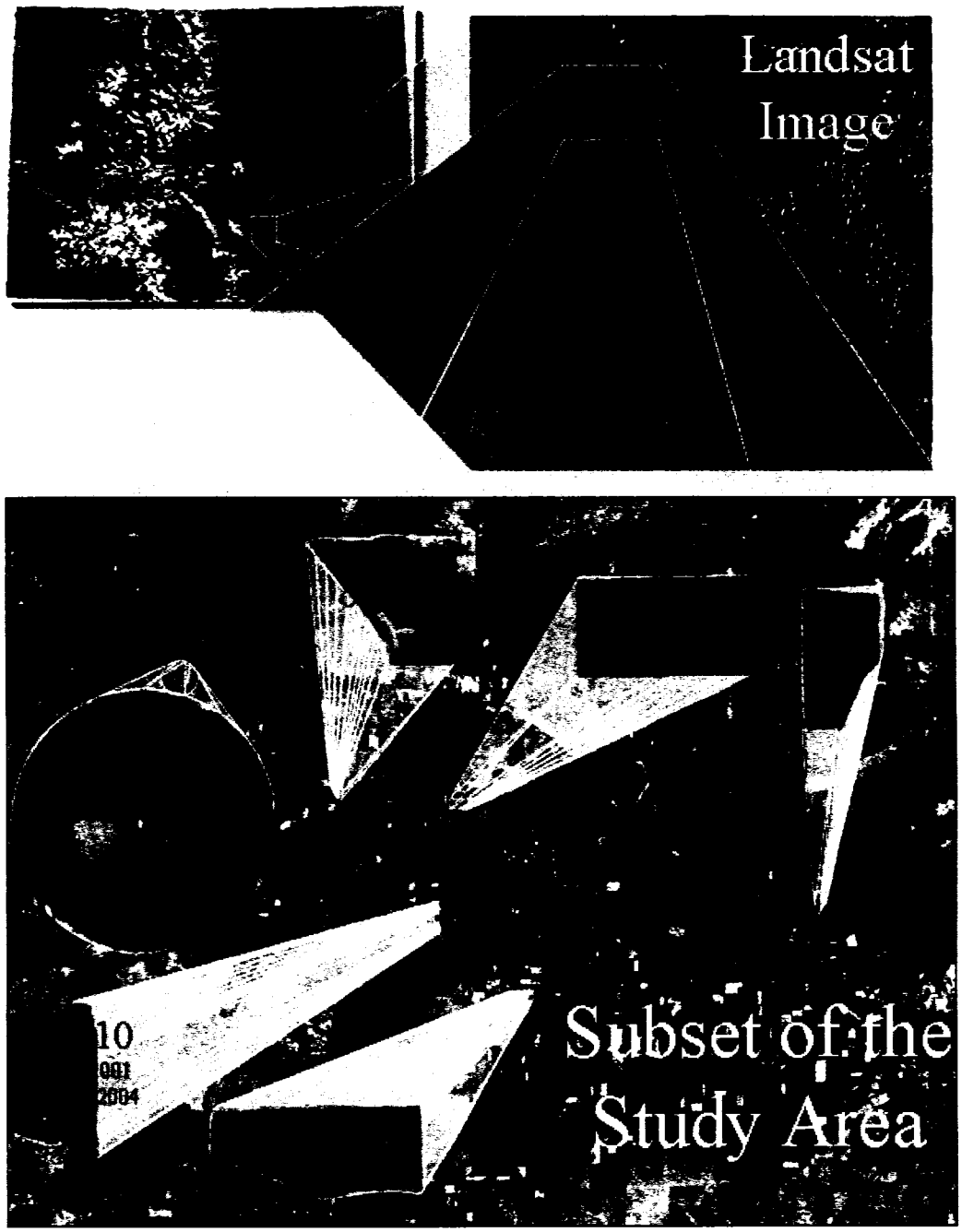


Figure 3.1: The study area of the salinity monitoring program near La Junta with Landsat image.

3.1 Data collection

Crop yield data and data pertaining to the conditions that can affect soil salinity, such as depth to water table or groundwater salinity, were collected in a number of fields. The fields for the study were selected to represent a variety of

scenarios with different soils, irrigation systems, and crops. Examples of data being collected are: depth to water table, groundwater salinity, soil salinity, crop samples, evapotranspiration, rainfall, and soil texture. During the growing season, data for water table depth, groundwater salinity, rainfall, and evapotranspiration are collected on a weekly basis. The same data are collected on a monthly basis during the rest of the year. Crop samples are collected in the form of number of plants, biomass and grain yield.

3.1.1 Soil salinity

Soil salinity is measured at the beginning, middle, and end of the irrigation season using an EM-38 instrument which is shown in Figure 3.2. These readings are transferred to dS/m (deciSiemens per meter) using a calibration equation (equation 4.1).

In addition to collecting soil salinity data with the EM-38, a number of soil samples are collected at four depths (1 foot, 2 feet, 3 feet, and 4 feet) and analysed in the laboratory using a Hach salinity kit. The EM-38 provides depths of exploration of 1.5 meters and 0.75 meters in the vertical and horizontal directions respectively. Very lightweight and only one meter long, the EM-38 provides rapid surveys with excellent lateral resolution. The readings of the EM-38 are affected by soil temperature and soil moisture content. In each field where soil salinity is measured using the EM-38, soil moisture content and soil temperature is measured in at least four different places. To calibrate the EM-38, a large number of soil samples were taken at four different depths (soil surface, 1 – 2 feet, 2 – 3 feet, and 3 – 4 feet) and analysed using the Hatch kit.

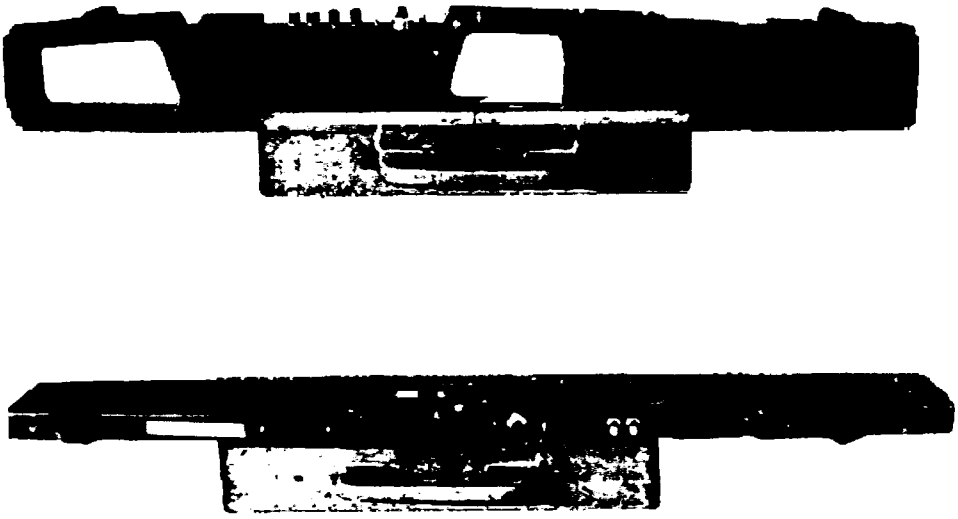


Figure 3.2: EM-38 that measures both vertical and horizontal readings individually.

3.1.2 Depth to water table

Water table has a great impact on soil salinity. Water table becomes a problem when it rises into the root zone or a little below the root zone. A high water table can prevent aeration in the submerged part of the root zone (waterlogging). Even if the water salinity is low or moderate, high water table can cause salinity problems in the long run because the water that directly evaporates from the soil surface is fairly pure and leaves salts in the soil profile. At selected wells, water table depth is measured on a continuous basis using an automatic level recorder (ALR) (figure 3.3).

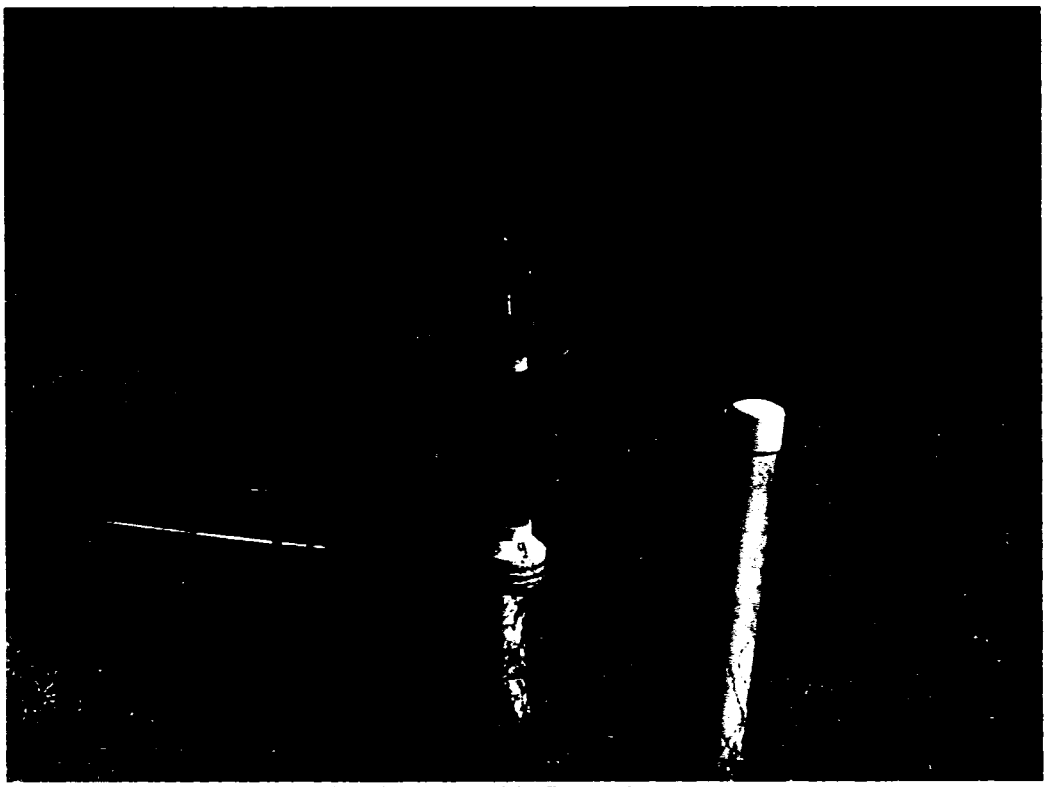


Figure 3.3: Tools for measuring the water table fluctuation.

3.1.3 Groundwater salinity

Sometimes groundwater salinity increases at the field scale due to the movement of groundwater coupled with the presence of marine deposits or due to upward flux and evapotranspiration. Groundwater salinity was measured in this study using YSI 30 conductivity meters (Figure 3.4) The YSI 30 has the ability to display reference temperature and the temperature coefficient for use when measuring temperature - compensated conductivity.

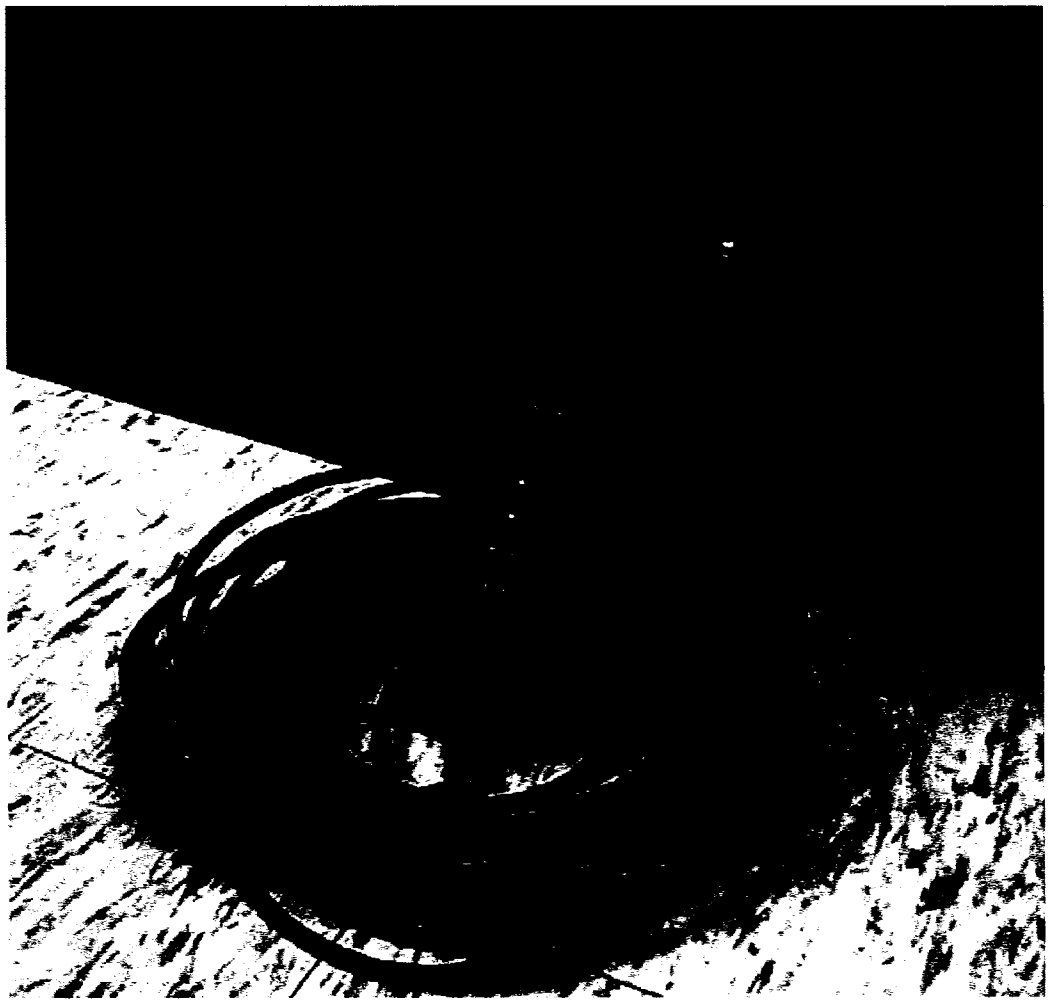


Figure 3.4: Instrument for measuring the groundwater table salinity.

3.1.4 Irrigation water

The quality and quantity of irrigation water is a crucial factor when soil salinity is an issue. When farmers apply more water than a crop requires, high water tables and waterlogginging can result. Also, when irrigation water has high salinity it can cause groundwater salinity and soil salinity to increase. In most of the monitored fields, irrigation water quantity was measured. For the center pivots, the irrigation water quantity was calculated based on flow meters connected to the pivot system. For furrow or border systems, flumes were used, as shown in figure 3.5. The irrigation water quality was measured using YSI 30 meters.

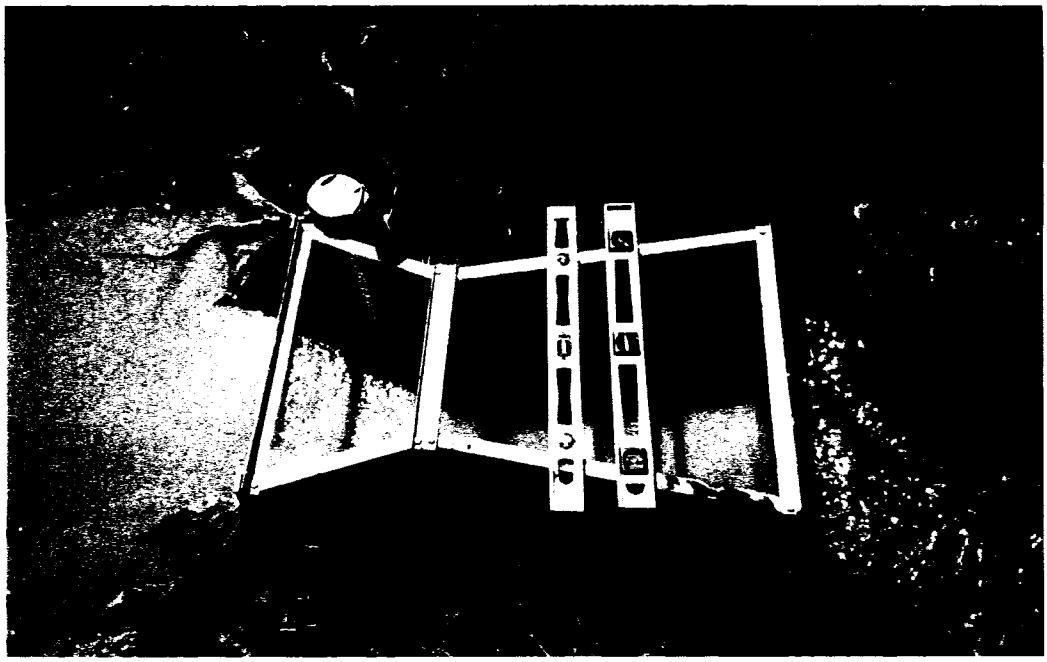


Figure 3.5: Typical flume setup for measuring irrigation quantity.

3.1.5 Rainfall

Rainfall affects the water balance both in quality and quantity. Figure 3.6 shows the type of rain gage which was used for measuring rainfall. A data logger, which is also shown in the same figure, is connected to the rain gage and records the data for rain events at 0.1 mm increments.

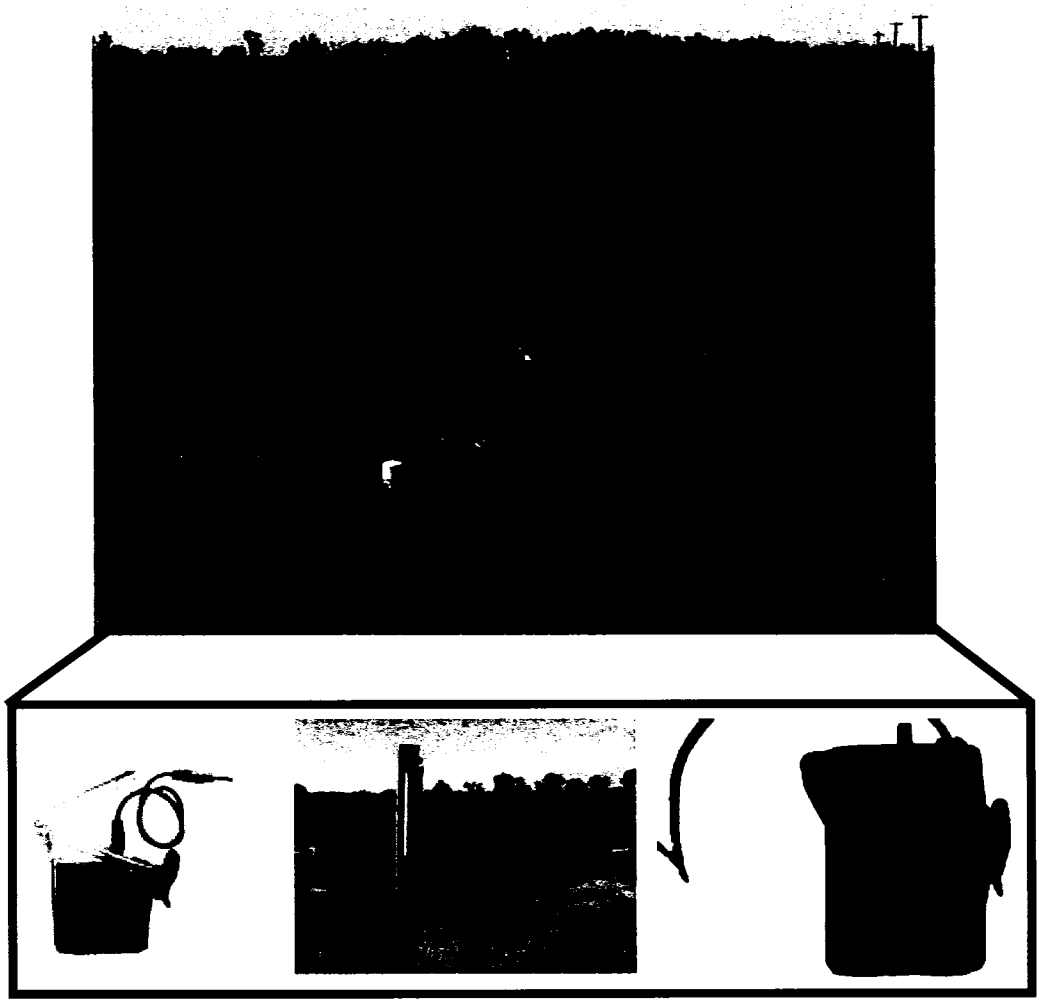


Figure 3.6: Tools for measuring rainfall.

3.1.6 Evapotranspiration

Evapotranspiration (ET) affects the water and salt balance. ET is the process where water evaporates leaving behind water with higher concentrations of salts. ET was measured using atmometers, as shown in Figure 3.7, where the level of water remaining in the reservoir is read directly from a scale on the reservoir in increments of 1 mm. Modified Bellani plate atmometers are used as "ET gages" (ETGs) since they can be setup to estimate the evapotranspiration by crops and have been studied for the past couple of decades. They consist of a ceramic plate covered by a non-fading green canvas to mimic albedo and surface resistance of green plants. The

ceramic plate is connected to a water reservoir made of PVC pipe with an attached glass sight tube.

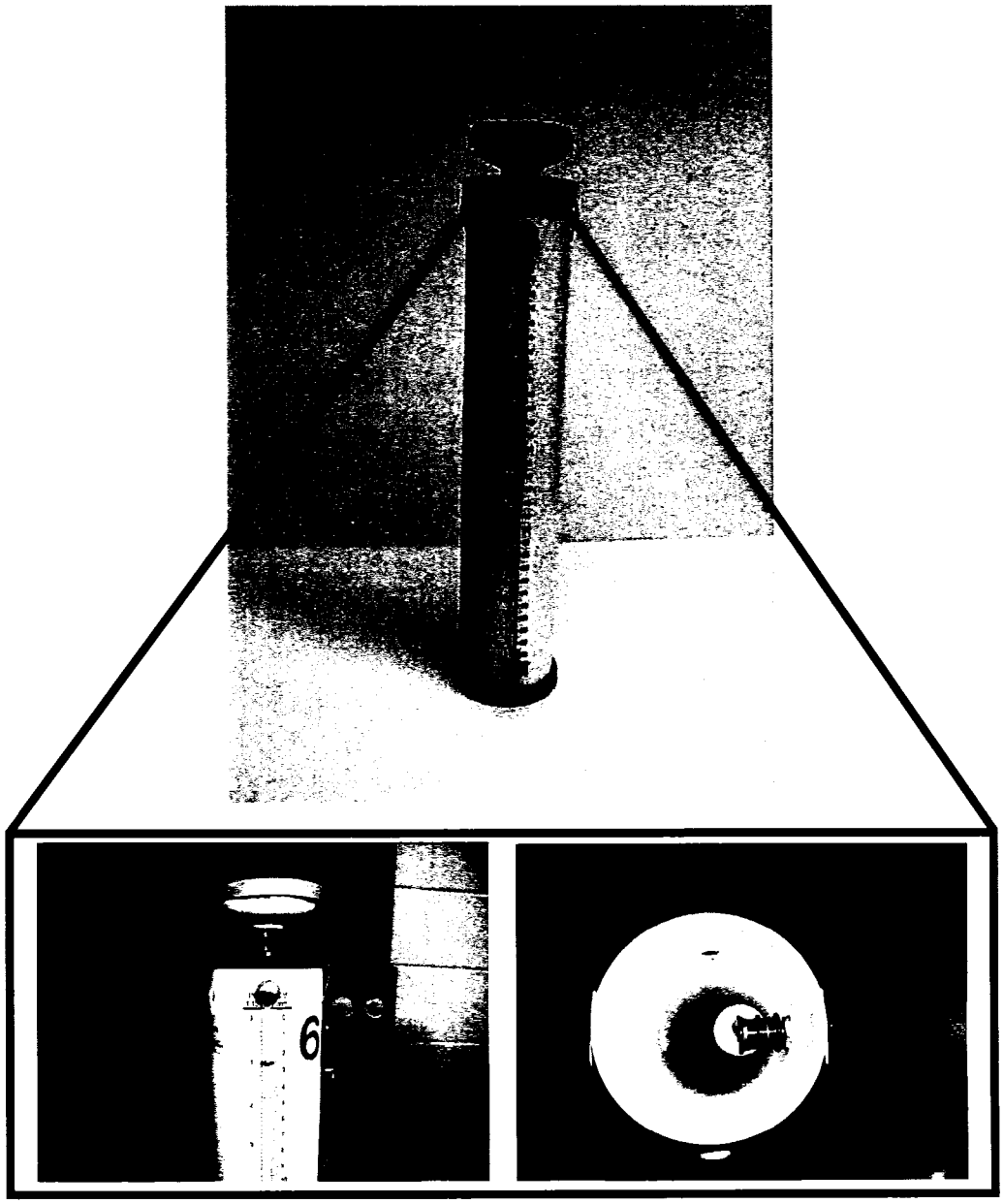


Figure 3.7: Measuring the evapotranspiration.

3.1.7 Soil moisture and temperature

Soil moisture content and soil temperature both affect the EM-38 probe readings. At the same time when the soil salinity data is measured by the EM-38 probe, samples were taken to measure soil moisture content and soil temperature in at

least four locations in each field. Soil samples were collected at four different depths (0 – 1 foot, 1 – 2 feet, 2 - 3 feet, and 3 – 4 feet). The weight of the soil samples wee measured in the field to determine the wet weight. These samples are then put in an oven with a temperature of 105 °C for 24 hours to remove all moisture from the samples and then weight again to obtain a dry weight. The gravimetric moisture content is obtained by dividing the difference between the wet weight and the dry weight of the soil samples.

3.2 Using GPS and GIS in data collection and producing maps

GPS and GIS play an important role in this study. Coordinates for the field borders, wells and sampling locations have been collected using GPS units and the data transfer to GIS for display and analysis.

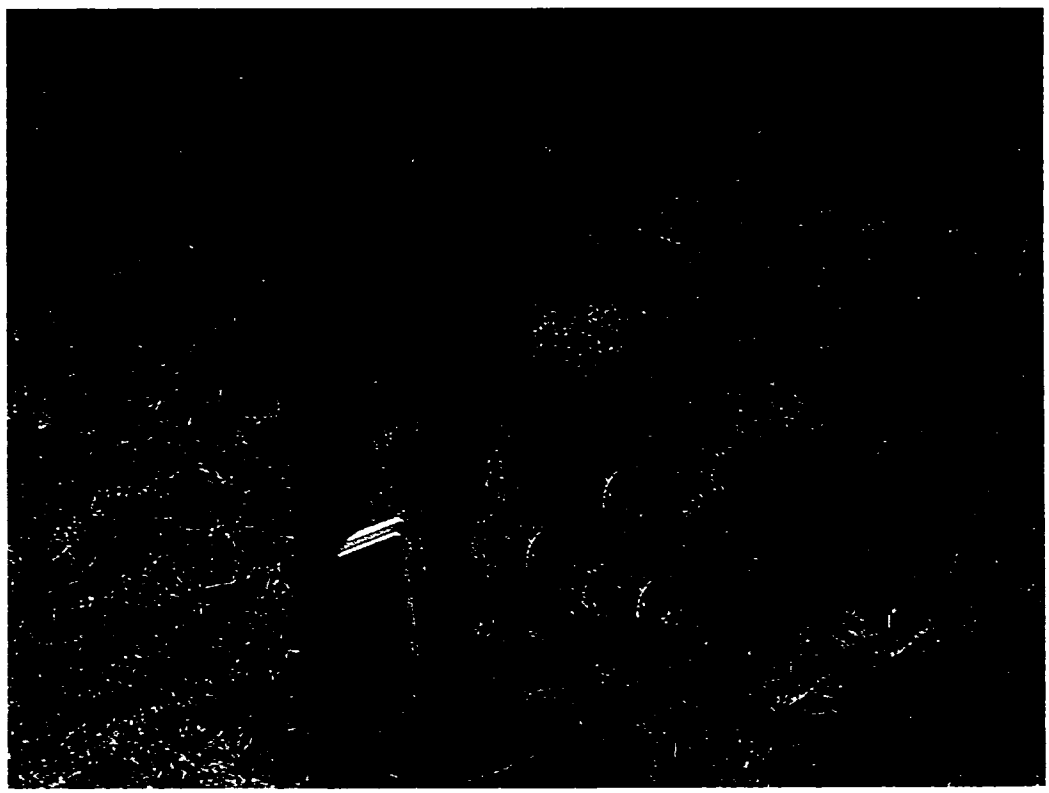


Figure 3.8: Measuring the soil moisture and soil temperature.

3.3 Remote sensing data

Five images will be tested in this study. Three images were acquired from Aster, Landsat, and Ikonos for the observed soil salinity in corn fields in the year 2001. The Aster, Landsat, and Ikonos images were acquired July 16, July 8, and July 11 of the year 2001 respectively. Two Ikonos image were acquired from Landsat and Ikonos for the observed soil salinity in alfalfa fields in the year 2004. The Landsat image was acquired in July 22 and the Ikonos image was aquiered in July 1 of the year 2004.

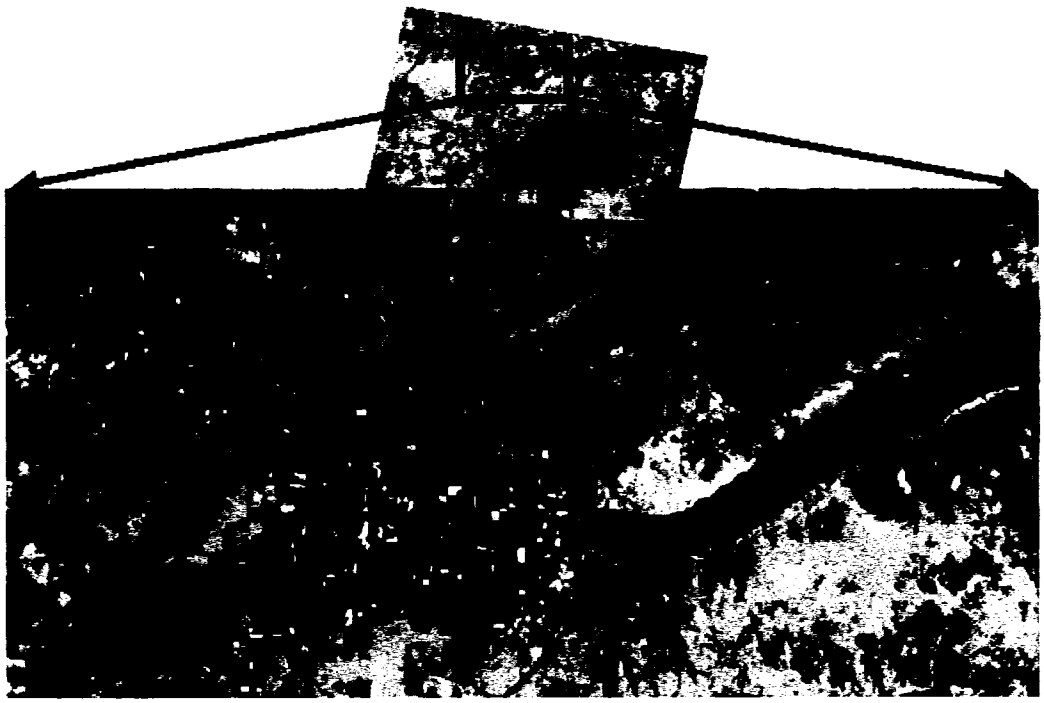


Figure 3.9: Landsat image acquired in July 8, 2001.

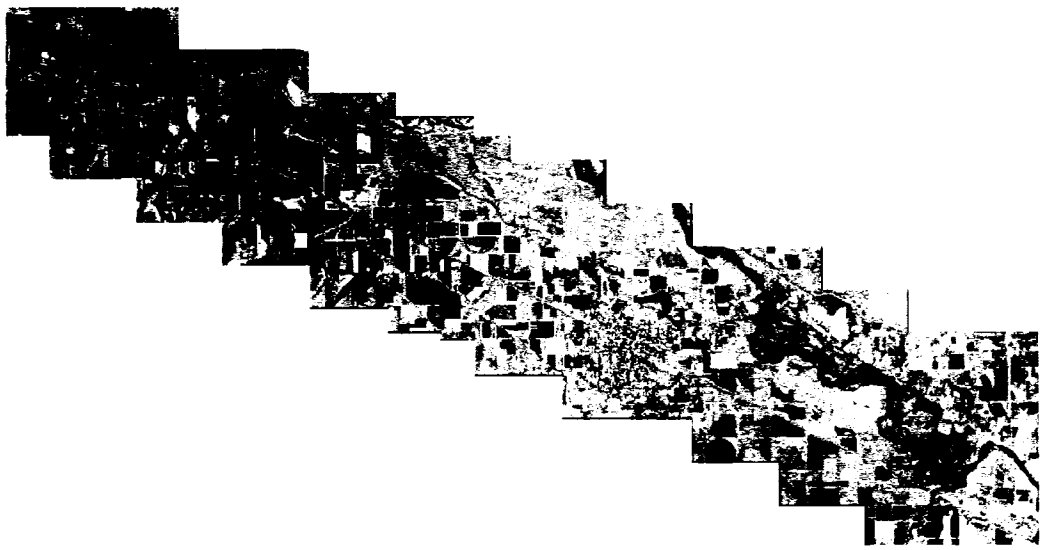


Figure 3.10: Ikonos image acquired in July 11, 2001.

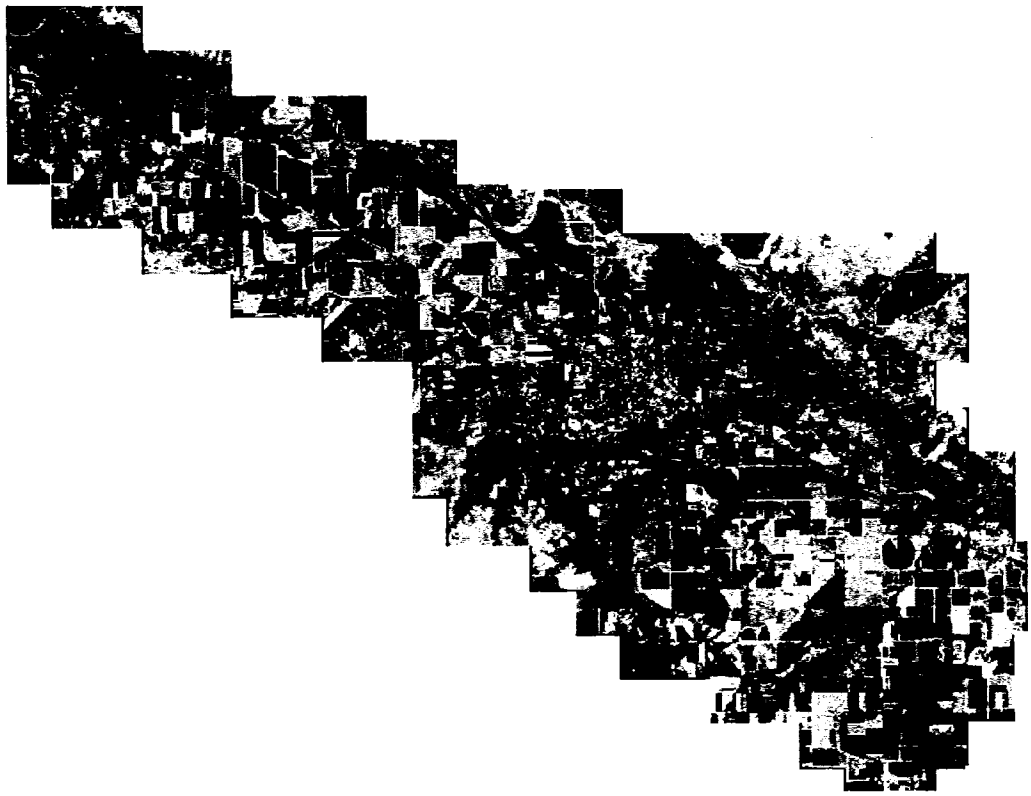


Figure 3.11: Ikonos image acquired in July 22, 2004.

CHAPTER 4. METHODOLOGY

4.1 Overview

In this study, the correlation between soil salinity and satellite imagery is tested using a variety of techniques that includes: ordinary least squares (OLS), spatial autoregressive model (SAR), and kriging. Five images were acquired from Aster, Ikonos, and Landsat to check to correlation between soil salinity and remote sensing data.

Most satellite images, including Aster, Ikonos, and Landsat, are unable to penetrate the soil surface. The reflectance of the satellite image is mainly affected by what is on the soil surface. Most irrigated crops, especially low to medium salt tolerant crops, can be used as a good indicator of soil salinity. The impact of soil salinity is displayed in the crop biomass on the soil surface. Alfalfa and corn, crops with medium salt tolerance, are the main crops grown in the area covered by the images. Therefore, alfalfa and corn are used as indicators for soil salinity.

To test for the cross correlation between the soil salinity data that has been observed with the EM-38 and the bands of the satellite imagery, three different approaches are used. The first technique is to fit the OLS model for the soil salinity data with the combination of bands, indices, and band ratios of the satellite imagery. By subtracting the observed responses from the predicted responses, estimates of error that are called residuals are obtained. Examining residuals is a key part of all statistical modelling. Residuals indicate whether the chosen model is appropriate. Therefore, residuals are inspected for normality and spatial autocorrelation. If the use

of the OLS model can not satisfy the selection criteria of the residuals which will be discussed later, then other techniques are used.

The same variables, soil salinity with a combination of satellite imagery bands which are used with the OLS model, are also used with the SAR model. The SAR model has the capability of removing some of the spatial autocorrelation in the residuals. If the use of the SAR model does not satisfy the model assumptions and the selection criteria, the third technique is used. The third technique is to krig the residuals of the OLS model and combine the surfaces generated by the OLS model and the kriged residuals. The model selected should meet the model assumptions and selection criteria which will be discussed later in this chapter. The following flowchart summarizes the procedure for model selection.

4.2 Adjusting EM-38 soil salinity

The EM-38 takes vertical and horizontal soil salinity readings. When collecting geo-referenced soil salinity data, the EM-38 probe is used in conjunction with a GPS unit to obtain geographic coordinates of the observed soil salinity data points. The EM-38 can cover large areas quickly without ground electrodes. The EM-38 provides depths of exploration of 1.5 meters and 0.75 meters in the vertical and horizontal dipole modes respectively. Measurement is normally made by placing this instrument on the ground and recording the meter reading. Digital meters are located on the top and the side of the EM-38 for the horizontal and vertical dipole measurements.

The EM-38 soil salinity readings are affected by soil moisture content and soil temperature. Therefore, some calibration processing needs to be done for each study area to acquire accurate readings of soil salinity. To achieve this, an EM-38 reading is

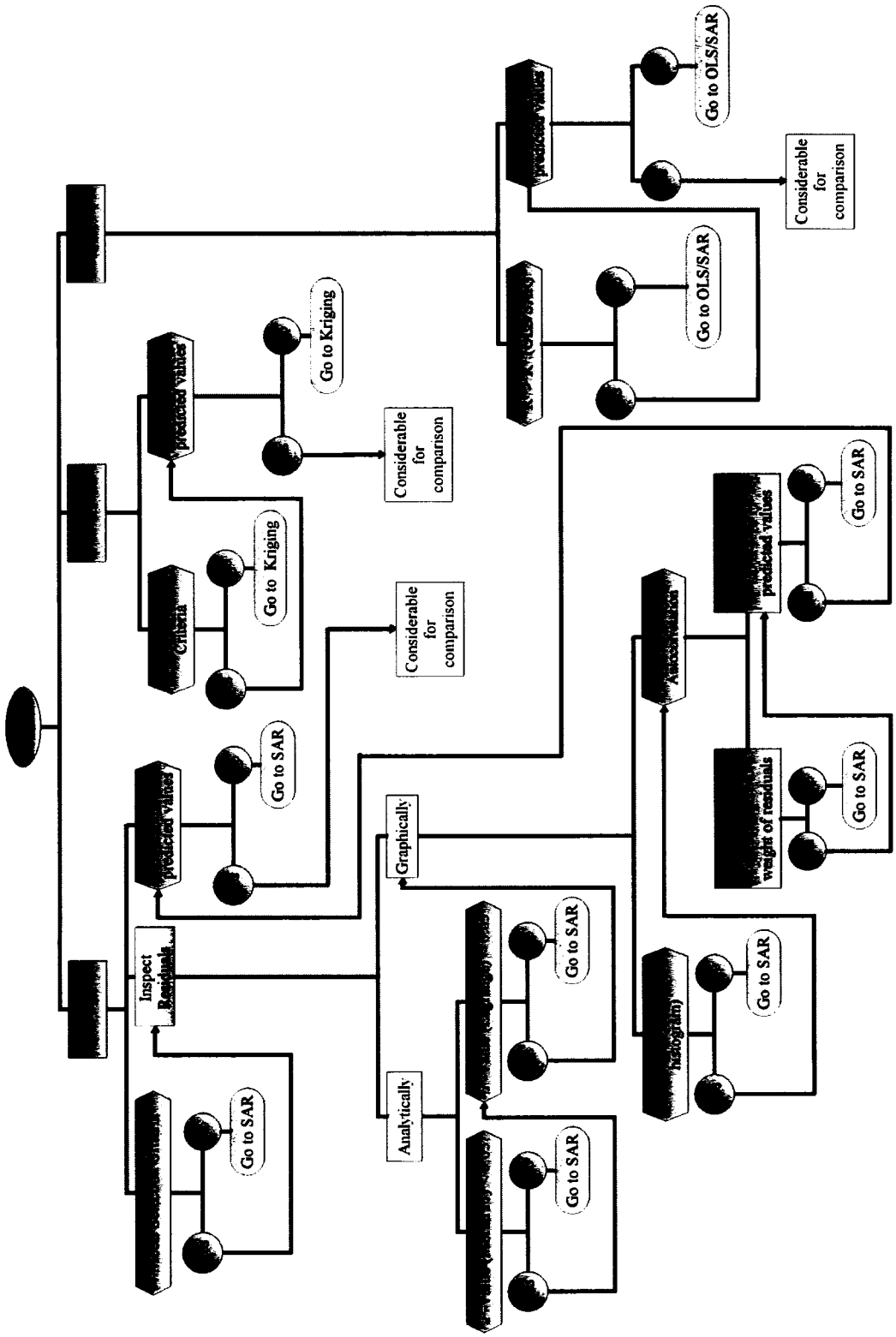


Figure 4.1: Flowchart for applying different spatial models to soil salinity data using satellite images.

taken along with four different soil samples at different depths (ground surface, 1 – 2 feet, 2 – 3 feet, and 3 – 4 feet) as well as soil moisture and soil temperature. The soil samples are analysed in the laboratory using a Hach salinity kit. The following equation was developed for our study area for the calibration of the EM-38 taking into consideration the EM-38 readings, soil moisture content, and soil temperature (Whittler et al., 2005).

$$\text{Soil Salinity (Adjusted)} = 0.45 + 7.23 * \left(\frac{\text{EM38v (co)}}{100}\right)^{1.78} + 19.54 * GMC - 34.06 * \left(\frac{\text{EM38v (co)}}{100}\right) * GCM \quad (4.1)$$

Where:

EM-38v(co): is the vertical reading of the EM-38 after being adjusted for soil temperature. (EM-38 readings are not adjusted for the soil temperature if it is 25°C.)

GMC: is the gravimetric moisture content of the soil.

4.3 Processing the remote sensing data

The necessary processing for images such as atmospheric correction, geometric correction, mosaic, and subsets was done for the images used in this study. The atmospheric correction was used since it compensates for the effects of selectivity scattered light in multispectral images. The geometric correction was used since it corrects for spatial distortion of the images and it guarantees that the same points on the image represent the same points on the ground. The Ikonos images were mosaic since each image covers a small portion of the study area. While the Aster and Landsat images were subset to take a portion of image that covers the study area. Figure 4.2 shows an example of the process of geometric correction for the 2001 Ikonos image. Points used for the geometric correction process should be scattered all

over the image with special attention to the field(s) of interest. To make sure that the rectification is accurate, the root mean square (RMS) should be less than 2.

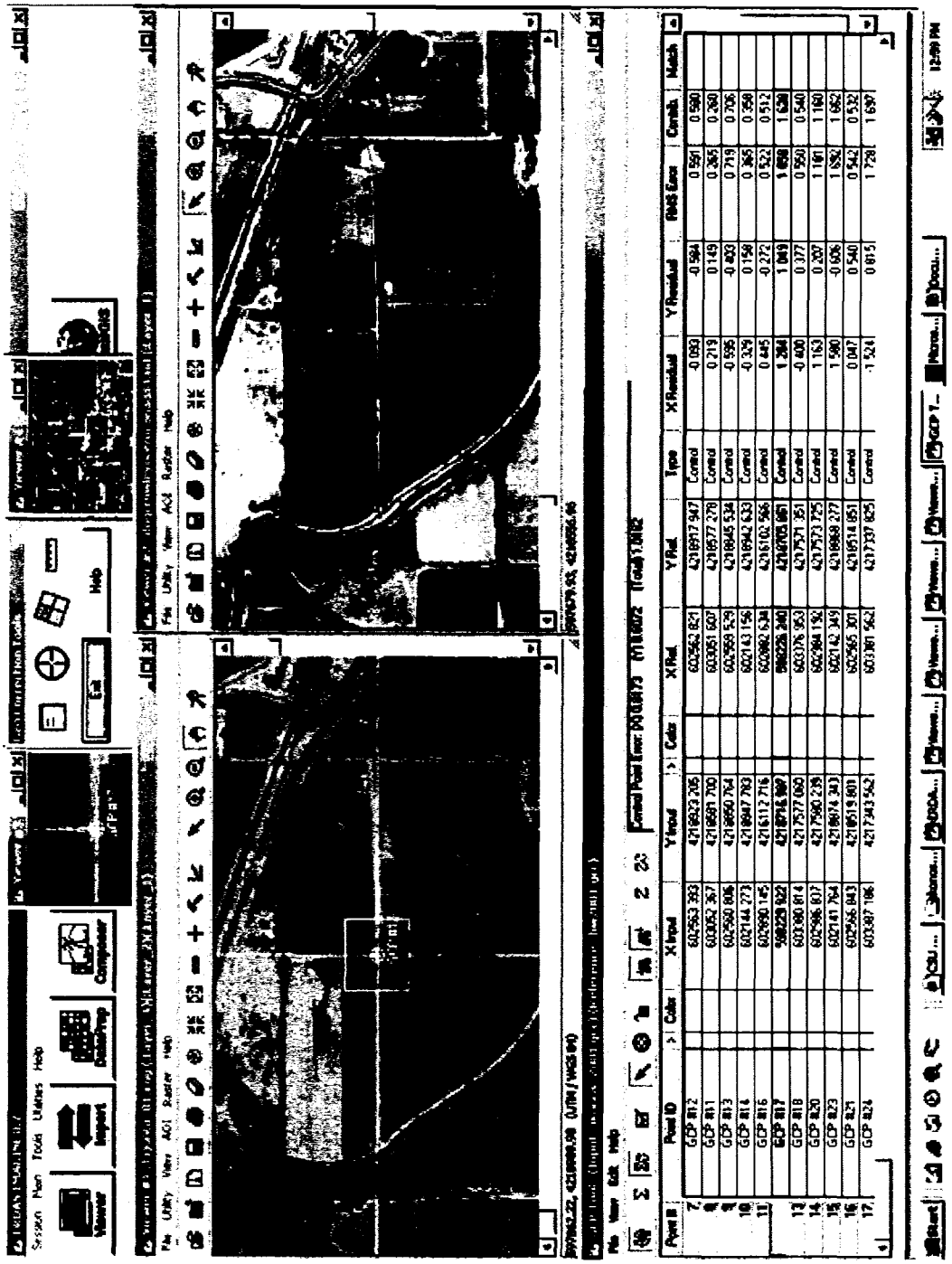


Figure 4.2: Rectifying the image using black and white reference image of one-meter resolution.

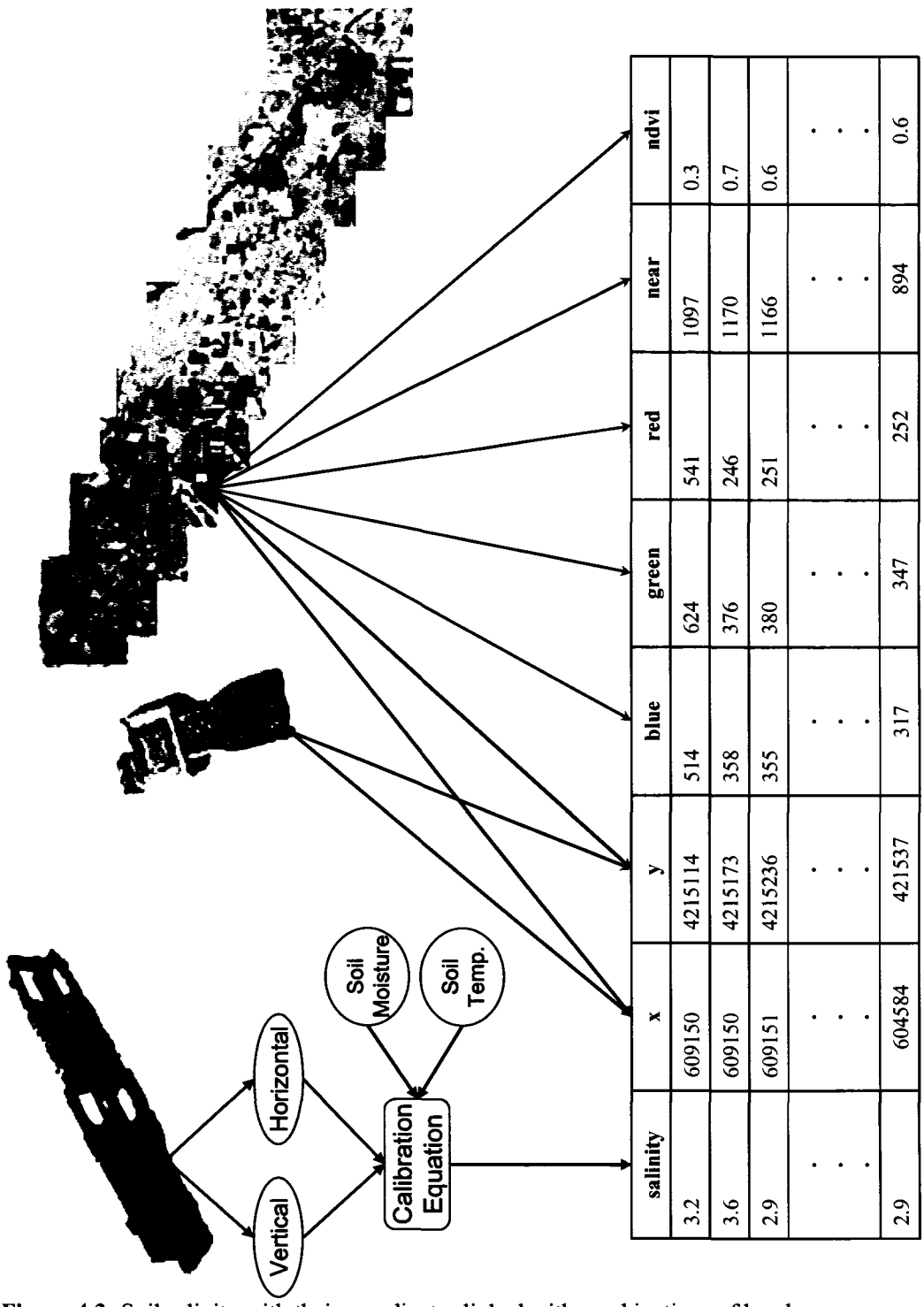
4.4 Combination of bands of the satellite images

All satellite images have more than one band and each band has its own spectral resolution and sometimes its own spatial resolution. With the unique design of the bands regarding spectral and spatial resolution, each band is specifically designed to detect something different. However, all bands do not necessarily have good correlation with the variable of interest such as soil salinity. Therefore, there is a need to determine which combination of bands has the best correlation with soil salinity data. In addition to the bands, some other indices such as normalized difference vegetation index (NDVI) which have proven helpful in predicting soil salinity in other studies were evaluated. Most previous studies used either all bands or selected only individual bands having the strongest correlation with soil salinity.

In this study, a combination of bands that has the best correlation with soil salinity is selected. The advantage of this technique is that it selects only the bands that have correlation with soil salinity and disregards the bands that do not correlate with soil salinity. This technique removes noise from the image that can be caused by the bands that do correlate with soil salinity. This technique also has an advantage over using a single band for predicting soil salinity because when using a combination of correlated bands each band can contribute some degree of correlation.

4.5 Linking soil salinity data with remote sensing data

The values of soil salinity which have been adjusted based on equation 4.1 are linked with the corresponding pixel values of the satellite image's combination of bands. Figure 4.3 shows how the coordinates of the field-observed soil salinity data were linked to points on the satellite image and used to extract the corresponding pixel values of different bands.



salinity	x	y	blue	green	red	near	ndvi
3.2	609150	4215114	514	624	541	1097	0.3
3.6	609150	4215173	358	376	246	1170	0.7
2.9	609151	4215236	355	380	251	1166	0.6
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
2.9	604584	421537	317	347	252	894	0.6

Figure 4.3: Soil salinity with their coordinates linked with combinations of bands.

4.6 Using OLS modeling techniques

4.6.1 OLS model description

Suppose that the observation y_i has been recorded on the variate $Y_i (i = 1, 2, \ldots, n)$. In vector notation, let $y'(= y_1, y_2, \ldots, y_n)$ be the $(n \times 1)$ vector of observed values corresponding to the variate vector $\boldsymbol{Y}'(=Y_1, Y_2, \ldots, Y_2)$. $\boldsymbol{Y}$ is assumed to have the mean vector $\boldsymbol{X}\beta$ and covariance matrix $\sigma^2 \boldsymbol{V}$; that is,

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}\beta \tag{4.2}$$

And

$$\mathrm{var}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{V} \tag{4.3}$$

Where $\boldsymbol{X}$ is a $(n \times 1)$ matrix of nonstochastic regressor variables,

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{12} & \cdots & X_{1k} \\ 1 & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n2} & \cdots & X_{nk} \end{bmatrix} \tag{4.4}$$

The vector β is of order $(k \times 1)$, and contains the parameters $\beta_i (i = 1, 2, \ldots, k)$; that is $\beta' = (\beta_1, \beta_2, \ldots, \beta_k)$, and **V** is a $(n \times n)$ matrix with elements $\{v_{ij}\}$.

The OLS model specifies the regression equation

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon} \tag{4.5}$$

Where $\boldsymbol{\varepsilon}$ is a $(n \times 1)$ vector of random error terms. It follows from the above equation and the assumption of nonstochastic $\boldsymbol{X}$ variables that

$$E(\boldsymbol{\varepsilon}) = 0 \tag{4.6}$$

4.6.2 OLS model selection criteria

The OLS model selection criteria will be based on the following: 1) multiple correlation coefficient R^2 is high; 2) p-value should be < 0.05 to guarantee that there is a strong cross correlation between soil salinity and the image; 3) Moran's I p-value should be > 0.05 to guarantee that there is no autocorrelation among the residuals; 4) Lagrange multiplier p-value should be < 0.05 as another test to guarantee that there is no autocorrelation among the residuals; and 5) the AICC of the model selected should have the smallest value among the tested models.

4.6.3 Inspection of residuals of the OLS model

Residuals are estimates of experimental error obtained by subtracting the observed responses from the predicted responses. Examining residuals is a key part of all statistical modelling. Carefully looking at residuals can tell whether assumptions are reasonable and the choice of model is appropriate. Residuals can be thought of as elements of variation unexplained by the fitted model. Since this is a form of error, the same general assumptions that apply to errors apply to the group of residuals: one expects them to be (roughly) normal and (approximately) independently distributed with a mean of 0 and some constant variance.

An analyst should expect a regression model to have errors in predicting a response in a random fashion; the model should predict values higher than actual and lower than actual with equal probability. In addition, the level of the error should be independent of when the observation occurred in the study, or the size of the observation being predicted, or even the factor settings involved in making the prediction. The overall pattern of the residuals should be similar to the bell-shaped pattern observed when plotting a histogram of normally distributed data. Departures

from these assumptions usually mean that the residuals contain structure that is not accounted for in the model. Identifying that structure and adding term(s) representing it to the original model leads to a better model. Any graph suitable for displaying the distribution of a set of data is suitable for judging the normality of the distribution of a group of residuals. The three most common types of graphs are: histogram, normal probability plots, and dot plots. The histogram, a frequency plot obtained by placing the data in regularly spaced cells and plotting each cell frequency versus the center of the cell, will be used in this study. Plotting residuals versus the value of a fitted response should produce a distribution of points scattered randomly about 0. The first measure of spatial autocorrelation was introduced by Moran (1950). The index is analogous to the conventional correlation coefficient, and its values range from 1 (strong positive spatial autocorrelation) to -1 (strong negative spatial autocorrelation). It is often used to measure the spatial autocorrelation of ordinal, interval or ratio data.

If the residuals of the OLS prove to have spatial autocorrelation, two approaches can be used to remove this autocorrelation or the spatial dependency. The first approach is to introduce the same variables to the SAR model which has the ability to remove the spatial autocorrelation among the residuals. The second approach is to krig the residuals of the OLS model and to combine this generated surface to the surface generated from the OLS model, which can get rid of the autocorrelation or spatial dependency in the residuals. The selection between the SAR model and the OLS model combined with kriged residuals is based on the higher R^2 value.

4.7 Using SAR modeling techniques

4.7.1 SAR model description

The SAR model is represented by the following equations:

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{4.7}$$

And

$$\varepsilon_i = \rho \mathrm{L}\varepsilon_i + u_i \qquad i = 1, 2, \ldots, n \tag{4.8}$$

Where L is the lag operator. These equations yield the transformed relationship,

$$\begin{aligned} y_i &= \alpha(1-\rho) + \beta x_i + \rho \mathrm{L} y_i - \rho\beta \mathrm{L} x_i + u_i \\ &= \gamma_1 + \gamma_2 x_i + \gamma_3 \mathrm{L} y_i + \gamma_4 \mathrm{L} x_i + u_i \end{aligned} \tag{4.9}$$

4.7.2 SAR model selection criteria:

The SAR model selection criteria will be based on the following: 1) multiple correlation coefficient R^2 is high; 2) p-value should be < 0.05 to guarantee that there is a strong cross correlation between soil salinity and the image; 3) Moran's I p-value should be > 0.05 to guarantee that there is no autocorrelation among the residuals; 4) maximum likelihood value should be < 0.05 and 5) the AICC of the model selected should have the smallest value among the tested models.

4.8 Modified kriging model:

The kriged residuals of the OLS model combined with the OLS model is another technique to over come the problem of autocorrelation in the residuals. The variogram selected is based on the smallest value of the AICC. Once the best variogram is chosen, the best number of the nearest neighbors must also be selected. The best nearest neighbors used in the kriging technique are selected based on the smallest variance. Then, the kriging technique is applied to the residuals of the OLS

model, and a three dimensional surface is created. The overall performance of the kriged surface can be evaluated by computing an R^2 value similar to that used in regression analysis:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\varepsilon_i - \bar{\varepsilon}\right)^2}{\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)^2} \tag{4.10}$$

Where ε_i is the residual from the kriging model associated with Z_i the observed response variable. The surface is converted to an ASCII file that can be imported into ArcGIS in order to resample it to the same cell size of the Ikonos image resolution (4 meters) or the Landsat image resolution (30 meters). After generating two surfaces with the same cell size, one from the OLS model and one from the kriged residuals, they are added together in ArcGIS using the raster calculator or in ArcView using the map calculator.

4.9 Main sets and subsets of data

The observed sets of data will be tested first with the OLS model and the derived equation will be applied to the combination of bands to generate the predicted surface using the OLS model. Then out of each main set three other subsets are selected randomly and the OLS model is applied to these subsets. The three subsets are selected randomly to represent 75%, 50%, and 25% of the observed data. For each subset, an equation is derived using the OLS model using only the points of the subset and then this equation is applied to the combination of bands. The kriged residual surface is generated for only the points of each subset and then combined with the OLS surface.

4.10 Validation of model selected

For this study the modified kriging model was applied to the 2001 dataset which are fields where corn was grown that year. The analysis was done using the Aster, Landsat, and Ikonos images. The modified kriging model is also applied to the 2004 dataset which are fields where alfalfa was grown. The analysis was done using the Landsat and Ikonos images. In addition the selected model was applied to the different subsets of data which added fifteen more cases (three subsets for each of the five images). Therefore, the methodology is applied to 20 sets of data to validate the model.

CHAPTER 5. RESULTS AND ANALYSIS

This chapter is divided into six parts. The first part describes how the OLS, SAR, and modified kriging models were applied and how they compare with the observed soil salinity data in conjunction with the remote sensing images. In the second part of the chapter, the output of the three different models is discussed and the predicted values are compared with the observed data. The third part discusses the results of applying the best selected model to the five images: Aster 2001, Landsat 2001, Ikonos 2001, Landsat 2004, and Ikonos 2004 using all the data and using 75%, 50%, and 25% of all the data. The fourth part consists of a comparison of the predicted values using the Aster, Ikonos, and Landsat images for corn fields during 2001 and Ikonos and Landsat images for alfalfa fields during 2004. The fifth section is a general comparison among the five images with corn and alfalfa data. The sixth part includes the validation of the selected model and some discussion and comments.

5.1 Applying the OLS, the SAR, and the modified kriging models

This section describes the testing of the OLS, SAR, and modified kriging models with different data sets. Five cases were tested with the three models with different data sets of the observed soil salinity data in 2001 and 2004 in conjunction with the five images. The three models were applied to the following five cases:

1. The 2001 Aster image applied to fields where corn was grown that year.
2. The 2001 Landsat image applied to fields where corn was grown that year.
3. The 2001 Ikonos image applied to fields where corn was grown that year.
4. The 2004 Landsat image applied to fields where alfalfa was grown that year.

5. The 2004 Ikonos image applied to fields where alfalfa was grown that year.

5.1.1 Using stepwise regression to select the combination of bands

Stepwise regression was used to select the combination of bands that have cross correlation with the observed soil salinity data. NDVI was added to the possible set of bands of the Ikonos or Landsat images since NDVI proved to be a good indicator for assessing the quantity of vegetation. As mentioned in the previous chapter, the vegetation or the biomass of the coverage was used as an indicator of soil salinity. The following bands were selected for each image:

1. The 2001 Aster image for corn fields, bands 4, 5, 6, 8, and 9 from the short wave infrared group in addition to band 3 from the visible infrared group were selected.
2. The 2001 Landsat image for corn fields, the blue, the near infrared, the thermal, the short wave bands, and NDVI were selected.
3. The 2001 Ikonos image for corn fields, the red band and NDVI were selected.
4. The 2004 Landsat image for alfalfa fields, the red, the near infrared, the thermal band, and the NDVI were selected.
5. The 2004 Ikonos image for alfalfa fields, the green, the red, the near infrared, and NDVI were selected.

5.1.2 Applying the OLS model

In each case, the combination of bands that was selected for use in the stepwise regression was tested with the OLS model to check whether the cross correlation between the soil salinity and the image was weak or strong and to check for the existence of autocorrelation among the residuals.

For the first case, the 2001 Aster image with corn fields, bands 3, 4, 5, 6, 8, and 9 have p-values of 0, 0.0001, 0.0, 0, 0.00097, and 0 respectively which means that there is strong cross correlation with soil salinity. Bands 3, 6, and 9 have negative cross correlation with interception values of - 0.08, - 0.5, and – 0.58 respectively. Bands 4, 5, and 8 have positive cross correlation interception values of 0.34, 0.49, and 0.3 respectively. The following equation was used to create the OLS model of the predicted soil salinity using the Aster image for the fields planted with corn:

Predicted Soil Salinity = 4.07 – 0.08 * band 3 + 0.34 * band 4
+0.49 * band 5 – 0.5 * band 6+ 0.3 * band 8 – 0.58 * band 9 (5.1)

For the second case, the 2001 Landsat image with corn fields, the blue, the near infrared, the thermal, the shortwave, and NDVI have p-values of 0.0, 0.0, 0.0136, 0.0303, and 0.0006 respectively meaning that there is strong cross correlation with soil salinity. The blue, the thermal, and the NDVI have positive cross correlation with interception values of 0.20, 0.61, and 13.34 while the near infrared and the short wave bands had negative cross correlations with interception values of - 0.17 and - 0.30. The following equation was used to create the OLS model or the trend surface of the predicted soil salinity using the 2001 Landsat image for the fields planted with corn:

Predicted Soil Salinity = -35 + 0.20 * blue band – 0.17 * near infrared band
+0.61 * thermal band – 0.3 * short band + 13.34 * NDVI

(5.2)

For the third case, the 2001 Ikonos image with corn fields, the red band and NDVI have p-values of 0.0024 and 0.0 respectively which means that there is strong cross correlation with soil salinity. Both the red band and NDVI have negative cross correlation with interception values of – 0.01 and – 14.68 respectively. The following equation was used to create the OLS model or the trend surface of the predicted soil salinity using the Ikonos image for the fields planted with corn:

Predicted Soil Salinity = 15 – 0.01 * red band – 14.68 * NDVI (5.3)

For the fourth case, the 2004 Landsat image with alfalfa fields, the red, the near infrared, the thermal band, in addition to NDVI have p-values of 0.0635, 0.0, 0.0012, and 0.0007 respectively. The NDVI, the red, and thermal bands show positive cross correlation with interception values of 0.24, 0.15 and 36.86 respectively. The near infrared band has a negative cross correlation with an interception value of – 0.17 with soil salinity. The following equation was used to create the OLS model or the trend surface of the predicted soil salinity using the Landsat image for the fields planted with alfalfa:

Predicted Soil Salinity = -17.48 + 0.24 * red band -0.17 * near infrared band + 0.15 * thermal band + 36.86 * NDVI (5.4)

For the fifth case, the 2004 Ikonos image for alfalfa fields, the green, the red, the near infrared, and NDVI have p-values of 0.0001, 0, 0.0004, 0.0, and 0.0132 respectively. The red band and NDVI show positive cross correlation with interception values of 0.08 and 32.14 respectively. The green and near infrared bands show negative cross correlation with soil salinity with interception values of – 0.08 and – 0.03 respectively. The following equation was used to create the OLS model or the trend surface of the predicted soil salinity using the Ikonos image for the fields planted with alfalfa:

Predicted Soil Salinity = 21.13 – 0.08 * green band + 0.08 * red band – 0.03 * near infrared band + 32.14 * NDVI (5.5)

Table 5.1: The OLS model parameters for the five cases.

Image	RSE	R^2	AICC	p-value Moran's I	p-value Lagrange
Aster 2001	2.54	0.34	1542	0.25	0
Landsat 2001	2.27	0.47	1475	0.31	0
Ikonos 2001	1.55	0.52	959	0.00	0
Landsat 004	3.83	0.26	1458	0.61	0
Ikonos 2004	3.55	0.37	1418	0.42	0

Table 5.1 summarizes the residual standard errors RSE, multiple R^2, AICC, p-value of Moran's I and p-value of Lagrange multiplier parameters of the OLS model for the five cases. It is clear from this table that the data set collected during 2001 in fields where corn was grown shows better results than that of the data collected during 2004 where alfalfa was grown. The AICC values are smaller, the R^2 values are higher, and the residual standard errors are smaller. The p-values of Moran's I show that there is no autocorrelation among the residuals except for the 2001 Ikonos image as all

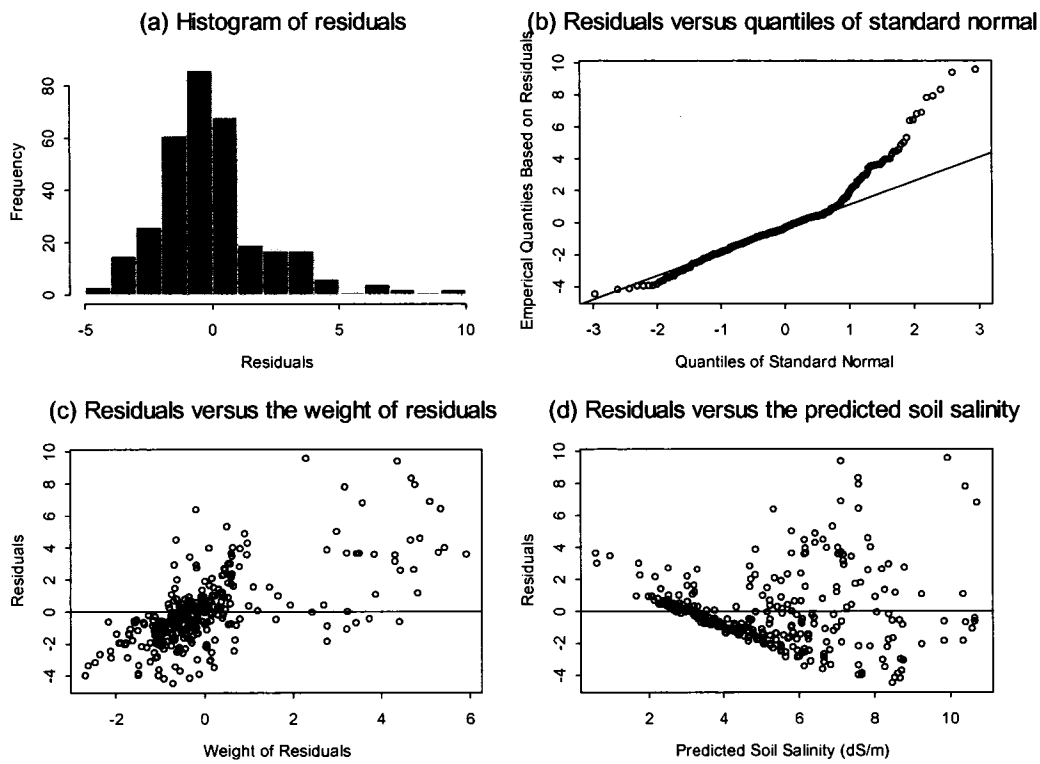


Figure 5.1: Graphical inspection of the residuals of the OLS models.

the values are larger than 0.05. However, the p-values of the Lagrange multiplier provide an indication of the existence of autocorrelation among the residuals where p-values of the four sets are less than 0.05.

Figure 5.1 shows the graphical representation of the residuals of the OLS model using the 2001 observed soil salinity data in conjunction with the 2001 Landsat

image. The two graphs at the top were used to evaluate the normality of the distribution of the residuals while the two graphs at the bottom were used to evaluate the homogeneity of the distribution of the residuals. The graph on the upper left (a) shows the histogram of residuals. This graph shows that the distribution is not entirely normal; it is not bell-shaped but skewed to the right. The upper right plot (Q-Q) graph (b) shows the empirical quantiles based on residuals versus the corresponding quantiles from a normal population. It is clear from the quantiles that all the points are very close to the line between the values -3 to 1 and that the points start to deviate from the line for values above 1. This result also indicates that the distribution of the residuals does not follow a normal distribution. The graph on the bottom left (c) shows the residuals versus the weight of the residuals. It is clear that there is a pattern and some clustering in the distribution which means that there is no homogeneity. The figure on the bottom right (d) shows the residuals versus the predicted values of soil salinity. It is clear from this figure that the distribution of points is not scattered randomly about 0, which also means that there is no homogeneity. Therefore, the presence of dependency and non homogeneity among the residuals implies that there is spatial autocorrelation among the residuals.

5.1.3 Applying the SAR model

The same sets of data were tested using the SAR model since the SAR model has the ability to remove some autocorrelation. For the first case, the 2001 Aster image applied to corn fields, the green, the near infrared, and the shortwave bands have p-values of 0.0065, 0.0, and 0.0374 respectively which means that there is strong cross correlation with soil salinity. The green band has a positive cross correlation while the near infrared and the shortwave have negative cross correlation. For the

second case, the 2001 Ikonos image for corn, the red, the near infrared bands, and NDVI have p-values of 0.0025, 0.0092, and 0.0 respectively which means that there is strong cross correlation with soil salinity. The near infrared band shows positive cross correlation while the red band and NDVI have negative cross correlation with soil salinity. For the third case, the 2004 Landsat image applied to alfalfa fields, the NDVI has a p-value of 0.18 which mean that it has weak cross correlation with soil salinity. For the fourth case, the 2004 Ikonos image for alfalfa fields, the near infrared band and NDVI have p-values of 0.0 and 0.0 which means a strong cross correlation with soil salinity. The NDVI shows positive cross correlation while the near infrared shows negative cross correlation with soil salinity.

Table 5.2: The SAR model parameters for the five cases.

Image	RSE	R^2	AICC	p-value Likelihood	λ
Aster 2001	1.81	0.05	1339	0	0.99
Landsat 2001	1.69	0.18	1300	0	0.98
Ikonos 2001	1.32	0.25	889	0	0.94
Landsat 2004	3.34	0.03	1398	0	0.94
Ikonos 2004	3.12	0.15	1366	0	0.91

Table 5.2 summarizes the residual standard errors (RSE), the multiple R^2, the AICC, the p-value of the likelihood ratio test and the λ parameters of the SAR model for the five cases. As in the OLS model, it is clear from this table that the data set collected during 2001 in fields planted with corn shows better results than that of the data collected during 2004 in fields planted with alfalfa. This is shown by the 2001 data set's smaller AICC values, higher R^2 values, and smaller residual standard errors. The values of λ and the likelihood ratio test p-values for the five cases, show that the significant autocorrelation among the residuals has been removed. The above table also shows an improvement in the AICC values and a reduction in the R^2 values, which are also results of the removal of autocorrelation among the residuals.

Figure 5.2 shows the graphical representation of the residuals for the SAR model using the observed data during 2001 for fields planted in corn in conjunction with the 2001 Landsat image. The figure on the upper right (a) shows the histogram of the residuals. With the SAR model, the distribution is getting closer to the desired bell shape than with the OLS model. This means that the SAR model was able to make the distribution of residuals closer to a normal distribution. The figure on the upper right (b) shows the quantiles. It seems that it behaves almost the same as that of the OLS model. The figure on the lower left (c) shows the residuals versus the weight of the residuals; the SAR model shows no clear pattern or trend which means that there is more homogeneity with the SAR model than with the OLS model. The figure on the lower right (d) shows the residuals versus the predicted values. This plot looks almost the same as the plot produced for the OLS model.

These results show that the SAR model was able to make some improvements over the OLS model regarding the normality and the homogeneity. These improvements remove some of the spatial autocorrelation among the residuals but not all. Furthermore, the R^2 values produced using the SAR model are not encouraging, indicating that the model would poorly predict soil salinity data using remote sensing data. Therefore, there is a need to test the other technique, the modified kriging model.

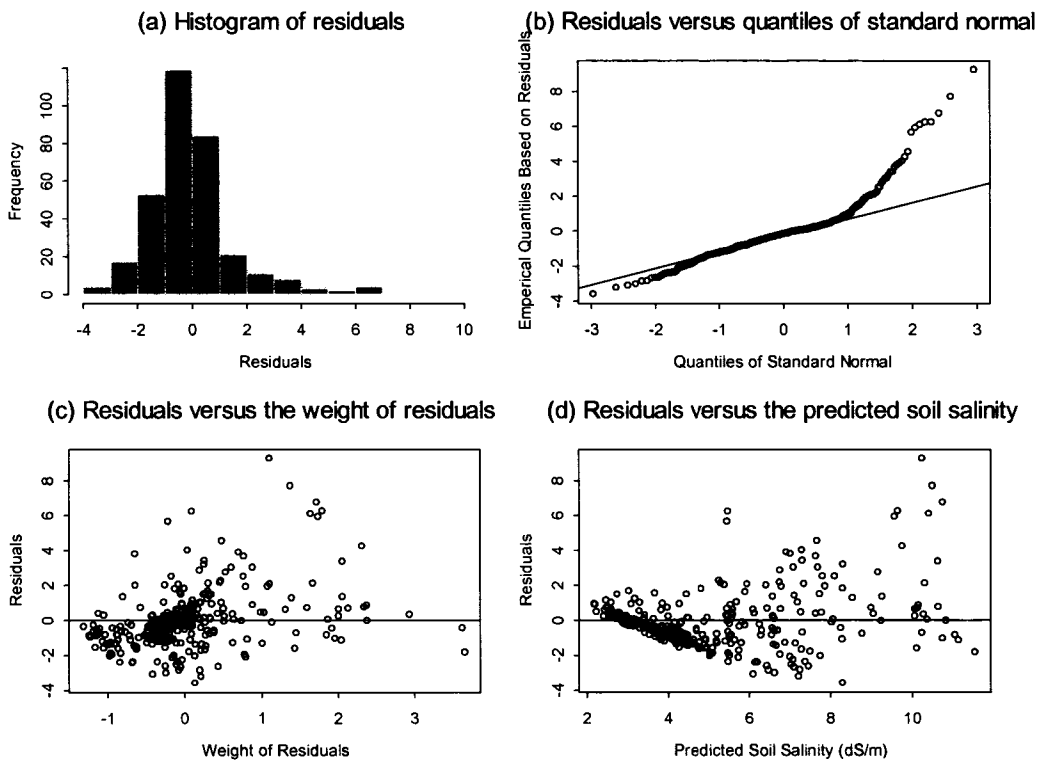


Figure 5.2: Graphical inspection of the residuals of the SAR models.

5.1.4 Applying the modified kriging model

The same sets of data that were tested previously using the OLS and the SAR models were tested using the kriging technique model. The kriging technique depends on the presence of autocorrelation or dependency among the data being kriged. Since the SAR model was able to remove most of the dependency or autocorrelation among the residuals, kriging the residuals of the SAR model would not be effective, consequently the OLS model was used with the kriging model in this research. In short, although the presence of autocorrelation among the residuals is considered a disadvantage when using the OLS model by itself, it can be considered an advantage when using the kriging technique. To apply the modified kriging model, the residuals of the OLS model were kriged and then combined with the surfaces produced by the OLS model alone.

To apply the kriging technique efficiently, three main issues should be considered. First, there must be dependency or autocorrelation among the residuals of the model. Second, a suitable variogram must be selected which best fits the data; this variogram should be selected based on the smallest AICC value. Third, the nearest number of neighbors of the selected variogram should be selected based on the smallest variance. In addition to these criteria, a reasonable cell size should be used. In this study, the selected cell size is the same as the resolution of the image being considered.

Figure 5.3 shows an example of the different variograms used in the kriging technique for the 2004 observed data in conjunction with the 2004 Ikonos image. It is clear from the figure that the exponential model is the closest to the points while the gaussian and the spherical models deviate from the data. In most of the cases tested, the exponential model performed the best. For the three variogram models, as the data decreases, the range becomes larger, which can clearly be observed in figure (d). This means that when the data decreases, the correlation between points decreases, which results in a decreasing the fit of the variograms to the data. Table 5.3 shows the parameters of the three variograms for the main set of data with the five different cases for the three models.

Table 5.3 shows the values of the nugget, the sill, the range, the standard error (s.e), and the AICC for the exponential, the gaussian, and the spherical models for the five cases. The AICC of the exponential model has the smallest value among the three variograms which makes it the best choice. There is no significant difference among the standard error values of the three models. From the values of the AICC, the standard error and the nugget it can be determined that the modified kriging model

behaves better for the Landsat and Ikonos images with the 2001 data set for corn crops than with the 2004 data set with alfalfa crops.

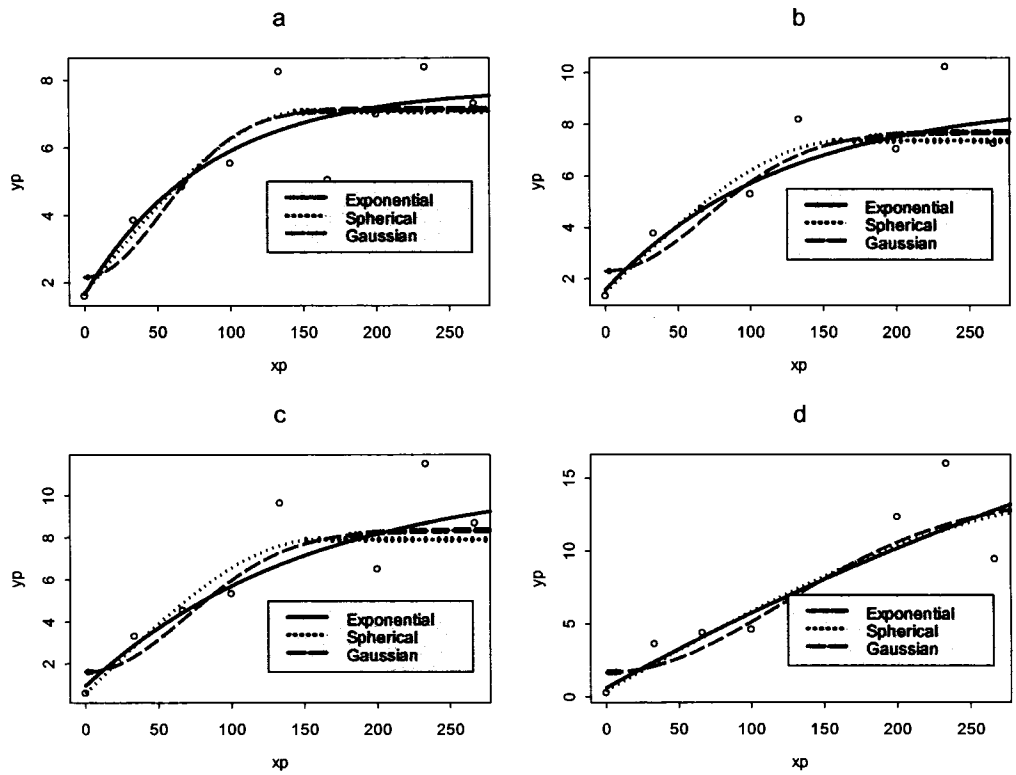


Figure 5.3: The variogram models used to krig the residuals of all data, 75% of the data, 50% of the data, and 25% of the data using the 2004 Ikonos image.

Table 5.3: Parameters of the Gaussian, Spherical, and Exponential models for the five cases.

	Model	Nugget	Sill	Range	s.e.	AICC
Aster 2001	Gaussian	1.8	4.5	132	2.11	27.7
	Spherical	1.6	4.5	174	2.12	21.9
	Exponential	1.4	4.7	71.1	2.16	18.23
Landsat 2001	Gaussian	1.8	4.3	134	2.1	27
	Spherical	1.7	4.4	185	2.1	21.7
	Exponential	1.4	4.6	78	2.1	15
Ikonos 2001	Gaussian	0.5	2.4	73	1.6	26.2
	Spherical	0.5	2.5	124	1.6	24.7
	Exponential	0.4	2.5	41	1.6	23.9
Landsat 2004	Gaussian	3.1	10.6	242	3.3	47.3
	Spherical	2.4	10.7	304	3.3	44.4
	Exponential	2.2	15.1	250	3.9	44.4
Ikonos 2004	Gaussian	2.2	7.1	126	2.7	48.9
	Spherical	1.8	7.1	149	2.7	47.8
	Exponential	1.7	7.6	77	2.8	47.4

5.2 Comparing the observed and the predicted data using the OLS, the SAR, and the modified kriging models

In this section a comparison between the observed and the predicted data for the three cases is discussed. The mean values of the observed data as well as the predicted data using the three models are compared. In some cases the mean is not a strong parameter for comparisons; therefore, another parameter, which is the mean absolute error, was also considered. In addition, all observed points were compared against all the predicted values of the three models using scatter plots. Some visual examples displaying the raster maps of the observed versus the predicted data are presented as well.

5.2.1 Observed and predicted soil salinity in fields planted with corn

5.2.1.1 2001 Aster image for detecting soil salinity in fields planted with corn that year

This case represents the analysis of the data observed during 2001 pertaining to the 2001 Aster image for fields planted with corn that year. This set of data consists of 328 observation points for four fields (US09, US10, US40, and US80).

Table 5.4: Mean and mean absolute errors values (dS/m) of the predicted soil salinity from the 2001 Aster image for fields planted with corn using the OLS, the SAR, and the modified kriging models.

Field #	Observed Mean	OLS model		SAR model		Modified kriging model	
		Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.55	0.91	2.52	0.67	3.03	0.48
US10	7.15	6.89	3.53	2.78	4.43	7.22	1.45
US40	6.19	5.80	1.70	2.98	3.23	6.15	0.96
US80	5.09	4.95	1.53	2.31	2.82	5.04	0.74
All Fields	5.36	5.30	1.92	2.65	2.79	5.36	0.91

Table 5.4 shows the mean and mean absolute error values of the predicted soil salinity from the 2001 Aster image for corn crops using the OLS, the SAR, and the

modified kriging models. From the table it can be determined that there are no significant differences in the mean values of all the models compared with the mean values of the observed data except in the case of the SAR model. The SAR model failed to predict the mean. However, the mean absolute errors of the different models differ significantly. The modified kriging model has the smallest value for the absolute mean errors both for individual fields and for all the fields combined. The OLS and the SAR models generate similar results and show reasonable values except for US10, which is a very small field with high variability. Table 5.4 also shows that US09 has the smallest mean absolute error amongst the fields; one possible explanation for this result is that US09 is a very homogenous field.

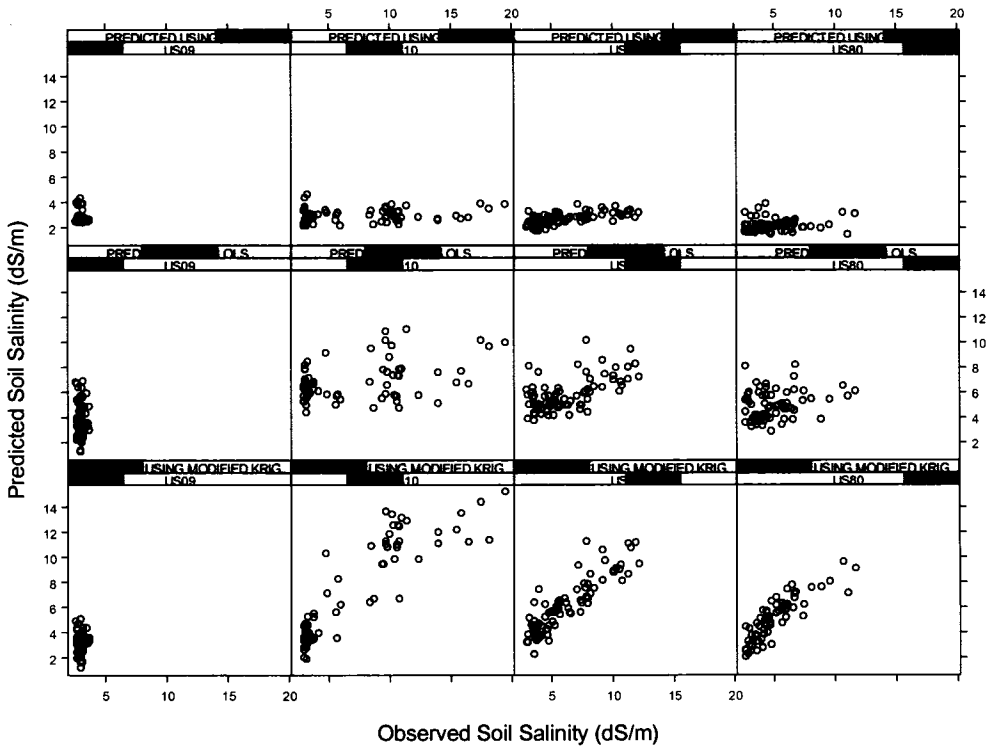


Figure 5.4: Comparison of predicted soil salinity (dS/m) from the 2001 Aster for the fields planted with corn for the OLS, the SAR, and the modified kriging models versus the observed values.

Figure 5.4 shows the comparison among the predicted soil salinity values using the three models (OLS, SAR, and modified kriging models) for the 2001 Aster image. This figure clearly shows that there is a significant improvement in the value of R^2 when using the modified kriged model over using the OLS and the SAR models. The average R^2 values of all fields were 0.34 for the OLS model, 0.05 for the SAR model, and 0.88 for the modified kriging model. The modified kriging model produces predicted points with a clearer trend and less scatter than those of the OLS and the SAR models.

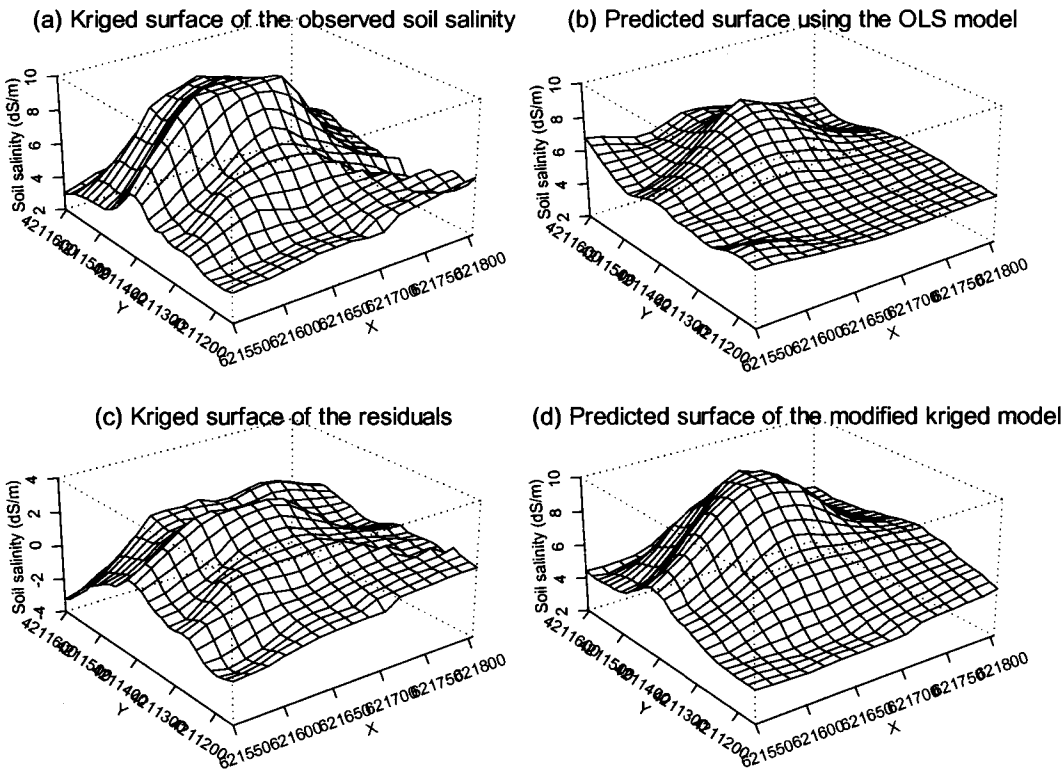


Figure 5.5: Generated surfaces for field US40 using the 2001 data in conjunction with the 2001 Aster image.

Figure 5.5 shows an example of the soil salinity surface for field US40 using with the data observed in 2001 in conjunction with the 2001 Aster image. The OLS model, shown in figure (b), behaves as a trend surface: the prediction of high soil

salinity values was underestimated while the prediction of the low soil salinity values was overestimated. Figure (c) represents the kriged surface of the residuals. Those residuals have positive values if they are above the OLS surface and have negative values if they are below the OLS surface. This kriged surface of the residuals can be used to correct the overestimated and the underestimated values generated using the OLS model. Therefore, when surfaces (b) and (c) are combined to produce figure (d), the resulting predictions are very close to the observed salinity measurements.

5.2.1.2 2001 Landsat image for detecting soil salinity in fields planted with corn that year

This case represents the analysis of the data observed during 2001 pertaining to the 2001 Landsat image for fields planted with corn that year. This set of data consists of 328 observation points for four fields (US09, US10, US40, and US80).

Table 5.5: Mean and mean absolute errors values (dS/m) of the predicted soil salinity from the 2001 Landsat image for fields planted with corn using the OLS, the SAR, and the modified kriging models.

Field #	Observed Mean	OLS model		SAR model		Modified kriging model	
		Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.33	0.70	3.13	0.46	3.08	0.34
US10	7.15	6.23	2.55	3.90	3.44	6.99	1.18
US40	6.19	6.58	1.92	4.92	1.82	6.14	0.93
US80	5.09	5.02	1.63	3.84	1.57	4.97	0.81
All Fields	5.36	5.29	1.70	3.94	1.82	5.29	0.82

Table 5.5 shows the mean and mean absolute error values of the predicted soil salinity from the 2001 Landsat image for corn crops using the OLS, the SAR, and the modified kriging models. From the table we can determine that there are no significant differences in the values of the mean of all the models compared with the mean of the observed data. However, there are significant differences between the mean absolute errors produced by the different models. Whether an individual field is

being considered or whether all the fields are looked at in combination, the modified kriging model shows the smallest values for the mean absolute errors. The OLS and the SAR models show reasonable values except for US10, again reflecting the small area and high variability of that field. US09 has the smallest mean absolute errors among the rest of the fields, and one possible explanation for this is the fact that this field is homogenous.

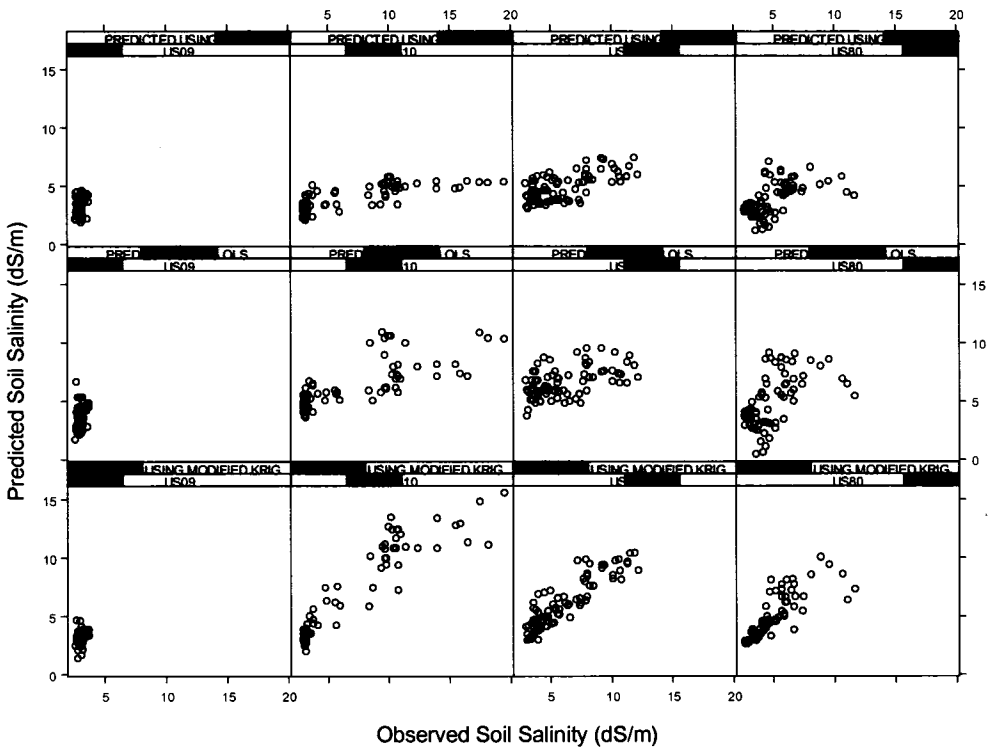


Figure 5.6: Comparison of predicted soil salinity (dS/m) from the 2001 Landsat for the fields planted with corn for the OLS, the SAR, and the modified kriging models versus the observed values.

Figure 5.6 shows the comparison among the predicted soil salinity values using the three models (OLS, SAR, and modified kriging models) using the 2001 Landsat image. This figure clearly shows that there is a significant improvement in the value of R^2 when using the modified kriging model over using the OLS and the SAR models. The R^2 values were 0.47 for the OLS model, 0.18 for the SAR model,

and 0.89 for the modified kriging model. The modified kriging model produces predicted points with a clearer trend and less scatter than those of the OLS and the SAR models.

5.2.1.3 2001 Ikonos image for detecting soil salinity in fields planted with corn that year

This section analyzes the 2001 observed data pertaining to the 2001 Ikonos image. This set of data consists of 257 observation points for three fields (US09, US40, and US80). Field US10 was not covered by the Ikonos image in 2001.

Table 5.6: Mean and mean absolute (dS/m) error values of the predicted soil salinity from the 2001 Ikonos image for the fields planted with corn using the OLS, the SAR, and the modified kriging models.

Field #	Observed Mean	OLS model		SAR model		Modified kriging model	
		Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.25	0.70	3.41	0.97	3.03	0.52
US40	6.13	5.86	1.81	5.00	1.96	6.12	0.58
US80	5.09	5.00	1.31	3.95	1.51	5.04	0.38
All Fields	4.75	4.70	1.28	4.12	1.48	4.73	0.49

Table 5.6 shows the mean and mean absolute error values of the predicted soil salinity using the 2001 Ikonos image for fields planted with corn using the OLS, the SAR, and the modified kriging models. The table shows that the absolute mean error values for each individual field and for all fields together resulting from the modifying kriging technique are the smallest of all the models. The OLS model performs second best, and the SAR model performs the worst. As in the previous case, field US09 shows the smallest mean absolute error values for the OLS model and the SAR model.

Figure 5.7 shows the predicted values using the three models (OLS, SAR, and kriging) versus the observed values of soil salinity when using the 2001 Ikonos image.

This figure clearly shows that there is significant improvement in the R^2 value when using the modified kriging model over using either the OLS or SAR models

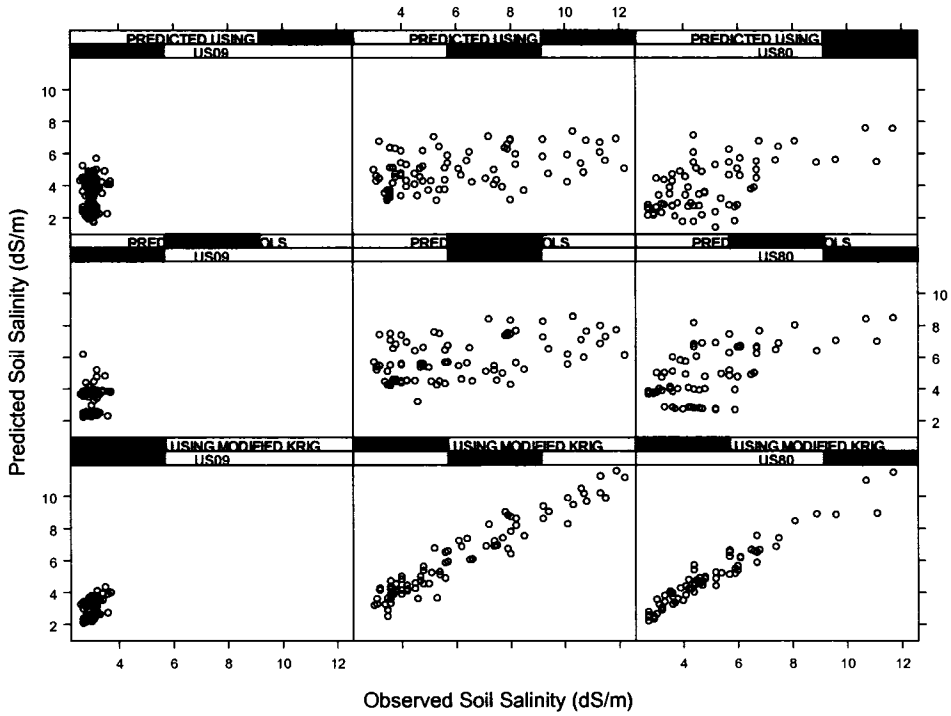


Figure 5.7: Comparison of the predicted soil salinity (dS/m) from 2001 Ikonos image for fields planted with corn using the OLS, the SAR, and the modified kriging models versus the observed values.

alone. The R^2 values were 0.45 for the OLS model, 0.26 for the SAR model, and 0.88 for the modified kriging model. The modified kriging model produces predicted points with a clearer trend and less scatter than those of the OLS and the SAR models.

5.2.2 Observed and predicted soil salinity for fields planted with alfalfa

5.2.2.1 2004 Landsat image for detecting soil salinity in fields planted with alfalfa that year

This section analyzes the data observed in 2004 in conjunction with the Landsat image for fields planted with alfalfa that year. This data set consists of 263 observation points for four fields US14, US04, US09, and US10.

Table 5.7: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2004 Landsat image for fields planted with alfalfa using the OLS, the SAR, and the modified kriging models.

Field #	Observed Mean	OLS model		SAR model		Modified Kriging model	
		Mean	MAE	Mean	MAE	Mean	MAE
US14	4.71	5.79	1.57	5.98	1.38	4.68	0.68
US04	9.29	7.92	4.98	5.36	5.36	9.39	1.88
US09	2.63	3.20	0.91	6.08	3.45	2.57	0.40
US10	6.09	6.12	3.06	5.84	3.01	6.15	1.64
All Fields	5.68	5.76	2.63	5.82	3.30	5.70	1.15

Table 5.7 shows mean and mean absolute error values of the predicted soil salinity using the 2004 Landsat image for fields planted with alfalfa that year using the OLS, the SAR, and the modified kriging models. The table shows that mean values predicted by the modified kriging model are the closest to the observed data. The mean values predicted by the OLS model are also reasonably close, but the values predicted by the SAR model are very poor for US04 and US09 and acceptable for US14 and US09. This means that the SAR model tends to have values close to average for all fields. The modified kriging model has the smallest mean absolute error for the individual fields and for all fields combined and, therefore, is the best model.

Figure 5.8 shows the predicted values of the three spatial models (OLS, SAR, and kriging models) versus the observed values of soil salinity when using the 2004 Landsat image. The figure shows that there is a significant improvement in the R^2

value when using kriged residuals combined with the OLS model over using the OLS or SAR models alone. The R^2 values were 0.16 for the OLS model, 0.02 for the SAR model, and 0.80 for the modified kriging model. The modified kriging model produces predicted points with a clearer trend and more scatter than those of the OLS and the SAR models. The above table shows the significant contribution resulting from adding the kriged residuals to the OLS model.

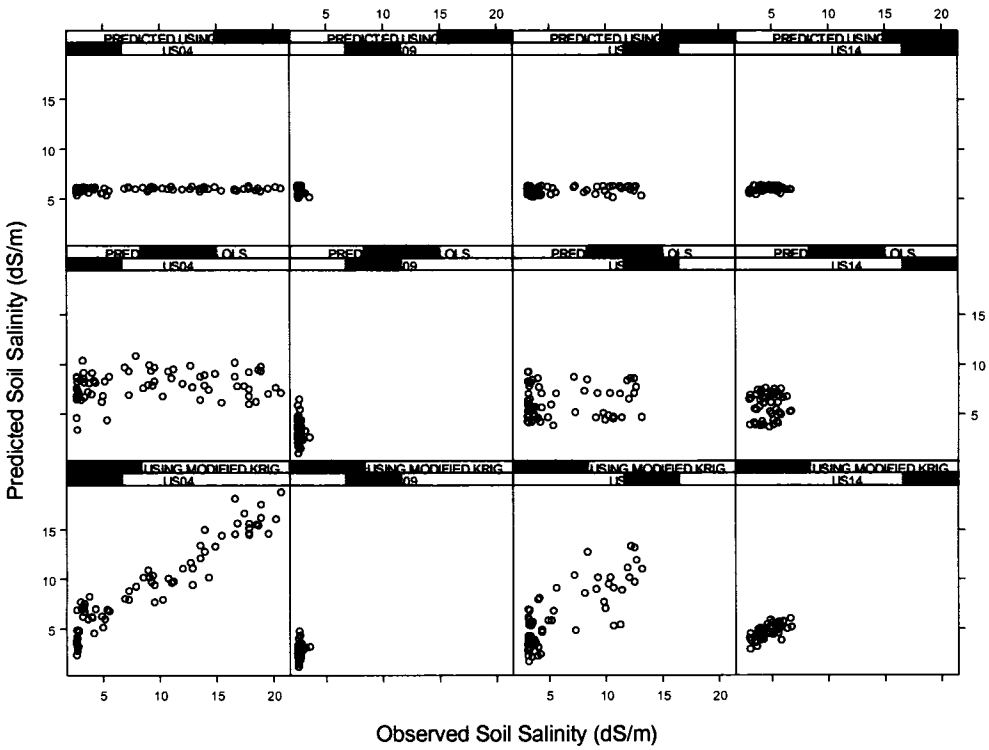


Figure 5.8: Comparison of the predicted soil salinity (dS/m) from the 2004 Landsat image for fields planted with alfalfa using the OLS, the SAR, and the modified kriging models versus the observed values.

5.2.2.2 2004 Ikonos image for detecting soil salinity in fields planted with alfalfa that year

This section analyzes the data observed in 2004 in conjunction with the 2004 Ikonos image for fields planted with alfalfa that year. This set of data consists of 263 observation points for four fields US14, US04, US09, and US10.

Table 5.8: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2004 Ikonos image for fields planted with alfalfa that year using the OLS, the SAR, and the modified kriging models.

Field #	Observed Mean	OLS model		SAR model		Modified Kriging model	
		Mean	MAE	Mean	MAE	Mean	MAE
US14	4.71	5.89	1.40	5.92	1.25	4.71	0.45
US04	9.43	8.33	3.61	7.05	4.80	9.42	1.16
US09	2.63	3.52	1.01	4.92	2.29	2.60	0.22
US10	6.13	5.35	1.94	5.51	2.40	6.18	1.04
All Fields	5.73	5.77	1.99	5.85	2.68	5.73	0.72

Table 5.8 shows the mean and mean absolute error values of the predicted soil salinity for the 2004 Ikonos image for the fields planted with alfalfa that year using the OLS, the SAR, and the modified kriging models. The OLS model presents reasonable values regarding the mean and the absolute mean errors for field US04. The SAR model provides some reasonable values regarding the mean soil salinity value while it deviates significantly for the mean absolute errors. The modified kriging model has the smallest mean absolute errors at the field scale and when applied to all the fields combined.

Figure 5.9 shows the predicted values of the three spatial models (OLS, SAR, and kriging models) versus the observed values of soil salinity when using the Ikonos image acquired in 2004. The figure shows that there is a significant improvement in the R^2 value when using kriged residuals combined with the OLS model over using the OLS or SAR models alone. The R^2 values were 0.31 for the OLS model, 0.15 for the SAR model, and 0.83 for the modified kriging model. The modified kriging model produces predicted points with a clearer trend and less scatter than those of the OLS and the SAR models. Figure 5.12 shows the significant contribution resulting from adding the kriged residuals to the OLS model.

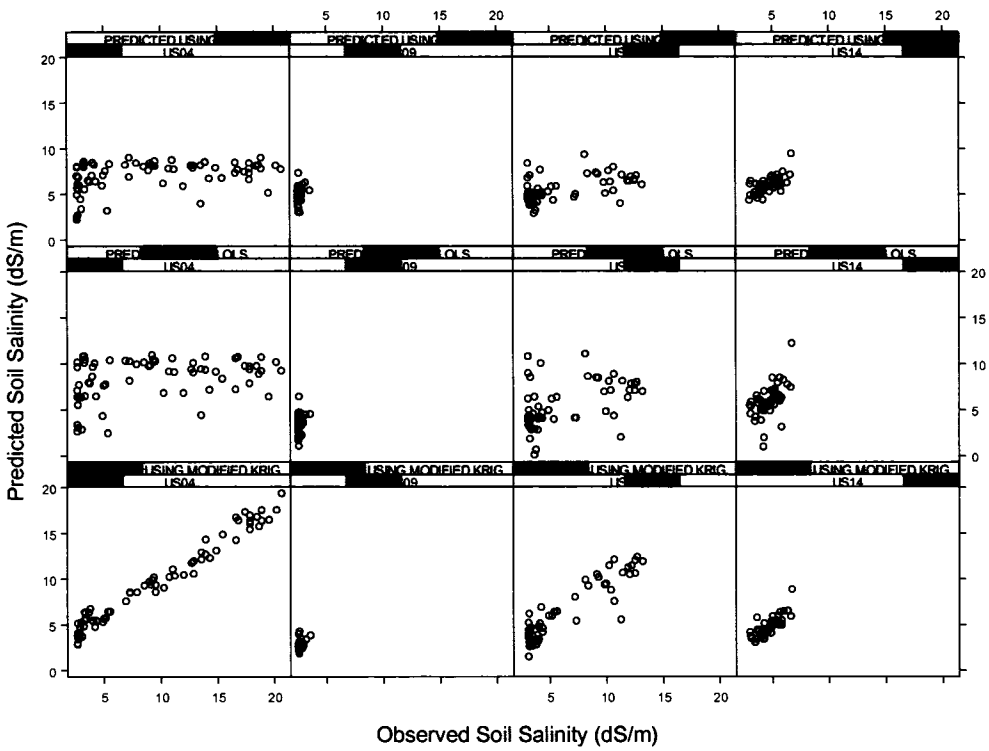


Figure 5.9: Comparison of the predicted soil salinity (dS/m) from the 2004 Ikonos image for fields planted with alfalfa using the OLS, the SAR, and the modified kriging models versus the observed values.

5.3 Applying the modified kriging model

The modified kriging model is applied to all the observed data (main set) in each of the five cases discussed previously. As part of validating the model, the modified kriging model is also applied to randomly selected subsets of data taken from the main set. In addition to helping to validate the model, this strategy also illustrates the efficacy of reducing the number of observed samples. The modified kriging model was tested with subsets of the whole set of the observed data. Three different subsets were tested, the first subset contains 75% of all data, the second subset contains 50 % of all data, and the third subset contains 25% of all data. The OLS model was tested and the equation derived from the model was applied to the whole set of data. A trend surface was generated based on the OLS model equation.

Then, a kriged surface was generated based on the OLS model residuals and added to the OLS model surface. The following section shows some statistical analysis followed by some examples of comparing the observed and the predicted data using the whole data set and the different subsets of the data for the 2004 Ikonos image.

5.3.1 2001 Aster image and for fields planted with corn that year

This section analyzes the data observed in 2001 in conjunction with the 2001 Aster image for fields planted with corn that year. The main set of data consists of 328 observation points for four fields US09, US10, US40, and US80. The three subsets consist of 75%, 50%, 25% of the main data which are 246, 164, and 82 points for the first, second, and third subset respectively.

Table 5.9: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Aster image for fields planted with corn using the all data, 90% of data, 50% of data, and 25% of data.

Field #	Observed Mean	All data		75% of data		50% of data		25% of data	
		Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.04	0.48	3.19	0.56	3.41	0.86	3.57	1.31
US10	7.15	7.21	1.45	7.38	1.45	7.56	1.87	7.39	2.51
US40	6.19	6.15	0.95	6.22	1.04	6.73	1.35	6.50	1.56
US80	5.09	5.04	0.74	5.10	0.82	5.72	1.13	6.30	1.93
All Fields	5.36	5.36	0.91	5.47	0.97	5.86	1.30	5.94	1.83

Table 5.9 shows the mean and mean absolute error values of the predicted soil salinity from the 2001 Aster image for fields planted with corn that year using the OLS, the SAR, and the modified kriging models for different sets of data. The mean values are almost the same for the whole data set, 75% of the data, 50% of the data, and 25% of the data. Using all the data reduces the mean absolute error to its smallest number, and the mean absolute error increases as the data sets decrease in size for both the individual fields and for all the fields combined. For fields with less variation such as US09 the results of using subsets are encouraging, but for US10, with its

small area and wide range of soil salinity values, the subsets of data are not sufficient to capture the variability.

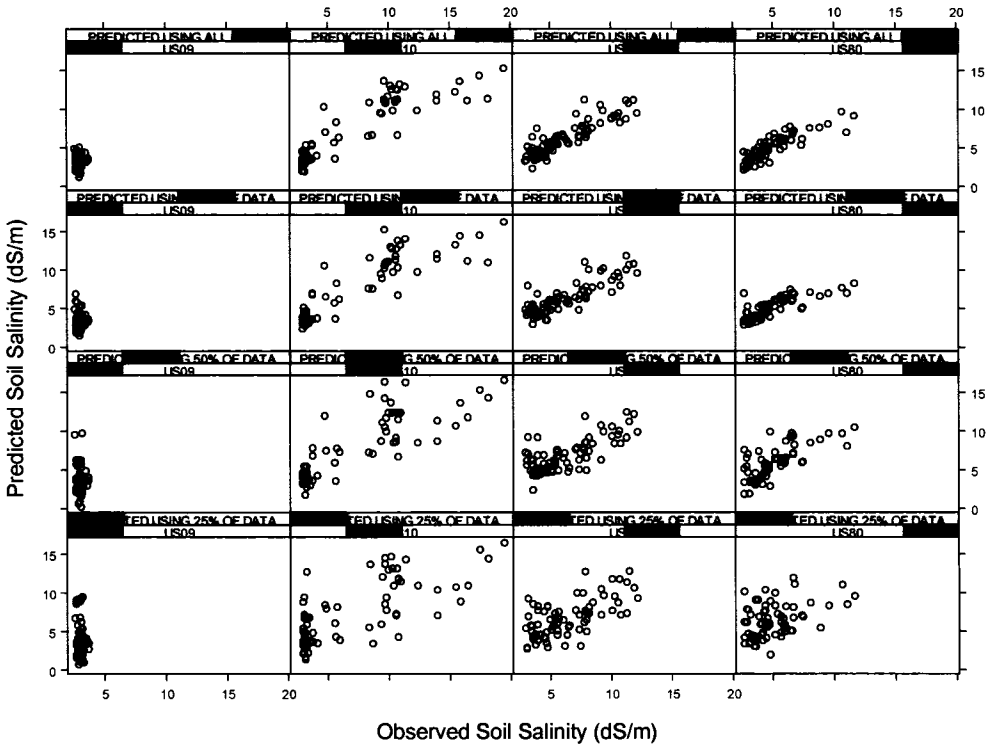


Figure 5.10: Predicted soil salinity (dS/m) vs. observed using from the 2001 Aster image for fields planted with corn usig all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data.

Figure 5.10 shows a comparison of the observed soil salinity versus the predicted values using the modified kriging model for all the data (328 points), 75% of data (246 points), 50% of data (164 points), and 25% of data (82 points). The R^2 of the predicted values was 0.81 for all the data, 0.72, 0.61, and 0.39 for 75%, 50%, and 25% of data respectively. Figure 5.10 shows that whole dataset has a clear trend and that it has less scatter than the three subsets. As expected there is significantly more scatter in the results of the reduced data sets, but there is still good predictive capability in these models as shown in the figure of US 40 where 75% and 50% of the data is used.

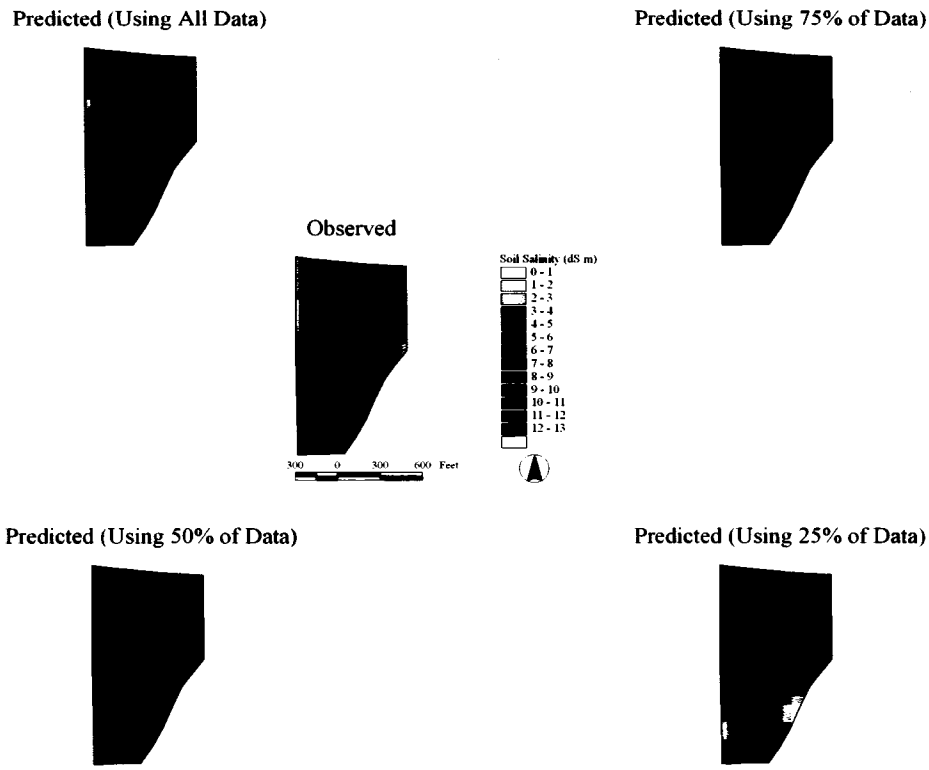


Figure 5.11: Observed and predicted surfaces of soil salinity (dS/m) of field US40 from the 2001Aster image for fields planted with corn using all the observed data, 75% of all the observed data, 50% of all the observed data, and 25% of all the observed data.

Figure 5.11 shows an example of soil salinity surface for US40 using the modified kriging model with all the data, as well as 75%, 50% and 25% of the data. The data set consists of 79 points for US40. The subsets of 75%, 50% and 25% of the data consist of 59, 40, and 20 points respectively. It is clear that very similar surfaces result from using all the data and using 75% of the data, and both of these surfaces are close to the observed data. The MAE in table 5.9, also shows that the using 75% of the data produces similar results to using the whole data set: the MAE is 0.95 when using all of the data and 1.04 when using 75% of the data. The two surfaces generated by using 50% and 25% of the data closely resemble each other but do not match the observed data as closely as the surfaces generated by using all the data and 75% the of data.

5.3.2 2001 Landsat image and different sets of data for the fields planted with corn in that year

This section analyzes the data observed in 2001 in conjunction with the 2001 Landsat image for fields planted with corn that year. The main set of data consists of 328 observation points for four fields US09, US10, US40, and US80. The three subsets consist of 75%, 50%, and 25% of the whole data set, 246, 164, and 82 data points respectively.

Table 5.10: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Landsat image for fields planted with corn using the OLS, the SAR, and the modified kriging models for different sets of data.

Field #	Observed Mean	All data		75% of data		50% of data		25% of data	
		Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.12	0.34	2.92	0.79	3.04	0.57	3.22	0.98
US10	7.15	6.95	1.22	7.16	1.41	7.44	2.19	6.77	2.68
US40	6.19	6.09	0.96	6.56	1.42	5.91	1.49	6.91	2.54
US80	5.09	4.93	0.84	5.15	1.09	5.35	1.17	5.29	1.65
All Fields	5.36	5.27	0.84	5.45	1.18	5.44	1.35	5.55	1.96

Table 5.10 shows the mean and mean absolute error values of the predicted soil salinity from the 2001 Landsat image for fields planted with corn that year using the OLS, the SAR, and the modified kriging models for different sets of data. The table shows that the mean absolute error generated using all the data is the smallest. As the subsets decrease in size, the MAE increases for all fields except for field US09 where the MAE of 50% of the data is smaller than that produced when using 75% of data. One of the reasons for this behaviour could be that this field is very homogenous and since the subsets are picked randomly, maybe the 50% subset was more representative of the variability of this field. As with the previous study using the Aster image, field US10 with its small area and large range of soil salinity values, is not well represented by the subsets.

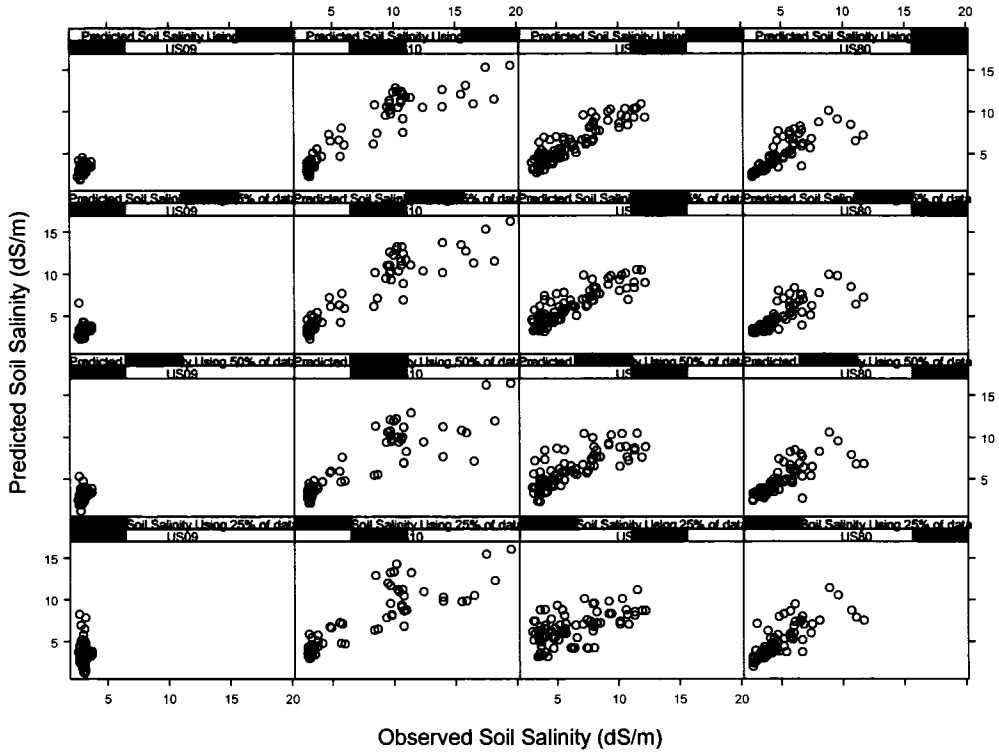


Figure 5.12: Predicted soil salinity (dS/m) vs. observed using from the 2001 landsat image for fields planted with corn using all the observed data, 75% of all the observed data, 50% of all the observed data, and 25% of all the observed data.

Figure 5.12 shows a comparison of the observed soil salinity versus the predicted values using the modified kriging model for all the data (328 points), 75% of data (246 points), 50% of data (164 points), and 25% of data (82 points). The R^2 of the predicted values were 0.83 for all the data and 0.74, 0.63, and 42 for the 75%, 50%, and 25% of the data respectively. Figure 5.12 shows that the dataset that includes all the data has a clear trend and that it has less scatter than the other three subsets. As expected there is significantly more scatter in the results of the reduced data sets, but there is still good predictive capability in these models. As in the previous case when using the Aster image, fields US40 and US80 show acceptable results when using 75% and 50% of the data.

5.3.3 2001 Ikonos image and different sets of data for fields planted with corn that year

This section analyzes the data observed during 2001 in conjunction with the Ikonos image for fields planted with corn that year. The complete data set consists of 257 observation points for four fields US09, US40, and US80. Field US10 was not covered by this image which makes the main data set smaller than those used with the Aster and Landsat images. The three subsets consist of 75%, 50%, and 25% of the main data which are 192, 128, and 64 sample points, respectively.

Table 5.11: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Ikonos image for fields planted with corn using the OLS, the SAR, and the modified kriging models for different sets of data.

Field #	Observed Mean	All data		75% of data		50% of data		25% of data	
		Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
US09	3.02	3.03	0.52	3.09	0.48	3.11	0.46	3.10	0.74
US40	6.19	6.17	0.62	6.19	0.66	6.82	1.02	6.91	1.27
US80	5.09	5.05	0.38	5.01	0.47	5.19	1.07	5.67	1.07
All Fields	4.77	4.75	0.51	4.76	0.54	5.04	0.85	5.23	1.03

Table 5.11 shows mean and mean absolute error values of the predicted soil salinity from the 2001 Ikonos image for fields planted with corn that year using the OLS, the SAR, and the modified kriging models. Comparing the results of this table and the previous two tables for the Aster and Landsat images, it is clear that in this case, reducing the number of observed points does not significantly reduce the accuracy of the results. The MAE values of the smallest subset of data (25% of all points) are even reasonably small.

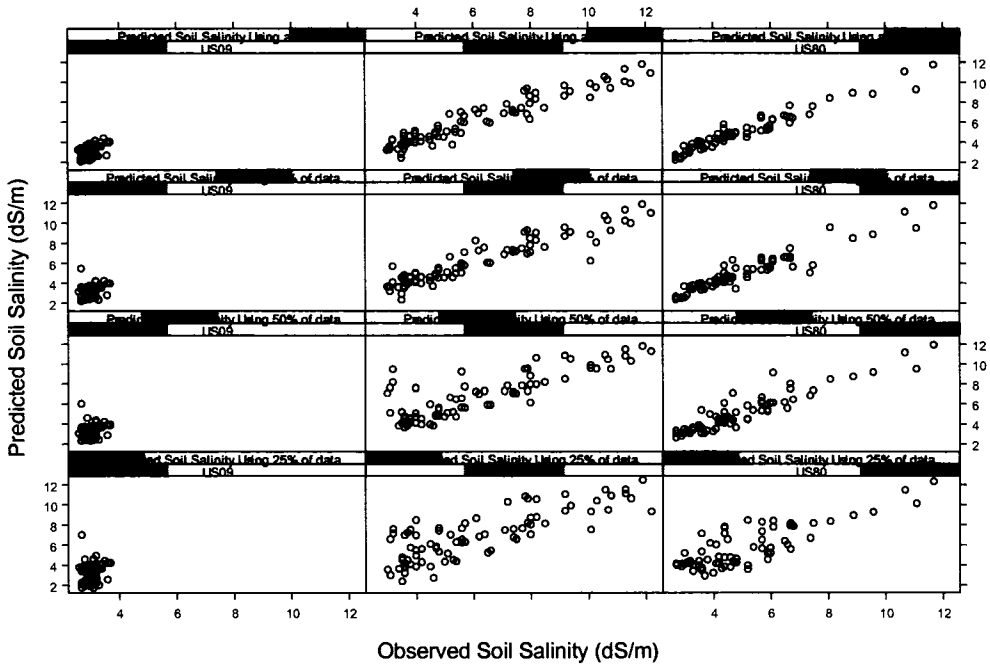


Figure 5.13: Observed versus predicted soil salinity (dS/m) for corn fields using the 2001 Ikonos image for fields planted with corn using all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data.

Figure 5.13 shows a comparison of the observed versus predicted soil salinity (dS/m) for corn fields applied to the 2001 Ikonos image using all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data. The first panel of the figure, which shows the scatter plots resulting from using all the data, has an average R^2 value of 0.91. The second panel, which shows the scatter plots generated by using 75% of the data, has an average R^2 of 0.82. The third panel showing the scatter plots generated by using 50% of the data, has an average R^2 of 0.74, and the bottom panel, which shows the scatter plots produced by using 25% of data, has an average R^2 of 0.60. It is clear that in this case the results are much better than the previous two cases when using Aster and Landsat images with the same sets

of data. This means that the observed data can be reduced dramatically in fields like US40 and US80.

5.3.4 2004 Landsat image and different sets of data for fields planted with alfalfa that year

This section analyzes the data observed in 2004 in conjunction with the 2004 Landsat image for fields planted with alfalfa. The whole set of data consists of 256 observation points for four fields US14, US04, US09, and US10. The three subsets consist of 75%, 50%, and 25% of the main data which are 192, 128, and 64 data points, respectively.

Table 5.12: Mean and mean absolute error values (dS/m) of the predicted soil salinity from the 2004 Landsat image for the fields planted with alfalfa using the OLS, the SAR, and the modified kriging models for all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data.

Field #	Observed Mean	All data		75% of data		50% of data		25% of data	
		Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
US04	9.43	9.52	1.94	9.67	2.70	9.85	3.45	12.96	4.73
US09	2.63	2.57	0.38	2.47	0.49	2.79	0.61	3.01	0.88
US10	6.13	6.21	1.65	6.52	1.75	5.95	1.85	7.86	2.91
US14	4.71	4.67	0.69	4.69	0.76	4.83	0.95	5.33	1.33
All Fields	5.73	5.74	1.16	5.84	1.43	5.86	1.71	7.29	2.46

Table 5.12 shows the mean and mean absolute error values of the predicted soil salinity for the 2004 Landsat image for the fields planted with alfalfa using the OLS, the SAR, and the modified kriging models for different sets of data. Table 5.12 strongly supports the conclusion from the previous cases that the subsets fails to capture all the variability when the fields are high variable. This can clearly be seen in the values of the mean absolute errors of fields US04 and US10 which have high variability. However the differences are not significant when using all the data or the 75%, 50% and 25% subsets for the fields with a small range of variability such as

US14 and US09. This means that the data collected for these two fields can be reduced significantly with acceptable results.

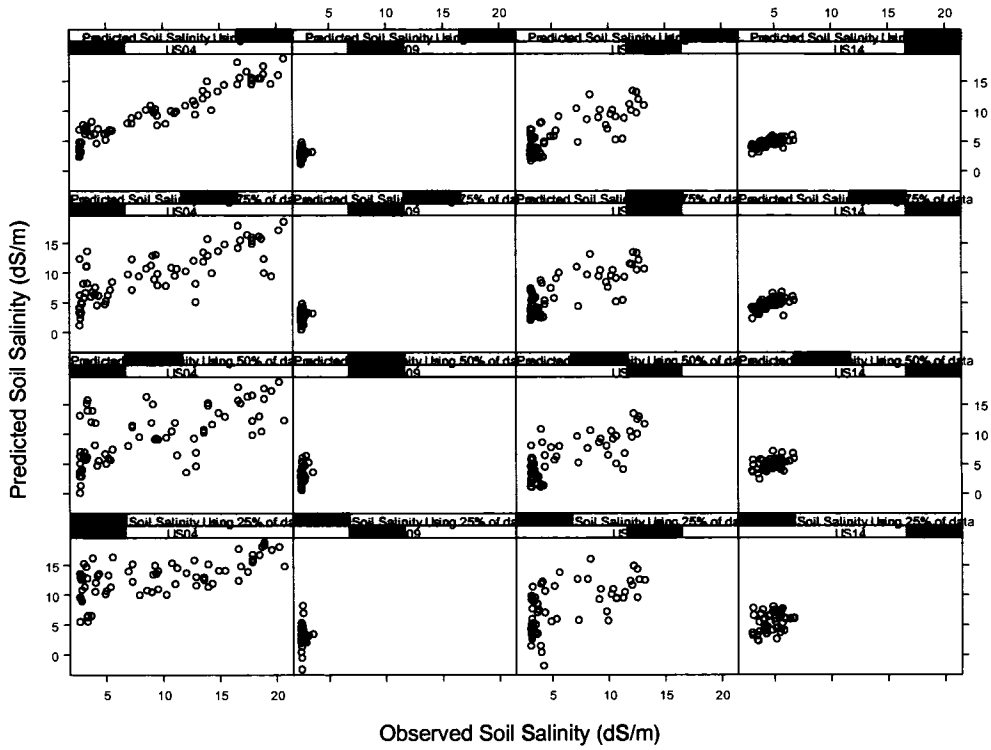


Figure 5.14: Observed versus predicted soil salinity (dS/m) for alfalfa fields using the 2004 Landsat image for fields planted with alfalfa using all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data.

Figure 5.14 shows the observed versus predicted soil salinity (dS/m) for alfalfa fields using the 2004 Landsat image for fields planted with alfalfa using all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data. The first panel of the scatter plot which shows when using all data have an average R^2 value of 0.60. The second panel which shows the scatter plots when using 75% of data have an average vv of 0.49. The third panel which shows the scatter plots when using 50% of data have an average R^2 of 0.42. The panel on the bottom which shows the scatter plots when using 25% of data shows an average R^2 of

0.19. It is obvious that using all data give the best results and 25% of give the worst results. It is obvious when comparing these results with the similar image of Landsat when it was used with corn that Landsat with corn give better results than with alfalfa.

5.3.5 2004 Ikonos image and different sets of data for the fields planted with alfalfa

This section analyzes the data observed in 2004 in conjunction with the 2004 Landsat image for fields planted with alfalfa. The data set consists of 256 observation points for four fields US14, US04, US09, and US10. The three subsets consist of 75%, 50%, and 25% of the main data which are 192, 128, and 64 sample points, respectively.

Table 5.13: Mean and mean absolute errors values (dS/m) of the predicted soil salinity from the 2004 Ikonos image for fields planted with alfalfa using the OLS, the SAR, and the modified kriging models for different sets of data.

Field #	Observed Mean	All data		75% of data		50% of data		25% of data	
		Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
US04	9.43	9.47	1.76	9.28	2.61	9.43	2.73	11.30	4.54
US09	2.63	2.69	0.30	2.56	0.34	2.84	0.46	2.43	0.67
US10	6.13	6.23	1.08	6.48	1.37	5.72	1.36	6.01	1.74
US14	4.71	4.76	0.49	4.71	0.63	4.52	0.77	4.66	1.86
All Fields	5.73	5.79	0.91	5.76	1.24	5.63	1.33	6.10	2.20

Table 5.13 shows the mean and mean absolute error values of the predicted soil salinity from the 2004 Ikonos image for fields planted with alfalfa using the OLS, the SAR, and the modified kriging models for different data sets. The results shown in the above table are similar to those obtained with the Landsat image, where for fields with high variability even the dataset with 75% of the sample points does not perform very well. However, for fields with little variability, reasonable results are achieved when using every dataset except for the smallest dataset (25% of the points).

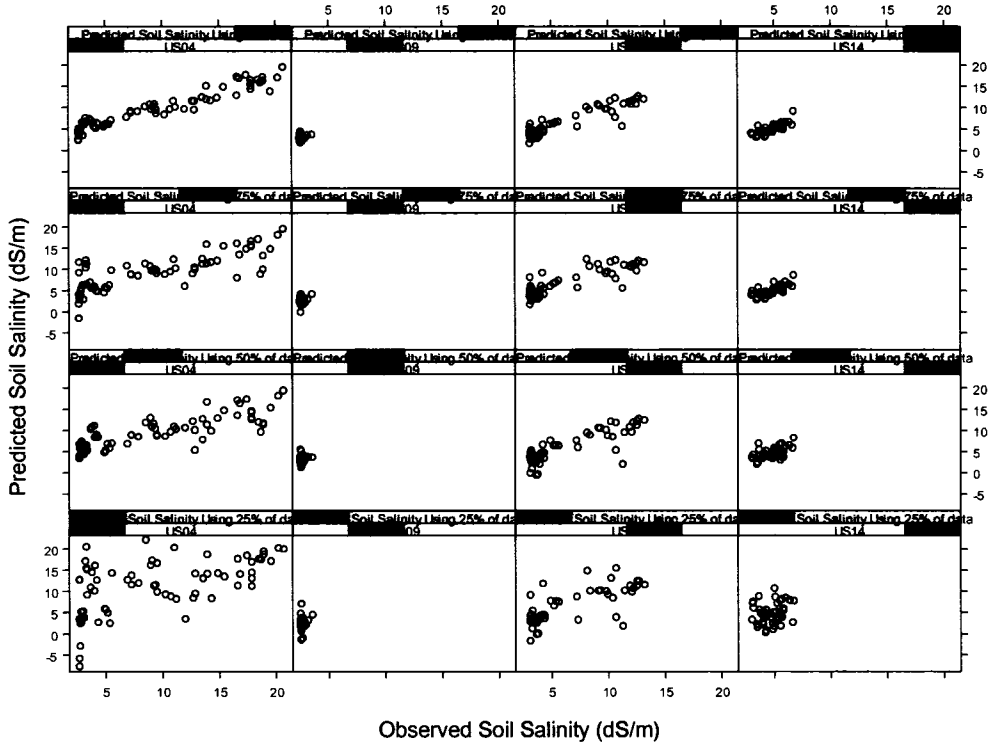


Figure 5.15: Observed versus predicted soil salinity (dS/m) for alfalfa fields using the 2004 Ikonos image for fields planted with alfalfa using all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data.

Figure 5.15 shows the observed versus predicted soil salinity (dS/m) for alfalfa fields using the 2004 Ikonos image in conjunction with all the observed data, 75% of the observed data, 50% of the observed data, and 25% of the observed data. The first panel of the scatter plot which shows the results of using all the data has an average R^2 value of 0.68. The second panel, which shows the scatter plots when using 75% of the data, has an average R^2 of 0.53. The third panel, which shows the scatter plots when using 50% of the data, has an average R^2 of 0.42. The panel on the bottom, which shows the scatter plots when using 25% of the data, shows an average R^2 of 0.22. Like the Landsat image, the Ikonos image gives better results when applied to corn fields than when used with alfalfa fields.

5.4 Comparison among predicted values from Aster, Landsat, and Ikonos images:

In this section the comparison among the predicted data from the Aster, Landsat, and Ikonos images will be presented in the form of scatter plots, histograms and box-and-whisker plots to compare between the fields planted with corn and alfalfa. Each kind of plot can be used to focus on one or more variable. The trilled scatter plot can be used to evaluate the trend and scatter of data, this type of scatter plot is more meaningful when more than one plot is displayed. The histogram is effective for displaying range. The box-and-whisker plot (Tukey, 1977) displays a statistical summary for a variable: median, quartiles, range and extreme values. It is especially useful for indicating whether a distribution is skewed and whether there are potential unusual observations (outliers) in the data set. It is also very useful when large numbers of observations are involved and when two or more data sets are being compared.

5.4.1 2001 Aster, Landsat, and Ikonos images for fields planted with corn

Figure 5.16 shows the scatter plot of the observed versus predicted soil salinity for the fields planted with corn using the Aster, Landsat, and Ikonos images. The upper four graphs show fields US09, US10, US40, and US 80 using the Landsat image, the four graphs in the middle show the same fields using the Ikonos image, while the bottom four show the same fields using the Aster image. It is clear that there is less scatter when using the Ikonos image than the Aster and Landsat images for fields planted with corn. The predicted data using the Ikonos image is very close to the observed values for fields US40 and US80 where the scattered points trend very

close to a straight line. The figure shows also that the range of soil salinity in field US09 is much smaller than that of the other two fields, a finding which was explained previously. Field US10 was not covered by the Ikonos image but was covered by both the Aster and Landsat images. The predicted values for this field are very similar when using both Aster and Landsat images.

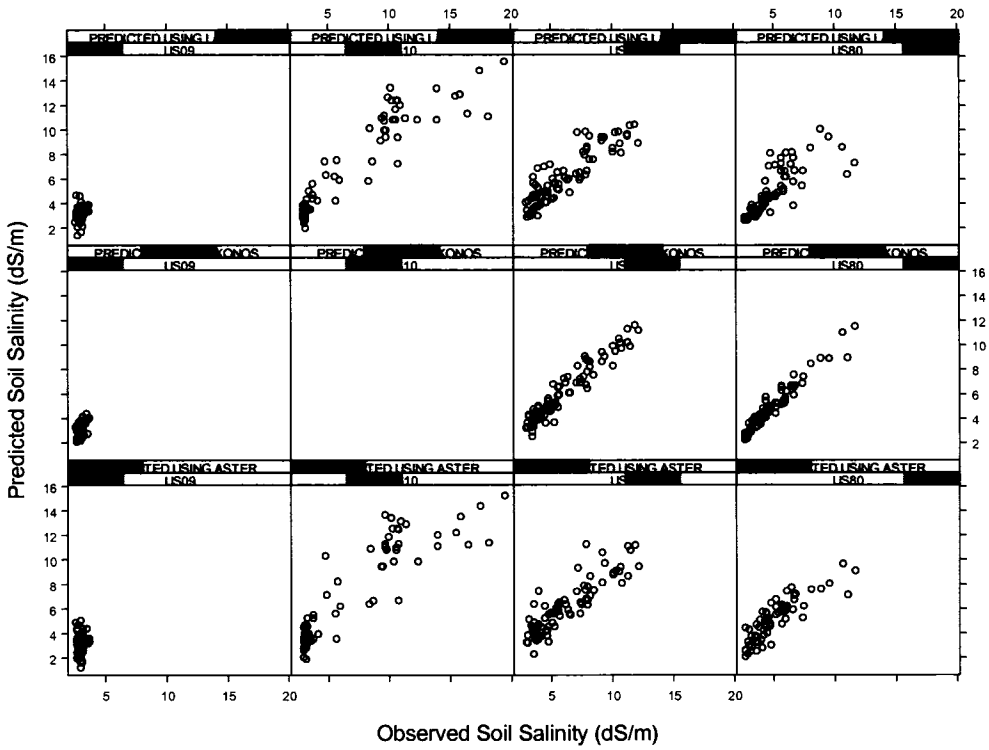


Figure 5.16: The scatter plot of observed versus predicted soil salinity (dS/m) for fields planted with corn for the 2001 using Aster, Ikonos, and Landsat images.

Figure 5.167 shows the histograms of the observed and predicted soil salinity of the fields planted with corn during 2001 using the Aster, Landsat, and Ikonos images. The four histograms on the first panel represent fields US09, US10, US40, and US80 developed from the Landsat image. The four histograms on the second panel represent fields US09, US10, US40, and US80 developed from the Ikonos image. The four histograms in the third panel represent the same fields for the Aster

image, and the four histograms on the bottom panel represent the observed soil salinity of the same fields. It is clear that the range of either the observed or predicted data is large for fields US40 and US80 while it is very small for field US09. Both the Aster and the Landsat images produce similar results but do not match the observed data as closely as the Ikonos image.

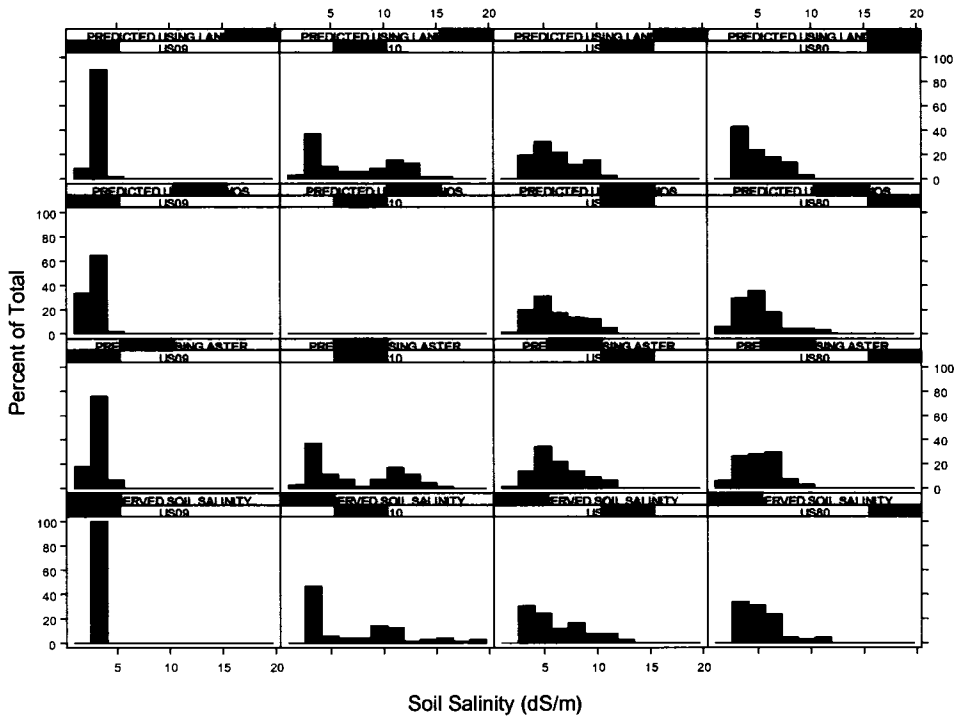


Figure 5.17: The histogram of the observed and predicted soil salinity (dS/m) for fields planted with corn during 2001 using Ikonos and Landsat images.

Figure 5.178 shows the box-and-whisker plot for the observed and predicted soil salinity of the fields planted with corn during 2001. The three box-and-whisker plots on the top represent the predicted soil salinity using the Landsat image of fields US80, US40, and US09 respectively. The three plots in the middle represent box-and-whisker plots of the predicted soil salinity for the same fields using the Ikonos image, while the three figures on the bottom represent the observed soil salinity of the same fields. Figure 5.18 show the whisker as the minimum and maximum points; the

median point is plotted inside the box, and the box represents 75% of the data. For field US80, the box for the Ikonos image is similar to the box of the observed data while it is larger for the Landsat image. The Ikonos image has a box and whiskers similar to the box and whiskers of the observed data while it has one outlier point higher than the maximum. For the same field, the Landsat image has a smaller box

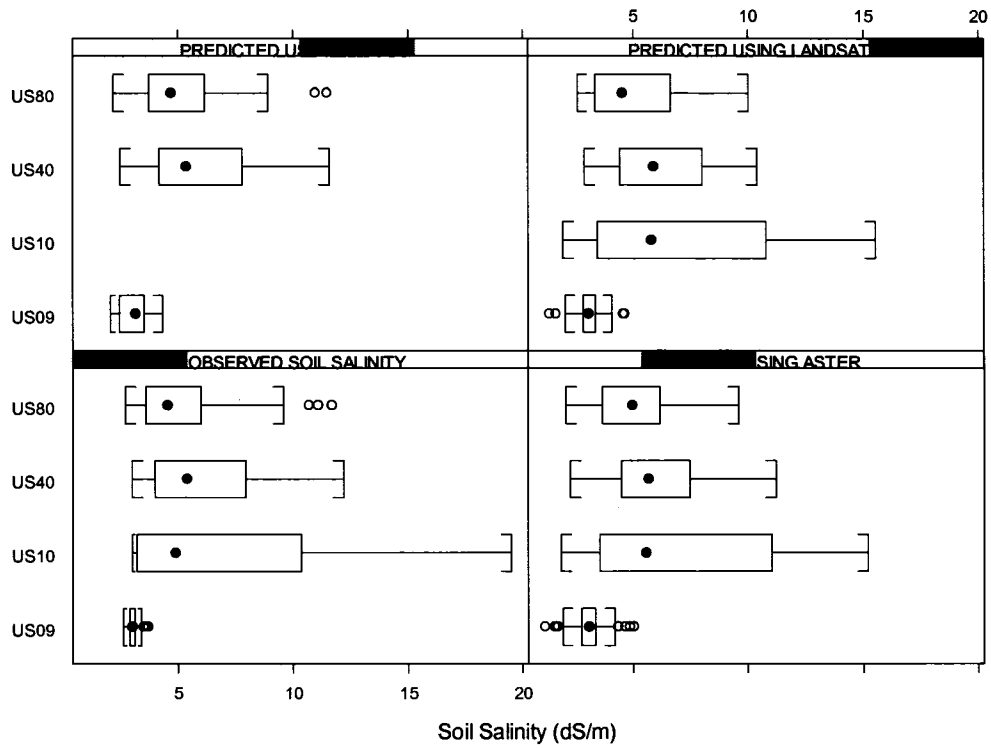


Figure 5.18: The box-and-whisker plot of the observed and predicted soil salinity (dS/m) for fields planted with corn during 2001 using Ikonos and Landsat images.

compared to the observed data while the small values are more than the observed data and it has one outlier point smaller than the minimum point. For field US10, the predicted data from the Ikonos and Landsat images are very similar. The box, the median, and the maximum points of both the Ikonos and Landsat images are close to the observed ones. The predicted minimum points of both the Ikonos and Landsat images are smaller than that of the observed ones. For field US09, the range, the

minimum, and the maximum of both the Ikonos and Landsat images are larger than that of the observed data. The median for the Ikonos and Landsat images is almost the same for both the observed and predicted data. There are more outliers in the predicted values than the observed ones. For field US04, the box for the observed data is larger than that of the predicted data using both the Ikonos and Landsat images.

5.4.2 2004 Landsat versus Ikonos images for fields planted with alfalfa

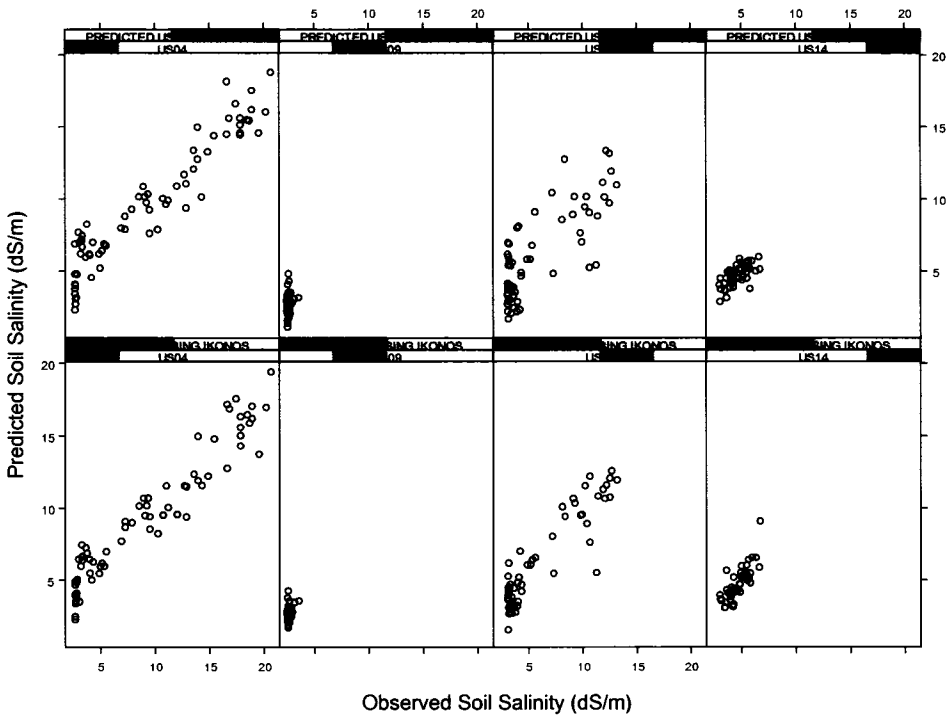


Figure 5.19: The scatter plot of observed versus predicted soil salinity (dS/m) for fields planted with alfalfa during 2004 using Ikonos and Landsat images.

Figure 5.19 shows the scatter plot of the fields planted with alfalfa during 2004. The upper four figures represent fields US04, US09, US10, and US14 when using the Landsat image while the lower four figures represent the same fields when using the Ikonos image. There is less scatter for the Ikonos image than for the Landsat

image. It is clear from the figure that the range of data in field US09 is the smallest while in field US04 it is the largest.

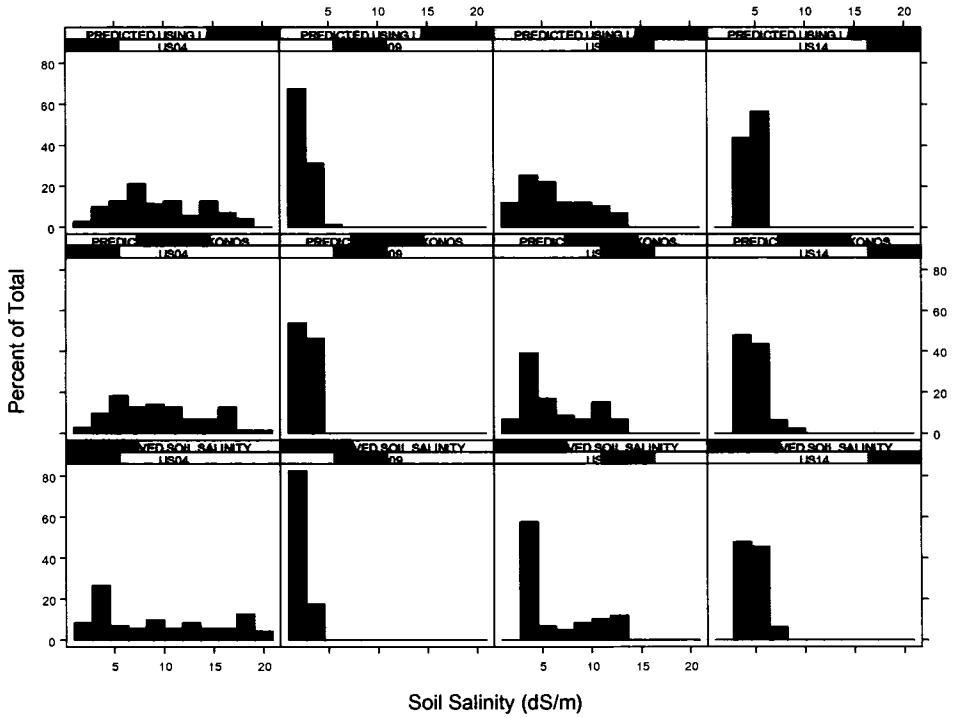


Figure 5.20: The histogram of the observed and predicted soil salinity (dS/m) for fields planted with alfalfa during 2004 using the Ikonos and Landsat images.

Figure 5.20 shows the histogram of the observed and predicted soil salinity of the fields planted with alfalfa using the Landsat and Ikonos images during 2004. The upper four histograms represent the predicted soil salinity of fields US04, US09, US10, and US14 when using the Landsat image, the middle four histograms represent the predicted soil salinity of the same fields when using the Ikonos image, while the lower four histograms represent the observed soil salinity of the same fields. Field US04 has the largest range; the distribution for the predicted values of the Ikonos image is a little smaller than the distribution of the predicted values from the Landsat image which is closer to the distribution of the observed values. Field US09 has the smallest distribution in both the observed and predicted data using the Landsat and

Ikonos images, and the distribution of the Ikonos image is closer to the distribution generated from the observed data than the Landsat image. For field US10, the distribution generated using the Ikonos image is closer to the distribution of the observed data than the distribution generated using the Landsat image. For field US14, both the distribution for the Ikonos and Landsat images are close to the distribution generated from the observed data.

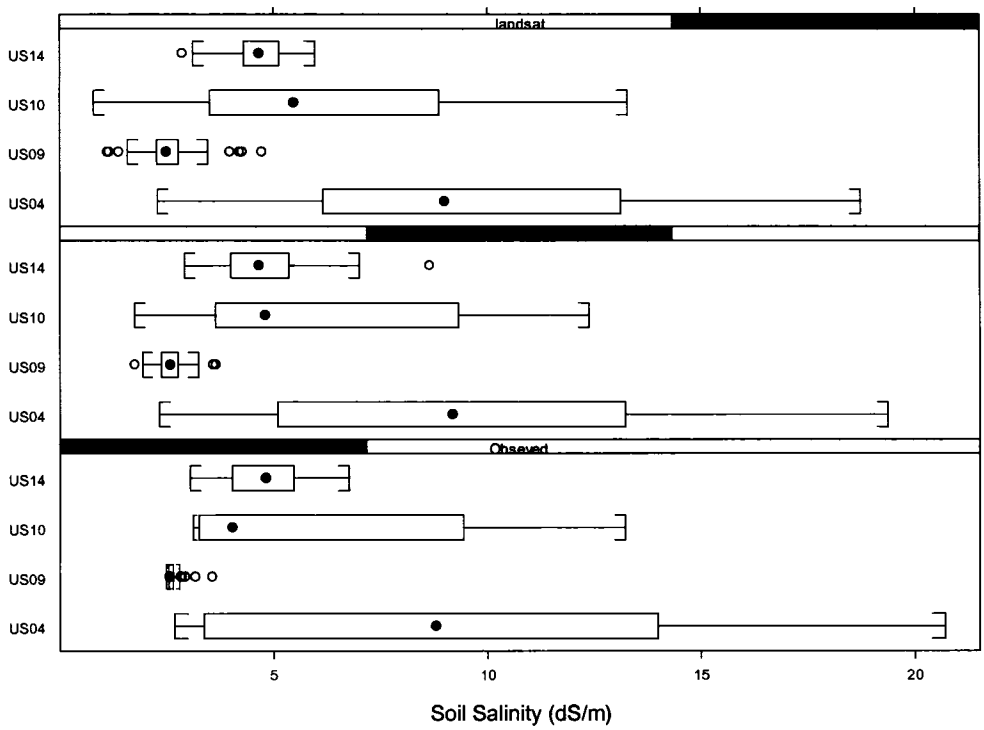


Figure 5.21: The box-and-whisker plot of the observed and predicted soil salinity (dS/m) for fields planted with alfalfa during 2004 using Ikonos and Landsat images.

Figure 5.21 shows the box-and-whisker plot of the observed and predicted soil salinity for the fields planted with alfalfa using the Ikonos and Landsat images during 2004. The four box-and-whisker plots on the top represent the predicted soil salinity for fields US14, US10, US09, and US04 using the Landsat image, the four box-and-whisker plots in the middle represent the predicted soil salinity for the same fields

using the Ikonos image, while the four box-and-whisker plots on the bottom represent the observed soil salinity for the same fields. For field US14, the Ikonos image has a box and whiskers plot similar to the box and whiskers of the observed data while it has one outlier point higher than the maximum. For the same field, the Landsat image has a smaller box compared to the observed data while the low values are lower than the observed. It has one outlier point lower than the minimum point. For field US10, the predicted data from the Ikonos and Landsat images are very similar. The box, median, and maximum points of both the Ikonos and Landsat images are close to the observed ones. However, the predicted minimum points of both the Ikonos and Landsat images are smaller than those of the observed data. For field US09, the range, minimum, and maximum of both the Ikonos and Landsat images are larger than those of the observed data. The median is almost the same in both the observed and predicted plots from both the Ikonos and Landsat images. For field US04, the box in the observed data is larger than that of the predicted using both the Ikonos and Landsat images.

5.5 Overall comparison of all images

This section provides a general comparison among the five images used in this study with the 2001 corn data and the 2004 alfalfa data. The comparison in this segment is mainly dependant on the R^2 values and the mean absolute error values to asses the best and the worst image type.

Table 5.14: R^2 and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Aster, 2001 Landsat, 2001 Ikonos, 2004 Landsat, and 2004 Ikonos images for combined fields using the OLS, the SAR, and the modified kriging models.

Field #	OLS model		SAR model		Modified kriging model	
	R^2	MAE	R^2	MAE	R^2	MAE
Aster 2001	0.34	1.92	0.05	2.86	0.81	0.91
Landsat 2001	0.47	1.70	0.18	1.82	0.83	0.82
Ikonos 2001	0.52	1.28	0.25	1.48	0.91	0.49
Landsat 2004	0.26	2.66	0.03	3.33	0.60	1.15
Ikonos 2004	0.37	1.96	0.15	2.68	0.68	0.78

The above table shows R^2 and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Aster, 2001 Landsat, 2001 Ikonos, 2004 Landsat, and 2004 Ikonos images for all fields combined using the OLS, the SAR, and the modified kriging models. It is clear that the values of the SAR models are the smallest, the OLS models are higher, and the modified kriging models are the highest. Also the mean absolute error values of the SAR models are the highest, the OLS models are smaller, and the modified kiging models are the smallest. The Ikonos 2001 image works best with all the models since it has the highest R^2 and smallest mean absolute error among all data sets. The Landsat 2004 image performs the worst since it produced the lowest R^2 and highest mean absolute error among all data sets.

Table 5.15: R^2 and mean absolute error values (dS/m) of the predicted soil salinity using modified kriging model from the 2001 Aster, 2001 Landsat, 2001 Ikonos, 2004 Landsat, and 2004 Ikonos images for combined fields using the all data, 90% of all the data, 50% of all the data, and 25% of all the data.

Field #	All data		75% of data		50% of data		25% of data	
	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
Aster 2001	0.81	0.91	0.72	0.97	0.61	1.30	0.39	1.83
Landsat 2001	0.83	0.82	0.74	1.06	0.63	1.30	0.42	1.33
Ikonos 2001	0.91	0.49	0.82	0.54	0.74	0.85	0.60	1.03
Landsat 2004	0.60	1.15	0.49	1.43	0.42	1.71	0.19	2.46
Ikonos 2004	0.68	0.78	0.53	1.24	0.42	1.33	0.22	2.20

The above table shows R^2 and mean absolute error values (dS/m) of the predicted soil salinity from the 2001 Aster, 2001 Landsat, 2001 Ikonos, 2004 Landsat, and 2004 Ikonos images for all fields combined using all the data, 75% of the data,

50% of the data, and 25% of the data. It is clear that the R^2 values worsen as the data sets become smaller while the mean absolute error values increase from the lowest to the highest as the data sets decrease. In the meantime, for some smaller data sets both R^2 and mean absolute error values are encouraging. Most values of the 75% and 50% of the data are very close to the numbers produced by all the data. The values produced using 25 % of the data and the Landsat 2001 and Ikonos 2001 images are also encouraging.

5.6 Model validation, discussion, and comments

5.6.1 Model validation

The modified kriging model was applied to five different images and in each case there was a significant improvement in the results over the OLS and the SAR models. The modified kriging model had the highest R^2 values compared to those of either the OLS or the SAR models. The mean values of the modified kriging model were the closest to the mean values of the observed soil salinity. Also the mean absolute error values of the modified kriging model were the smallest compared with the OLS and the SAR models. When the modified kriging model was applied to the other subsets of data, it also gave better results than the OLS and the SAR models. The modified kriging model was applied in 20 cases (four cases for each image: all data, 75% of data, 50% of data, and 25% of data) and in all cases it performed better than the OLS and the SAR models. Also for the field scale and all fields combined, the modified kriging model performed the best.

5.6.2 Discussion and comments

The results and analysis presented in this chapter show that the modified kriging model performs significantly better than the other models in trying to develop

soil salinity surfaces. Also selecting the best variogram as well as selecting the best number of nearest neighbors plays an important role in achieving better results. When using Ikonos images the results were slightly better than when using Aster or Landsat images. This can be attributed to the fact that the resolution of the Ikonos image is finer than that of the Aster and Landsat images. Since the improvement was not significant, it is a matter of judgement when deciding to select one over the other. For determining soil salinity it is strongly recommended that Aster and Landsat images be used since one scene covers a much larger area at a much cheaper price. Only in cases where there is a need for greater detail (4 meter pixels) would the Ikonos image be needed.

Determining soil salinity using satellite images was more successful in fields planted with corn than with alfalfa. A possible reason for this is that alfalfa is cut and harvested several times during the growing season making it difficult to capture images of alfalfa fields that are at the same growth stage. Also, since alfalfa is a shorter crop than corn, the presence of weeds could impact the results. Weeds growing in areas affected by salts would register as vegetative growth on the satellite image and be combined with the alfalfa data to bias the results.

CHAPTER 6. CONCLUSION

This study has shown the possible benefit of using satellite images in generating accurate soil salinity maps. It is clear that a key aspect of the success of using this methodology is the selection of the appropriate model, the modified kriging model. It is also important to note that the residuals can significantly affect the results. To generate high quality soil salinity maps, it is necessary to check for normality and spatial autocorrelation among the residuals. In most cases the OLS model produced higher R^2 values than the SAR model but it usually associates some autocorrelation in the residuals. The SAR model was able to remove some of the autocorrelation in the residuals but the R^2 values were reduced which means that the predicted values are not as close to the observed ones as those of the OLS model. It can be concluded that the inflation of the R^2 value associated with the OLS model is related to the autocorrelation in the residuals, which can lower the accuracy of the generated maps. When the residuals of the OLS model were kriged and combined with the OLS model, the results show a significant improvement in the R^2 over the other models.

The OLS model combined with the kriged residuals of the OLS performed the best among the models evaluated for all the cases evaluated in this research. The reason for this is that the OLS model considers the residuals while it adds the kriged residuals to the OLS model. The kriged residuals generate a surface that has positive and negative values which are added to the generated surface from the OLS model. This makes a significant improvement by adding positive values to the underestimated values and adding negative values to the overestimated values.

Selecting a combination of bands is better than using all the bands in the image since all the bands do not necessarily strongly correlate with soil salinity. Selecting only the correlated bands clears the image of unwanted bands which create noise.

- Bands 4, 5, 6, 8, and 9 from the short wave infrared group in addition to band 3 from the visible infrared group can give better results when using Aster image for fields planted with corn.
- The blue, the near infrared, the thermal, the short wave bands in addition to the NDVI can give better predictions of soil salinity when using Landsat images for fields planted with corn.
- The red band and NDVI can give better predictions of soil salinity when using Ikonos images for fields planted with corn.
- The red, the near infrared, the thermal band in addition to the NDVI can give better predictions of soil salinity when using Landsat images for fields planted with alfalfa.
- The green, the red, and the near infrared, in addition to the NDVI can give better predictions of soil salinity when using Landsat images for fields planted with alfalfa. This conclusion does not mean that the other bands which do not appear in these combinations are not correlated with soil salinity, but they are less correlated than the ones used in these combinations.

Remote sensing data can save a lot of time and effort for the soil salinity researcher by reducing the amount of data that needs to be collected in the field. This study shows that in some fields, accurate predicted soil salinity maps can be generated using as few as 12 or 20 observation points. A key issue which should be considered

in selecting the number of observation points is the range of soil salinity in a particular area. The larger the soil salinity range, the more observed data is needed.

It is better to predict soil salinity from remote sensing when a crop is present than when the soil is bare soil. Bare soil presents many challenges such as crop residue, snow, and weeds. Cover types with low to moderate tolerance to salinity such as corn can give a better indication of soil salinity since the roots penetrate the soil profile and the impacts of soil salinity can be detected through the biomass of the crop. Each cover type has its own reflectance therefore when trying to predict soil salinity, one specific cover type should be used at a time instead of using a mixture of cover types.

Statistical measures are very important in generating high quality soil salinity maps. It can be concluded that introducing more measurements and tests to the selected variables can lead to producing more accurate soil salinity maps. It is useful to evaluate the residuals using different data presentation forms. In this study, the residuals are examined using Moran's I p-value and Lagrange multiplier. Another way to evaluate residuals is to look at them graphically by analysing the residuals versus the weight of residuals and residuals versus the predicted values of soil salinity.

The best model selected is subjected to several tests such as: the model assumptions and selection criteria test of dependency and spatial autocorrelation among the residuals, and comparing the predicted soil salinity values with the observed values. For all cases, the kriged residuals combined with the OLS model proved to be the best model.

Mapping and assessing soil salinity requires integrating field data, GIS, remote sensing, and spatial modelling techniques. Any integration of field data, GIS, and

remote sensing is considered weak unless some strong statistical measures are introduced. The model that satisfies the assumptions, selection criteria and has no autocorrelation in residuals is not considered the best unless the predicted values of soil salinity match up relatively well with the observed values.

CHAPTER 7. FUTURE RESEARCH

A comprehensive study is needed to decide the best type of satellite image to use in estimating soil salinity for each crop type as well as for bare soil. Other important aspects to study in future research would include looking at timing of imagery and determining if there might be better bands or combinations of bands to use in detecting soil salinity. One source of error when using remote sensing data to assess either crop yield or soil salinity is the presence of weeds in the field, therefore future research should be conducted to determine the spectral signature of different types of weeds. If the signature of pig weeds or night shade weeds were known, then these spectral signatures could be considered as noise in the image which could be removed to make the interpretation of the image more accurate.

There is need for evaluating the accuracy of comparison methods for evaluating the different interpolation techniques such as inverse distance weight, spline, kriging, cokriging, etc. If an interpolation technique is determined to be superior to others then a clear set of guidelines should be developed on how use the interpolation model. For example: What degree polynomial should we use in describing the large-scale variability? How many nearest neighbors should we use in kriging or cokriging models? How do we select the secondary variables to include in

the cokriging models? What criteria do we use for comparing the different kriging and cokriging models?

Soil salinity maps can be enhanced by using cokriging interpolation for other data observed with soil salinity. In most instances, the data set contains information on the variable of interest along with one or more secondary variables, such as water table and groundwater salinity. Often, these variables prove to have strong cross correlation with soil salinity. Thus, it seems reasonable that additional research should be devoted to investigating how this additional information might be used to improve our estimates.

Disjunctive kriging technique might be very useful when conducting soil salinity mapping assessments in helping produce high quality probability maps. Disjunctive kriging is a nonlinear estimation technique that allows for the conditional probability that a value of a spatially variable management unit is greater than some threshold. Since disjunctive kriging is a nonlinear estimator, in general, it provides a more accurate estimate of the variable of interest. The conditional probability can be used as input to a management decision making model to provide a quantitative means for determining whether management actions are necessary and would be very useful when dealing with soil salinity and its impact on crops.

Categorical kriging might be very useful when conducting soil salinity mapping assessments to produce high quality classified maps. Zones of uncertainty between soil salinity types can be obtained by identifying locations with low probabilities, for a given threshold, that they might belong to a specific soil salinity level. For example, we can define a zone of uncertainty as being the lowest 25 percent of the probabilities of belonging to a particular soil salinity level. A soil salinity map can be categorized into three levels. The first level would indicate areas of low

salinity which should not worry farmers, the second level would show areas of moderate salinity which require careful management, and the third level would highlight areas of high salinity where farmers need to implement management practices such as leaching, drainage systems, and/or cropping changes.

Anisotropy might be important when dealing with soil salinity. Anisotropy is present when the spatial autocorrelation of a process changes with direction; the underlying physical process evolves differently in space. Unlike a variogram from an isotropic process, the variogram from an anisotropic process is not purely a function of the distance, but is a function of the magnitude and direction of the distance. There are two types of anisotropy: geometric anisotropy occurs when the range of the variogram changes in different directions while the sill remains constant, and zonal anisotropy exists when the sill of the variogram changes with direction. Therefore, where soil salinity might trend along the river, as for instance when conducting research on basin scale salinity issues, anisotropy might be a significant issue.

BIBLIOGRAPHY

1. Amer, M.H. and de Ridder, N.A. eds. (1989). *Land Drainage in Egypt.* Delta Barrage, Cairo: Drainage Research Institute.
2. Anselin, L. (1988). Lagrange Multiplier test diagnostics for spatial dependence and spatial autoregressive modelsa: some Monte Carlo results. *Journal of Regional Sciences*, **26**, 267-84.
3. Barr, N. 2003. Future agricultural landscapes. Australian Planner, 40,(2),123-127.
4. Barrett-Lennard, E.G. 2003. The interaction between waterlogging and salinity in higher plants: causes, consequences and implications. Plant and Soil 253: 35-54.
5. Barrett-Lennard, E.G., Malcolm, C.V. and Bathgate, A. 2003. Saltland Pastures in Australia – a Practical Guide, Second Edition. Sustainable Grazing of Saline Lands (a sub-program of Land, Water and Wool).
6. Bonham, C.D., Reich, R.M., and Leader, K.K. (1995). "A spatial cross-correlation of *Bouteloua gracilis* with site factors." *Grassland Science*, 41, 196-201.
7. Burkhalter, J.P., and Gates, T.K. 2005a. "Agroecological impacts from salinization and waterlogging in an irrigated river valley." *Journal of Irrigation and Drainage Engineering*, ASCE, 131(2): In press.
8. Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A*, **13**, 795-800.
9. Burrough, P. (1989). "Fuzzy mathematical methods for soil survey and land evaluation." *Journal of Soil Science*, 40 (3), 477–492.

10. Chamberlain, T., and Wilkinson, K. E., (Eds), 2004. Salinity investigations using airborne geophysics in the Lower Balonne area, Southern Queensland. Queensland Department of Natural Resources and Mines Technical Report QNRM 04255.

11. Chong, G.W., Reich, R.M., Kalkhan, M.A., and Stohlgren, T.J. (2001). "New approaches for sampling and modeling native and exotic plant species richness." *Western North American Naturalist*, 61 (3), 328-335.

12. Clarke, J. D. A., and Riesz, A. L., 2004. Fluvial architecture of the subsurface of the Lower Balonne area, Southern Queensland, Australia. CRC LEME Open File Report 162.

13. Cliff, A.D. and Ord, J.K. (1981). *Spatial Processes: Models and Applications.* London: Pion.

14. Clifton, C. AND Heislers, D. (2004). ASSALT: an asset based salinity priority setting approach. In 9th Murray-Darling Basin Groundwater Worskhop, 2004.

15. Craig, J. C., Shih, S. F., Boman, B. J., and Carter, G. A. (1998). *Detection of salinity stress in citrus trees using narrow-band multispectral imaging.* ASAE paper no. 983076. ASAE Annual International Meeting, Orlando, FL, USA, 12–16 July 1998. 10 pp.

16. Craig, R.G. (1979). Autocorrelation in LANDSAT data. *Proceedings of the 13th International Symposium on Remote Sensing of the Environment,* pp. 1517-24. Ann Arbor, Michigan.

17. Craig, R.G. and Labovitz, M.L. (1980). Sources of variation in LANDSAT autocorrelation. *Proceedings of the 14th International Symposium on Remote Sensing of the Environment,* pp. 1 755-67. San Jose, Costa Rica.

18. Cressie, N. A. (1988). Variogram. Entry in *Encyclopedia of Statistical Sciences, Vol.* 9, S. Kotz and N. L. Johnson, eds. Wiley, New York, 489-491.

19. Cressie, N. A. (1991). *Statistics for Spatial Data*. New York: John Willey & Sons.

20. Csillag, F., Pásztor, L. and Biehl, L. (1993). "Spectral band selection for the characterization of salinity status of soils." *Remote Sensing of Environment*, 43, 231–242.

21. Czaplewski, R.L., and Reich, R.M. (1993). "Expected value and variance of Moran's bivariate spatial autocorrelation statistic under permulation," *USDA For. Serv. Res. Paper. RM-309,* Fort Collins, Colorado.

22. Dregne, H., Kassas, M., and Razanov, B. (1991). "A new assessment of the world status of desertification." *Desertification Control Bulletin,* (United Nations Environment Programme), 20, 6-18.

23. Dwivedi, R. S., and Sreenivas K. (1998). "Image transforms as a tool for the study of soil salinity." *International Journal of Remote Sensing*, 19(4), 605-619.

24. Dwivedi, R. S., Ramana. K. V., Thammappa S. S., and Singh A. N. (2001). "The utility of IRS-1C LISS-III and PAN-merged data for mapping salt-affected soils". *Photogrammetric Engineering and Remote Sensing,* 67(10), 1167-1175.

25. Dwivedi, R.S., and Rao, B.R.M., (1992). "The selection of the best possible Landsat TM band combination for delineating salt-affected soils." *International Journal of Remote Sensing,* 13 (11), 2051–2058.

26. Eastman, J.R., (1992). *Time series map using standard principal components.* ASPRS/ACM/RT 92 Technical Papers, col. 1: Global Change and Education, 3-8 August, Washington DC, pp. 195-204.

27. Evans, F. (1998). "*An investigation into the use of maximum likelihood classifiers, decision trees, neural networks and conditional probabilitistic networks for*

mapping and predicting salinity." MSc thesis, School of Computer Science, Curtin University of Technology, Perth, Australia.

28. Fakuhara, M., Hayashi, H., Yasuda, Y., Asanuma, I., Emori, Y., and Lisaka, J., (1979). "Extraction of soil information from vegetated area." *Proceedings of International Symposium on Machine Processing of Remotely Sensed Data,* 242-251.
29. Farifteh, J., Farshad, A., George, R.J. (2006). "Assessing saltaffected soils using remote sensing, solute modeling, and geophysics." *Geoderma*, 130, 191-206.
30. Fernández-Buces, N., Siebe, C., Cram, S., Palacio, J.L. (2006). "Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico." *Journal of Arid Environment*, 65, 644-667.
31. Fisher, R. A. (1935). *The Design of Experiments.* Oliver and Boys, Edinburgh.
32. Fitzpatrick. A., Lawrie, K. C., Clarke, J., Coram, J., Wilkinson, K. and Herczeg, A., 2004. Mapping 'nested' groundwater flow systems, aquifers, water tables and salinity systems in the Lower Balonne, SW Queensland. Aust. Soc. Explor. Geophys. Conference, 2004.
33. Forster, B.C. (1980). Urban Residential ground cover using Landsat digital data. *Photogrammetric Engineering and Remote Sensing,* 46, 547-58.
34. Galloway, R. and Davidson, N.J. 1993. The response of Atriplex amnicola to the interactive effects of salinity and hypoxia. Journal of Experimental Botany 44: 653-663.
35. Gandin, L. S. (1963). *Objective Analysis of Meteorological Fields.* Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965).

36. Garthwaite, A.J., von Bothmer, R. and Colmer, T.D. 2003. Diversity in root aeration traits associated with waterlogging tolerance in the genus Hordeum. Functional Plant Biology 30: 875-889.

37. Gates, T.K.; Burkhalter, J.P., Labadie, J.W., Valliant, J.C., and Broner, I. (2002). "Monitoring and modeling flow and salt transport in a salinity-threatened irrigated valley," *Journal of Water Resources Planning and Management.* 128(2), 87-99.

38. Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association,* **70**, 320-328.

39. George, R. and Woodgate, P., 2002. Critical factors affecting the adoption of airborne geophysics for management of dryland salinity. Explor. Geophys., 33, 84-89.

40. George, R., Lawrie, K. C., and Woodgate, P., 2003. "Convince me all your bloody data and maps are going to help me manage salinity any better ?" A review of remote sensing methodologies for salinity mapping. 9th PURSL (Productive Use of Saline Land) Conference, Yepoon, Qld.

41. Ghassemi, F., Jackeman, A.J., and Nix, H.A. (1995). *Salinization of land and water resources : human causes, extent, management and case studies.* CAB International, Wallingford Oxon, UK.

42. Gleeson, A., Turner, C. and Douglas, J. 2002. Beyond agriculture, changing patterns of farm household income. Rural Industries Research and Development Corporation, Canberra

43. Golovina, N.N., Minskiy, D.Ye., Pankova, I., and Solov'yev, D.A. (1992). "Automated air photo interpretation in the mapping of soil salinization in cotton-growing zones." *Mapping Sciences and Remote Sensing,* 29, 262-268.

44. Goward, S.N., Waring, R.H., Dye, D.G., and Yang, J.L. (1994). "Ecological remote sensing at OTTER: Satellite macroscale observation." *Ecological Applications*, 4 (2), 322-343.

45. Greenway, H. and Munns, R. 1980. Mechanisms of salt tolerance in nonhalophytes. Annual Review of Plant Physiology 31: 149-190.

46. Griffith, D. A. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In S. L. Arlinghaus and D. A. Griffith (eds.), *Practical Handbook of Spatial Statistics.* Boca Raton, FL: CRC Press, pp. 65-82.

47. Haggett, P. (1976). Hybridizing alternative models of an epidemic diffusion process *Economic Geography,* 52, 136-146. [3.4]

48. Heynike, J.J.C. (1981). *The economic effects of the mineral content present in the Vaal River Barrage on the community of the PWVS Complex (A desk study).* Water Research Commission, Pretoria.

49. Hick, P.T., and Russell, W.G.R. (1990). "Some spectral considerations for remote sensing of soil salinity." *Australian Journal of Soil Research,* 28, 417–431.

50. Hick, P.T., Davies, J.R., and Steckis, R.A. (1984). "Mapping dryland salinity in Western Australia using remotely sensed data." *Satellite remote sensing: review and preview*, Remote Sensing Society, Reading, UK, 343–350.

51. Hill, M. J., and Donald, G. E. (2003). "Estimating spatio-temporal patterns of agricultural productivity in fragmented landscapes using AVHRR NDVI time series." *Remote Sensing of Environment,* 84 (3), 367-384.

52. Hillel, D. (2000). Salinity management for sustainable irrigation: integrating science, environment, and economics. The World Bank: Washington D.C.

53. Howari, F.M. (2003). "The use of remote sensing data to extract information from agricultural land with emphasis on soil salinity." *Australian Journal of Soil Research,* 41, 1243-1253.

54. Isaaks, E.H., and Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

55. Johnston, R.J. (1980). *Multivariate statistical analysis in geography.* Longman, New York, N.Y.

56. Jowett, G.H. (1952). The accuracy of systematic sampling from conveyor belts. *Applied Geostatistics*. **1**, 50-59.

57. Kalkhan, M. A., Stohlgren, T. J., Chong, G. W., Schell, L. D., and Reich, R. M. (2000). *A predictive model of plant diversity: Integration of remotely sensed data, GISand spatial statistics.* Presented at the Eighth Biennial Remote Sensing Applications Conference, Albuquerque, New Mexico, April 10-14, 2000.

58. Kalkhan, MA, Martinson, E.J. Omi, P.N., Stohlgren, T.J., Chong, G.W., and Hunter, M.A. (2002). "Integration of spatial information and spatial statistics: a case study of invasive plants and wildfire on the Cerro Grande fire, Los Alamos, New Mexico, USA." *Proc., 22nd Tall Timbers Fire Ecology Conference: Fire in Temperature, Boreal, and Montane Ecosystems,* Tall Timbers Research Station, Tallahassee, FL.

59. Kolmogorov, A. N. (1941). The local structure of turbulence in an incompressible fluid at very large Reynolds numbers. *Doklady Akademii Nauk SSSR,* 30, 301-305. Reprinted (1961), in *Turbulence: Classic Papers on Statistical Theory,* S. K. Friedlander and L. Topping, eds. Interscience Publishers, New York, 151-155.

60. Krige, D.G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, **52**, 119-139.

61. Lane, R., Brodie, R., and Fitzpatrick, A., 2004. Constrained inversion of AEM data from the Lower Balonne area, Southern Queensland, Australia. CRC LEME Open File Report 163.

62. Lawrie, K. C., Munday, T .J., Dent, D. L., Gibson, D. L., Brodie, R. C., Wilford, J., Reilly, N. S, Chan, R. A. and Baker, P. 2000. A 'Geological Systems' approach to understanding the processes involved in land and water salinization in areas of complex regolith- the Gilmore Project, central-west NSW. AGSO Res. News., v. 32, p. 13-32, May 2000.

63. Lawrie, K. C., Please, P. and Coram, J., 2003. Groundwater quality (salinity) and distribution in Australia., 2003. In "Water, the Australian Dilemma". Proceedings of the Annual Symposium of the Academy of Technological Sciences and Engineering, 17-18th Nov., Melbourne.

64. Lawrie, K. C., Wilford, J., Gibson, D., Pain, C., Roberts, L., Coram, J., Halas, L., Apps, H., Fitzpatrick, A., Tan, K. P., Sorenson, C. James, J. and Tate, S., 2004. Regolith constraints on modeling salt movements in upland landscapes in the Murray-Darling Basin. In Roach, I (ed.) CRC LEME Regional Regolith Symposia, 2004. p. 193-197.

65. Matern, B. (1960). Spatial Variation. *Meddelanden fran Statens Skogsforkskinginstitut*, **49**, No. 5. [Second edition (1986), *Lecture Notes in Statistics,* **No. 36**, Springer, New York].

66. Mather, P. M. (1987). Computer processing of remotely-sensed images: an introduction, Wiley, New York.

67. Matheron, G. (1962). *Traite de Geostatistique Appliquee, Tome I. Memoircs du Bureau de Recherches Geologiques et Minieres,* No. 14. Editions Technip, Paris.

68. Matheron, G. (1963b). Principles of geostatistics. *Economic Geology,* 58, 1246-1266.

69. McFarlane, D. J., George, R. J., and Caccetta, P.A., 2004. The extent and potential area of saltaffected land in Western Australia estimated using remote sensing and digital terrain models. In Engineering Salinity Solutions Conference, Perth, 2004.

70. Metternicht, G. (1998). "Fuzzy classification of JERS-1 SAR data: an evaluation of its performance for soil salinity mapping." *Ecological Modeling,* 111, 61–74.

71. Metternicht, G., and Fermont, A. (1998). "Estimating erosion surface features by linear mixture modeling." *Remote Sensing of Environment,* 64 (3), 254–265.

72. Metternicht, G.I. and Zinck, J.A. (1996). "Modeling salinity–alkalinity classes for mapping salt-affected topsoils in the semi-arid valleys of Cochabamba (Bolivia)." *ITC Journal* 1996 (2), 125–135.

73. Metternicht, G.I. and Zinck, J.A. (1997). "Spatial discrimination of salt- and sodium-affected soil surfaces." *International Journal of Remote Sensing*, 18 (12), 2571–2586.

74. Moran, P. A. P., 1950. "Notes on continuous stochastic phenomena", *Biometrika,* 17-23. [1.3.2]

75. Moran, P.A.P. (1948). The Interpretation of Statistical Maps, Roy. Stat. Soc., Ser. **B.** 10, 243-351.

76. Munday, T., 2004. Application of Airborne Geophysical Techniques to Salinity Issues in the Riverland, South Australia. DWLBC Report 2004/ 3. 71p.

77. Nelson, R., Alexander, F., Elliston, L. and Blias, A. 2004. Natural Resource Management on Australian Farms. ABARE eReport 04.7 Prepared for the Australian Government Department of Agriculture, Fisheries and Forestry, Canberra.

78. Olsen, A., R. and Schreuder, H.T. (1997). "Perspectives on large-scale natural resources surveys when cause-effect is a potential issue." *Environ. and Ecol. Stat.*, 4, 167-180.

79. Pannell, D. 2001. Explaining non-adoption of practices to prevent dryland salinity in Western Australia: Implications for policy. In 'Land Degradation'. (Ed. A Conacher), pp. 335-346. (Kluwer: Dordrecht)

80. Passioura, J. 2004. From propaganda to practicalities – the progressive evolution of the salinity debate. Australian Journal of Experimental Agriculture (submitted)

81. Peck, A.J., Thomas, J.F., and Williamson, D.R. (1983). "Salinity issues, effects of man on salinity in Australia." *Water 2000 Consultants Report, no. 8*. Canberra: Australian Government Publishing Service.

82. Peng, W.L. (1998). "Synthetic analysis for extracting information on soil salinity using remote sensing and GIS: a case study of Yanggao Basin in China." *Environmental Management,* 22 (1), 153–159.

83. Postel, S. (1999). *Pillar of Sand: Can the Irrigation Miracle Last?* W.W. Norton and Co., New York, NY.

84. Purcell, L. C., Ball, R. A., Reaper, J. D., III, and Vories, E. D., (2002). "Radiation use efficiency and biomass production in soybean at different plant population densities." *Crop Science*, 42 (1), 172-177.

85. Reich, R.M., Czaplewaki, R.L., and Bechthold, W.A. (1994). "Spatial cross-correlation of undisturbed natural shortleaf pine stands in northern Georgia." *Journal of Environmental and Ecological Statistics,* 1, 201-217.

86. Reisner, M. (1986). Cadillac desert: The American West and its disappearing water. Viking, New York, N. Y.

87. Richards, L. (1954). "United States salinity laboratory staff, diagnosis and improvement of saline and alkali soils." *Agriculture handbook* 60 Richards, L., ed, 1954. US Department of Agriculture, USA.

88. Richardson, A.J., and Weigand, C.L. (1977). "Distinguishing vegetation from soil background information." *Photogrammetric Engineering and Remote Sensing,* 43, 1541-1552.

89. Ridley, A. and Pannell, D.J. 2005. The role of plants and plant-based RandD in managing dryland salinity in Australia, Australian Journal of Experimental Agriculture (forthcoming).

90. Robbins, C.W., and Wiegand, C.L. (1990). "Field and laboratory measurements." *Agricultural Salinity Assessment and Management,* American Society of Civil Engineers, New York.

91. Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. (1973). "Monitoring vegetation systems in the Great Plains with ERTS." *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, vol. I,* Greenbelt, MD, 309-317.

92. Scheaffer, R.L., Mendenhall, W., and Ott, L. (1990). *Elementary survey sampling,* 4th ed., PWS-Kent Publishing Company, Boston, MA.

93. Serrano, L., Filella, I., and Peñuelas, J. (2000). "Remote sensing of biomass and yield of winter wheat under different nitrogen supplies." *Crop Science,* 40 (3), 723-731.

94. Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., Schlemmer, M. R., and Major, D. J. (2001). "Use of Remote-Sensing Imagery to Estimate Corn Grain Yield." *Agronomy Journal*, 93 (3), 583-589.

95. Short, D.C. and Colmer, T.D. 1999. Salt tolerance in the halophyte Halosarcia pergranulata subsp. pergranulata. Annals of Botany 83: 207-213.

96. Smith, D., Ustin, S., Adams, J. and Gillespie, A. (1990). "Vegetation in deserts: I. A regional measure of abundance from multispectral images." *Remote Sensing of Environment,* 31 (1), 1-26.

97. Spies, B. and Woodgate, P.,2004. Salinity Mapping in the Australian context. Technical Report. Land and Water Australia. 153p.

98. Srivastava, A., Tripathi N. K., and Gokhale, K. V. G. K. (1997). "Mapping groundwater salinity using IRS-1B LISS II data and GIS techniques." *International Journal of Remote Sensing*, 18 (13), 2853-2862.

99. Stohlgren, T.J., Chong, G.W., Kalkhan, M.A., and Schell, L.D. (1997). Multiscale sampling of plant diversity: effects of minimum mapping unit size. *Ecological Application,* 7(3), 1064-1074.

100. Stone, M. (1974). Cross-validatory choise and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, **36**, 111-133.

101. Szabolcs, I. (1989). *Salt-Affected Soils.* CRC Press, Boca Raton, FL.

102. Tanji, K., Läuchli, A., and Meyer, J. (1986). "Selenium in the San Joaquin Valley." *Environment.* 28(6), 6-11 and 34-39.

103. Taylor, G., and Dehaan, R. (2000). Salinity mapping with hyperspectral imagery. Available at www.geology.unsw.edu.au/research/R-Sensing.

104. Toth, T., Csillag, F., and Buttner, G. (1991). Satellite remote sensing of salinity—alkalinity in the Great Hungarian plain. *Impacts of Salinization and Acidification on Terrestrial Eco-system and Its Rehabilitation. Proceedings,* Fuchu, Tokyo.

105. Triantafilis, J., Odeh, I.O.A., and McBratney, A.B. (2001). "Five Geostatistical Models to Predict Soil Salinity from Electromagnetic Induction Data Across Irrigated Cotton." *Soil Science Society of America Journal*, 65, 869-878.

106. Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example.* Chichester: John Willey and Sons.

107. Utset, A., Ruiz, M., Herrera, J. and Deleon, D. (1998). "A geostatistical method for soil salinity sample site spacing." *Geoderma,* 86 (1-2), 143–151.

108. van Bueren, M. and Price, R 2004. Breaking ground – Key findings from 10 years of Australia's National Dryland Salinity Program. (Land and Water Australia: Canberra)

109. Vanclay, F. 1997. The social basis of environmental management in agriculture. In 'Çritical Landcare'(Eds S Lockie, F Vanclay) pp. 9-28. (Centre for Rural Social Research, Charles Sturt University, Wagga Wagga)

110. Wartenberg, D. (1985). "Multivariable spatial correlation: a method for exploratory geographical analysis." *Geographical Analysis*, 17 (4), 263-283.

111. Watson, G. S. (1972). Trend-surface analysis and spatial correlation. *Geological Society of America, Special Paper*, **146**, 39-46.

112. Whitten, E. H. T. (1970). Orthogonal polynomial trend surfaces for irregularly spaced data. *Journal of the International Association for Mathematical Geology,* **2**,

113. Wiegand, C.L., Rhoades, J.D., Escobar, D.E., and Everitt, J.H. (1994). "Photographic and videographic observations for determining and mapping the response of cotton to soil salinity." *Remote Sensing of Environment*, 49 (3), 212-223.

114. Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* MIT Press, Cambridge, MA.

115. Wilford, J., R., Dent, D. L., Dowling, T. and Braaten, R., 2001. Rapid mapping of soils and salt stores using airborne radiometrics and digital elevation models. AGSO Res. News., May 2001, p. 33-40.

116. Wilhelm, W.W., Ruwe, K., and Schlemmer, M.R. (2000). "Comparison of the three leaf area index meters in a corn canopy." *Crop Science,* 40 (4), 1179-1183.

117. Williams, W.D. (1987). "Salinization of rivers and streams: an important environmental hazard." *Ambio*, 16(4), 180-185.

118. Wold, H. (1938). *A Study in the Analysis of Stationary Time Series.* Almqvist and Wiksells, Uppsala.

119. Yates, F. (1938). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika*, **30**, 444-466.

120. Zadeh, L., (1965). "Fuzzy sets." *Information and Control*, 8 (3), 338–353.

APPENDICIES

APPENDIX A: MORE EXAMPLES OF GENERATED SURFACES FROM OTHER IMAGES

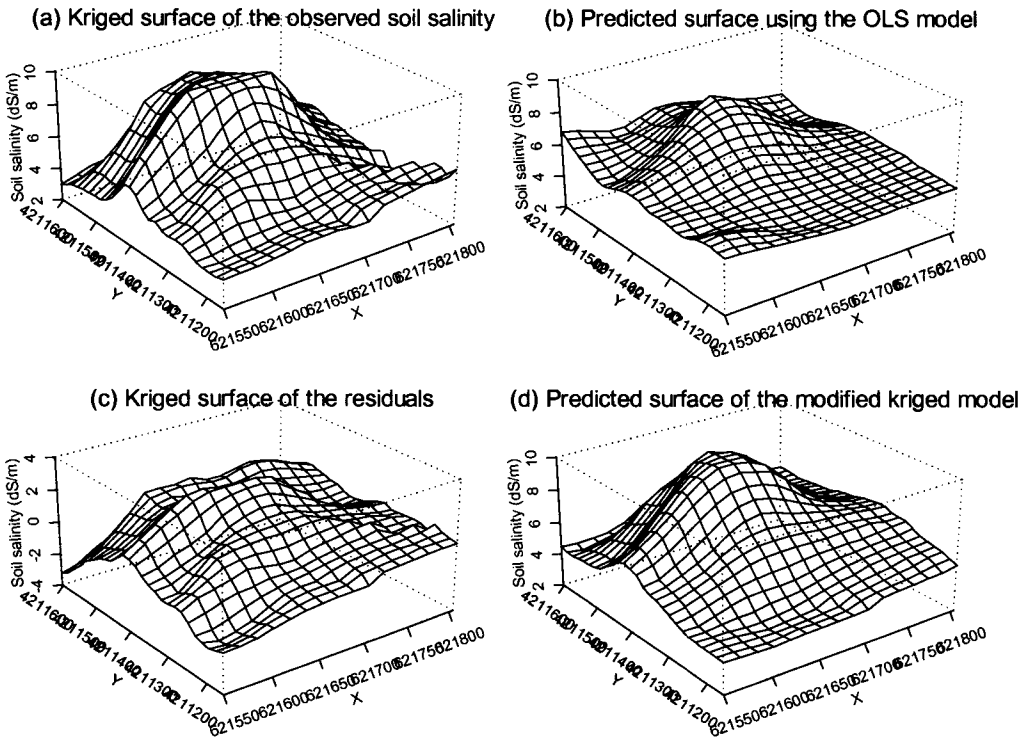


Figure A 1: Generated surfaces for field US40 using the 2001 data from the 2001 Aster image.

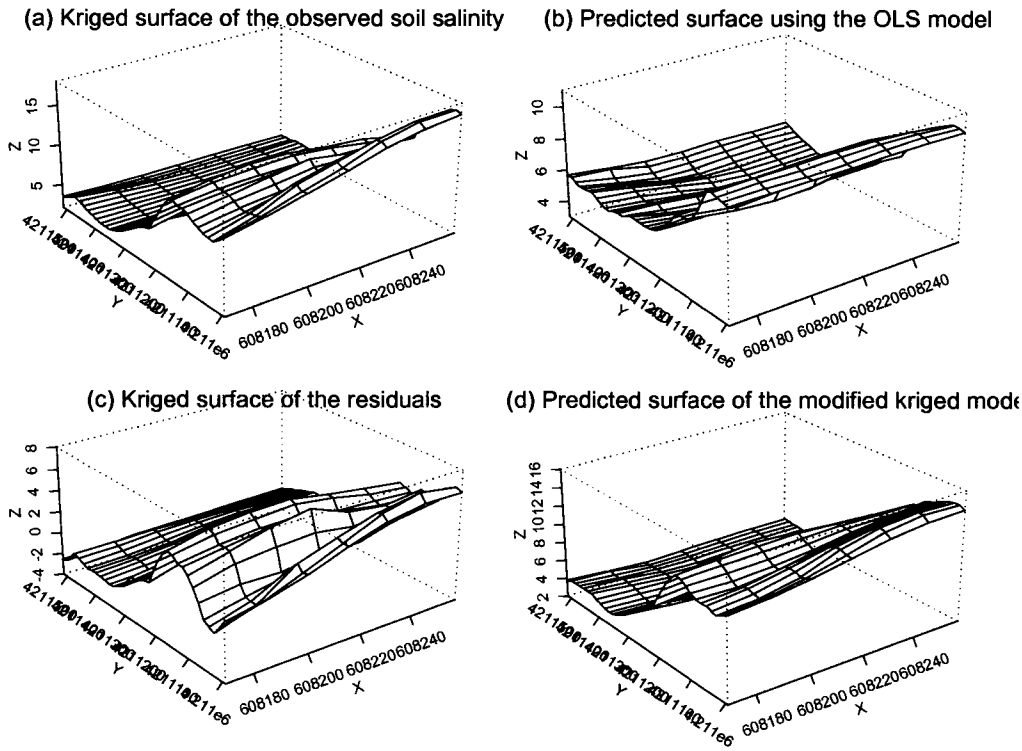


Figure A 2: Generated surfaces for field US10 using the 2001 data from the 2001 Landsat image.

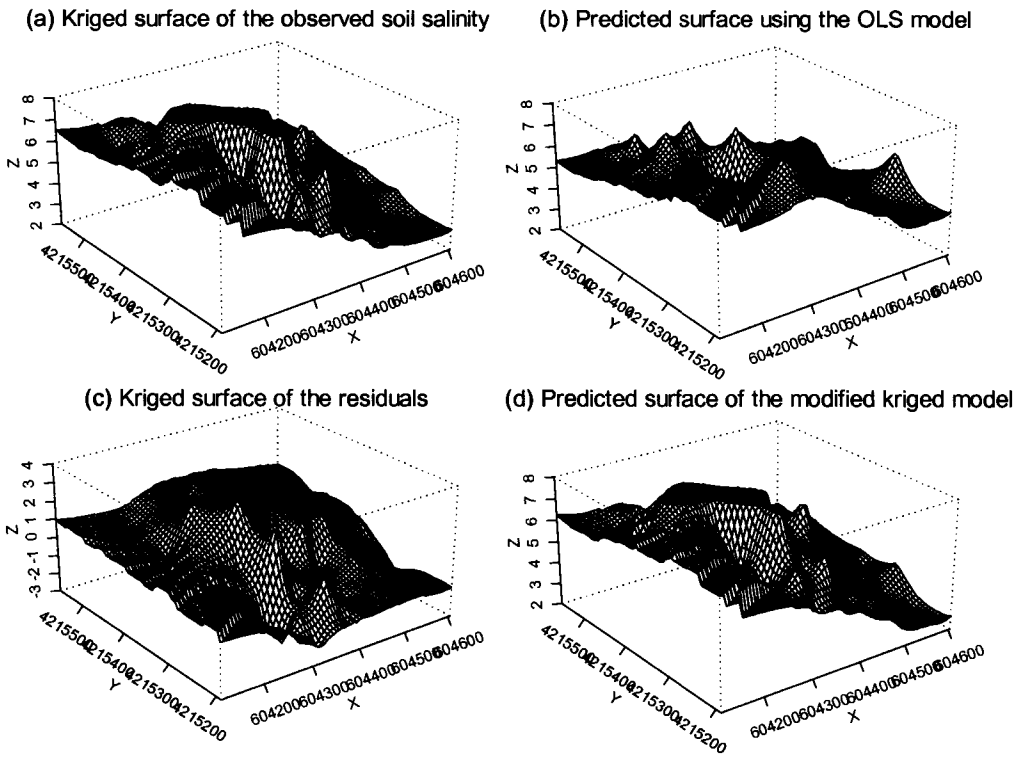


Figure A 3: Generated surfaces for field US80 using the 2001 data from the 2001 Ikonos image.

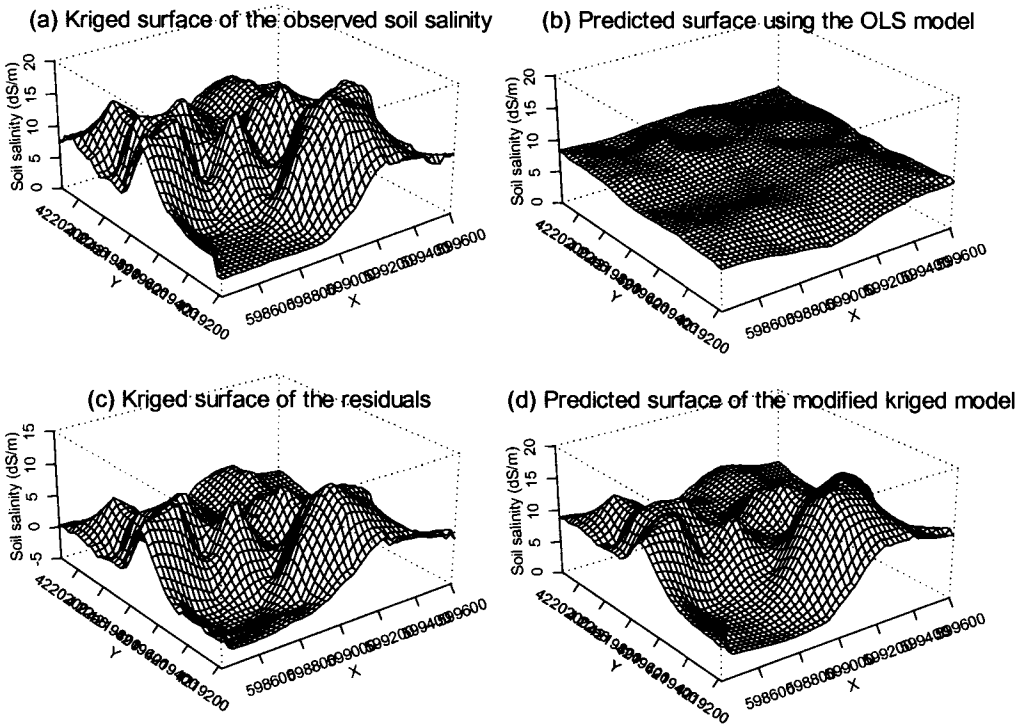


Figure A 4: Generated surfaces for field US04 using the 2004 data from the 2004 Landsat image.

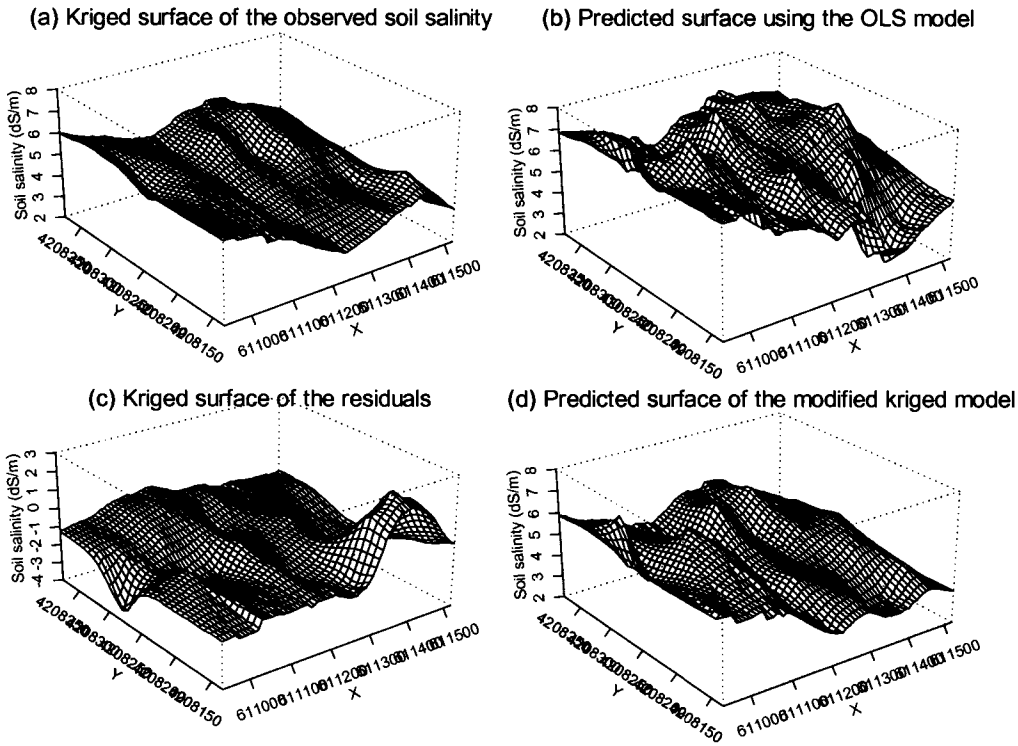


Figure A 5: Generated surfaces for field US14 using the 2004 data from the 2004 Ikonos image.

APPENDIX B: MORE EXAMPLES OF PREDICTED SURFACES USING DATA SETS

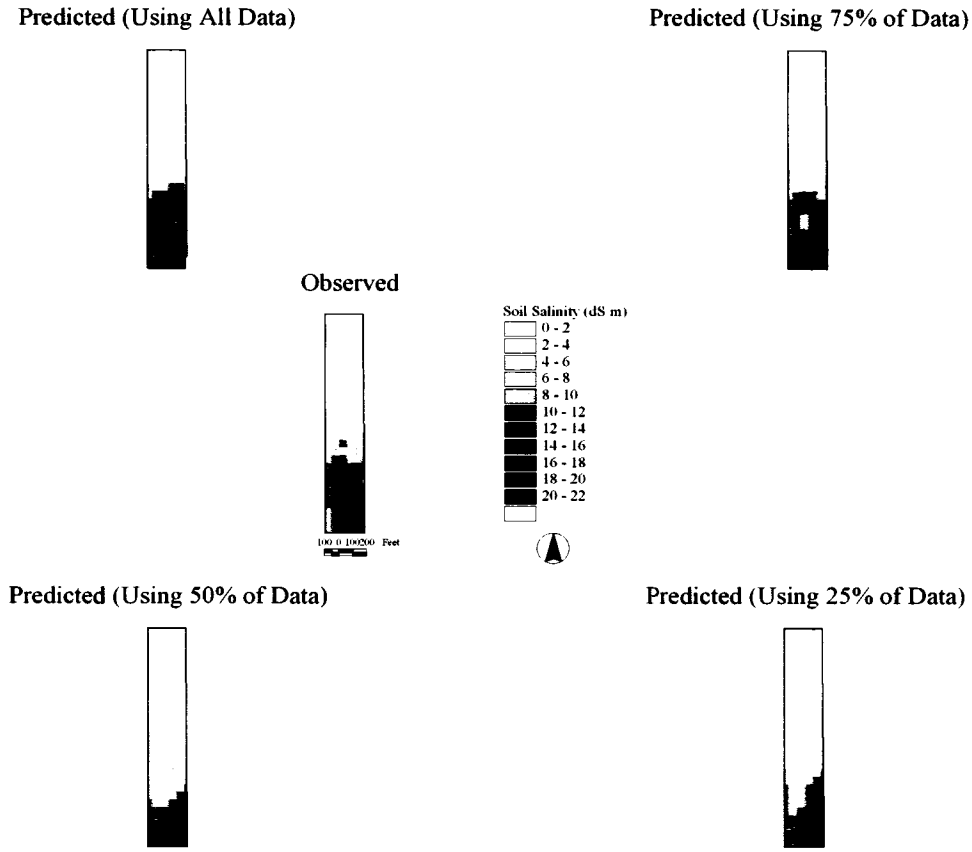


Figure B 1: Observed and predicted surfaces of soil salinity (dS/m) of field US04 from the 2001 Landsat image for fields planted with corn using all the observed data and data sets.

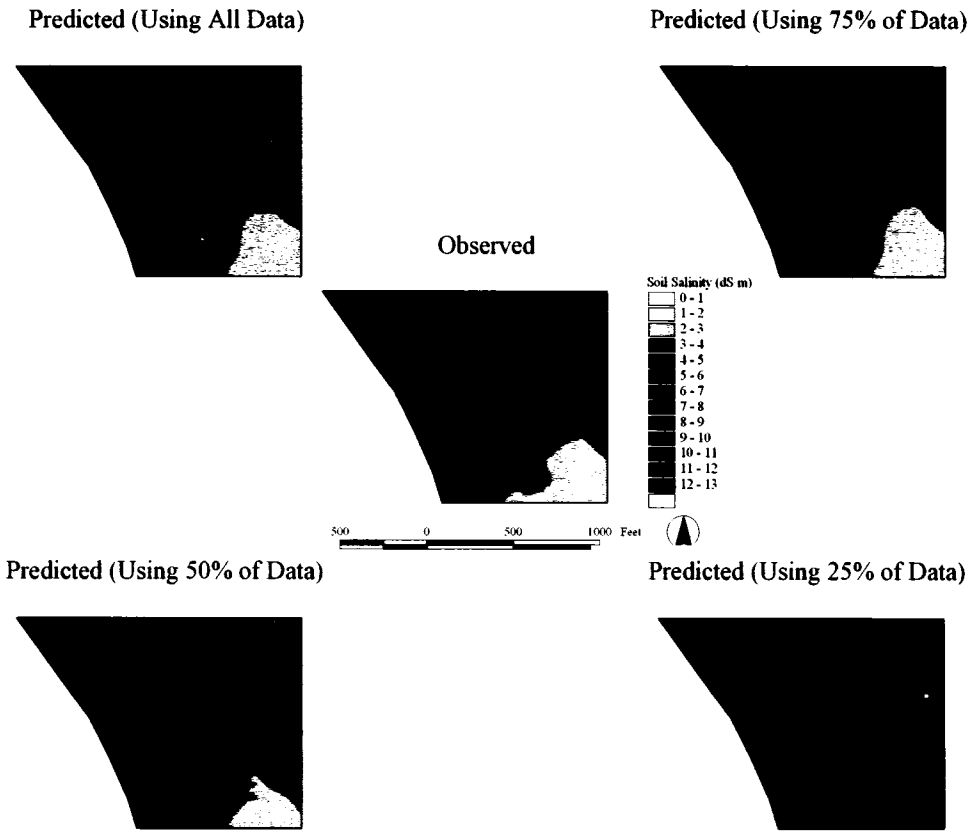


Figure B 2: Observed and predicted surfaces of soil salinity (dS/m) of field US80 from the 2001 Ikonos image for fields planted with corn using all the observed data and data sets.

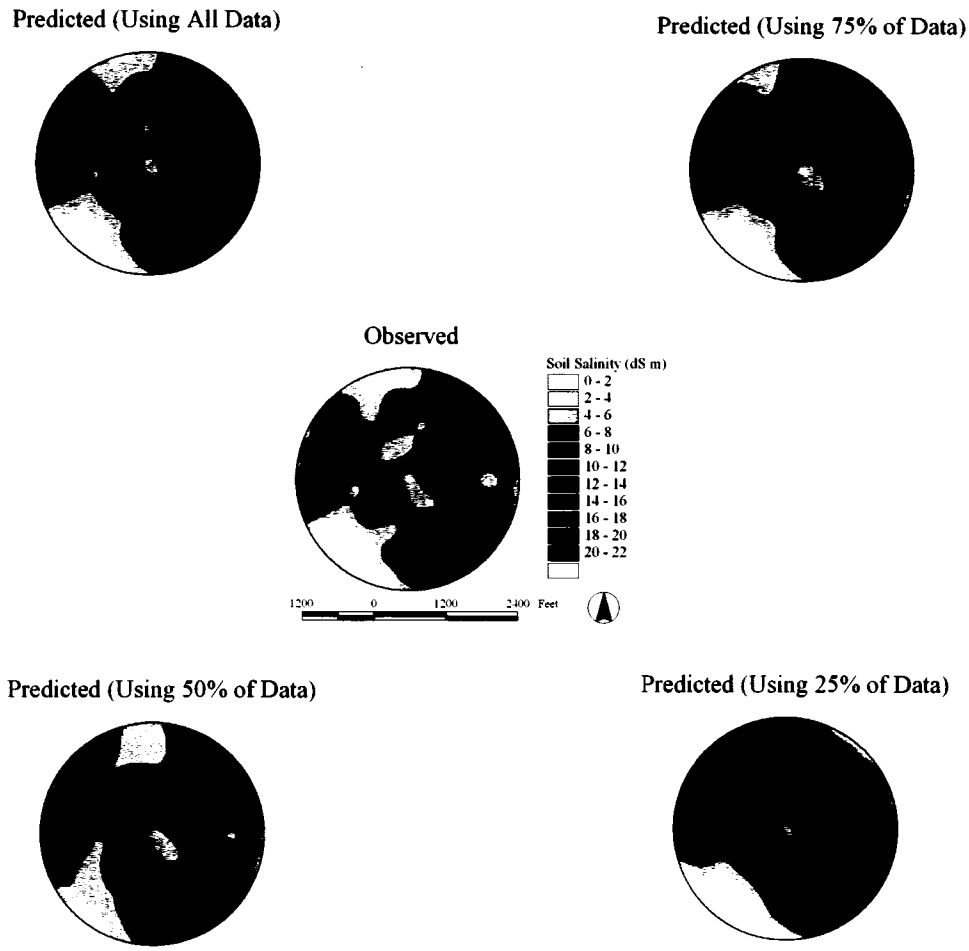


Figure B 3: Observed and predicted surfaces of soil salinity (dS/m) of field US04 from the 2004 Landsat image for fields planted with alfalfa using all the observed data and data sets.

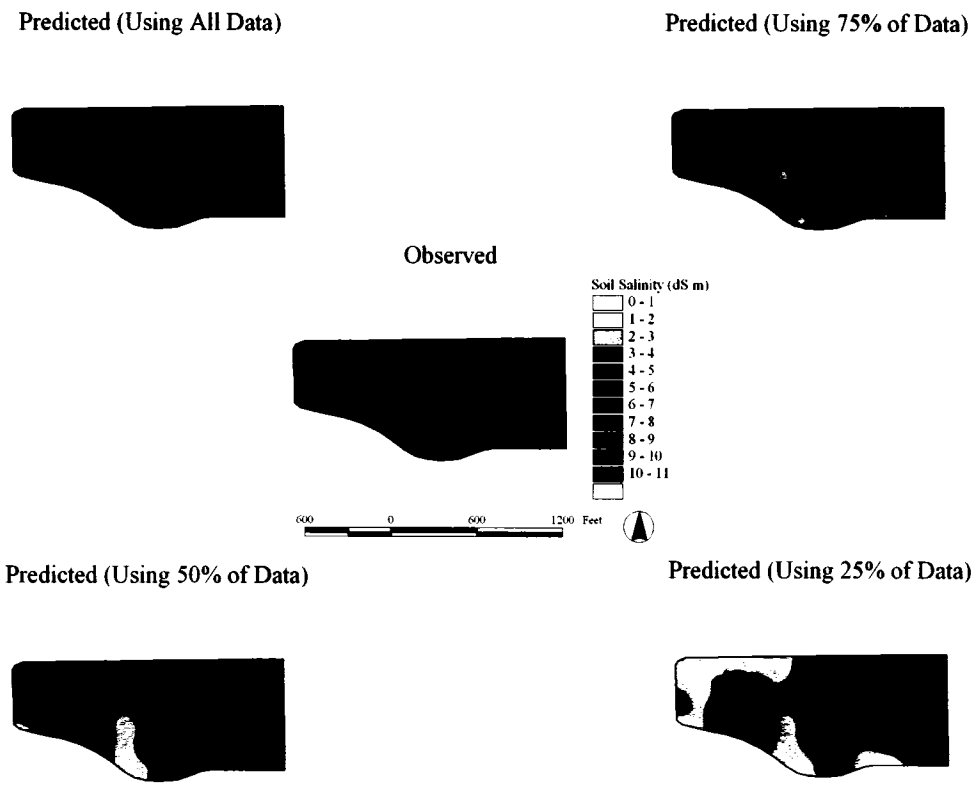


Figure B 4: Observed and predicted surfaces of soil salinity (dS/m) of field US14 from the 2004 Ikonos image for fields planted with alfalfa using all the observed data and data sets.

APPENDIX C: FIGURES OF MEAN ABSOLUTE ERROR VALUES OF THE MODELS AND DATA SETS

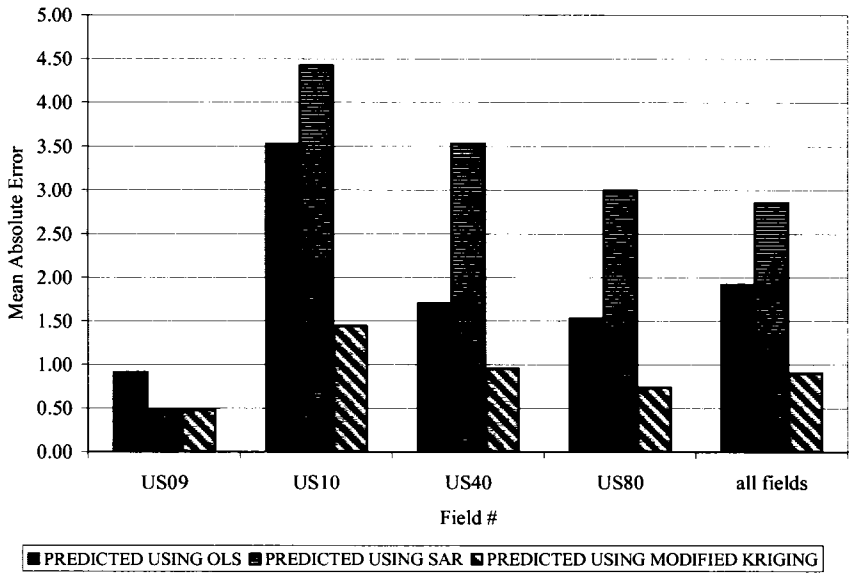


Figure C 1: Mean absolute error values of the OLS, SAR, and modified kriging models using 2001 data set of corn fields in conjunction with the 2001 Aster image.

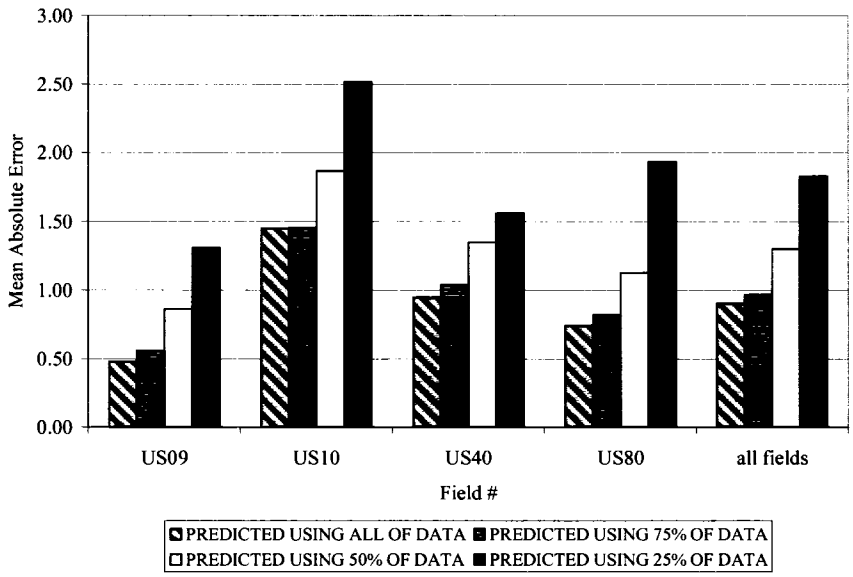


Figure C 2: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2001 data set of corn fields in conjunction with the 2001 Aster image.

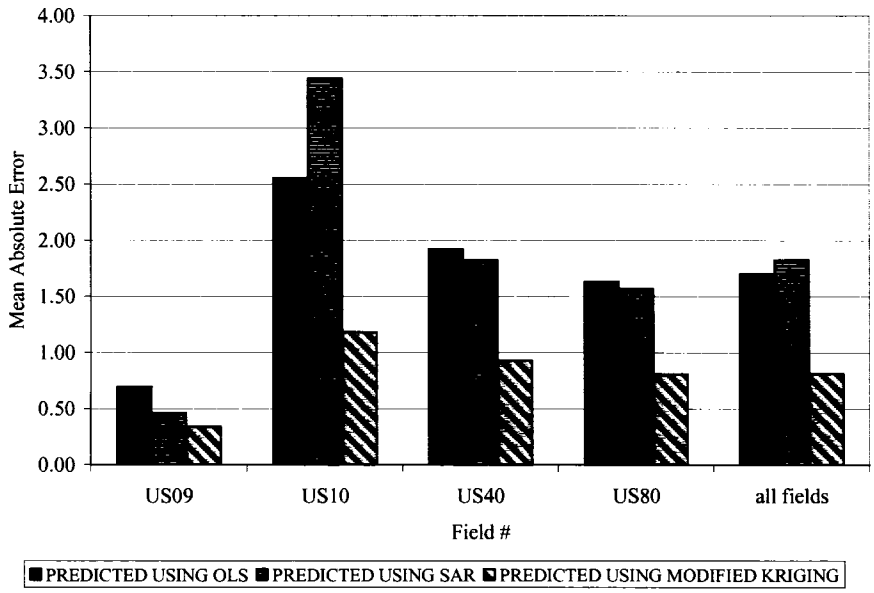


Figure C 3: Mean absolute error values of the OLS, SAR, and modified kriging models using 2001 data set of corn fields in conjunction with the 2001 Landsat image.

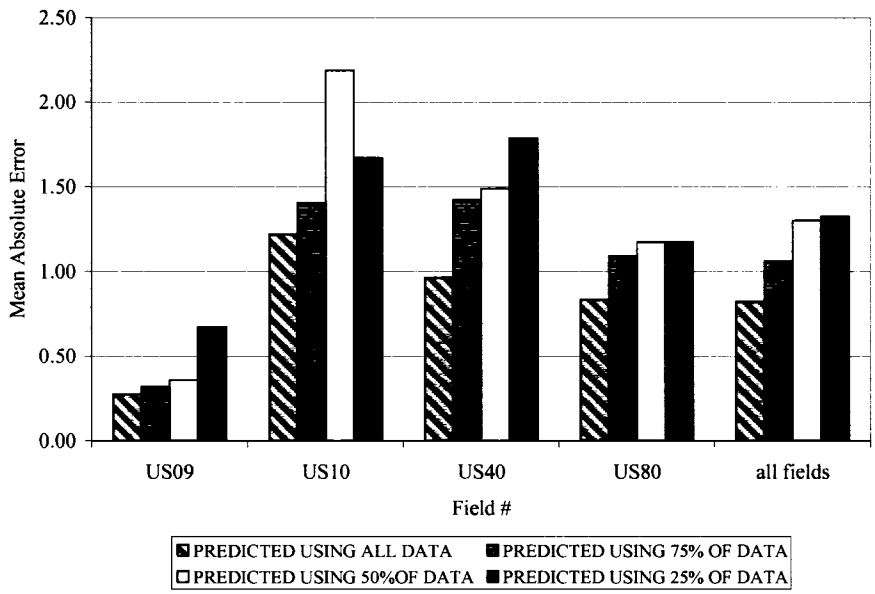


Figure C 4: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2001 data set of corn fields in conjunction with the 2001 Landsat image.

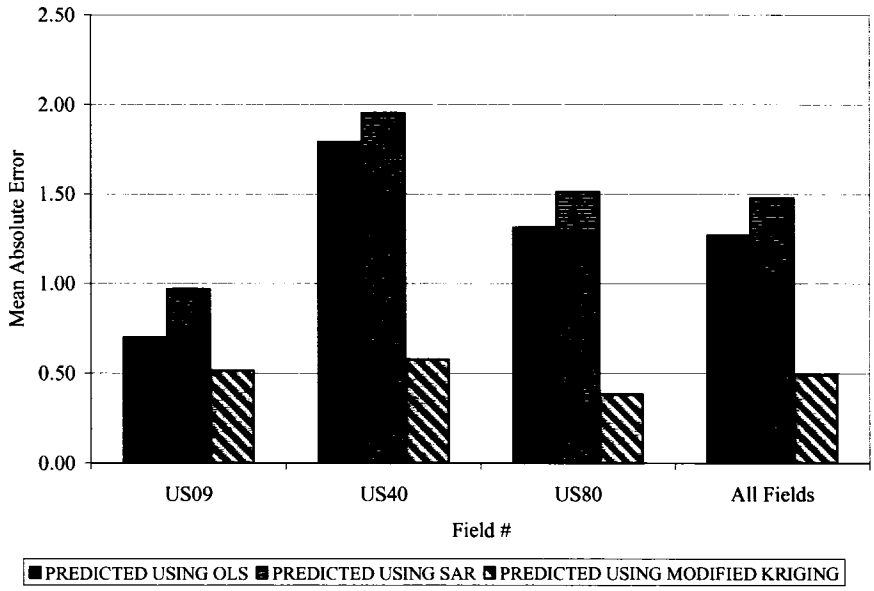


Figure C 5: Mean absolute error values of the OLS, SAR, and modified kriging models using 2001 data set of corn fields in conjunction with the 2001 Ikonos image.

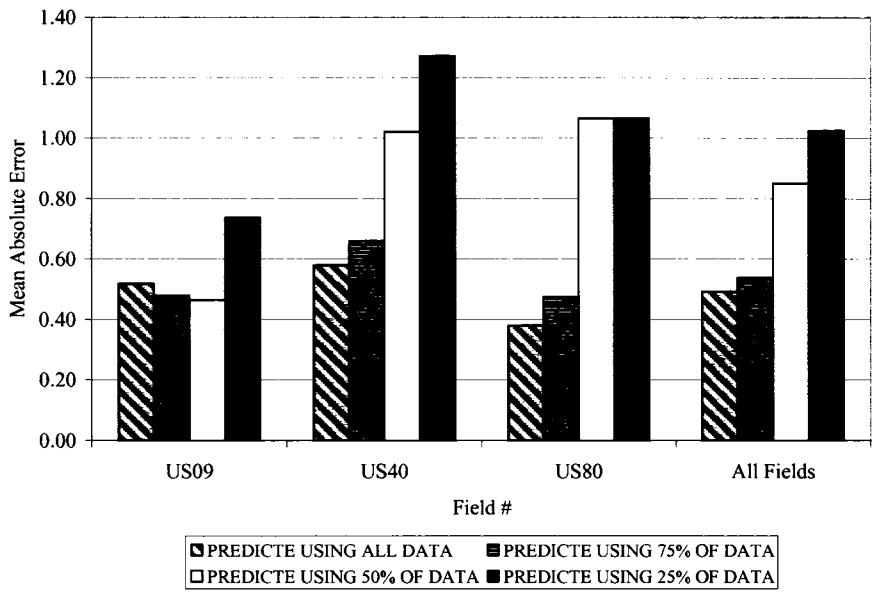


Figure C 6: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2001 data set of corn fields in conjunction with the 2001 Ikonos image.

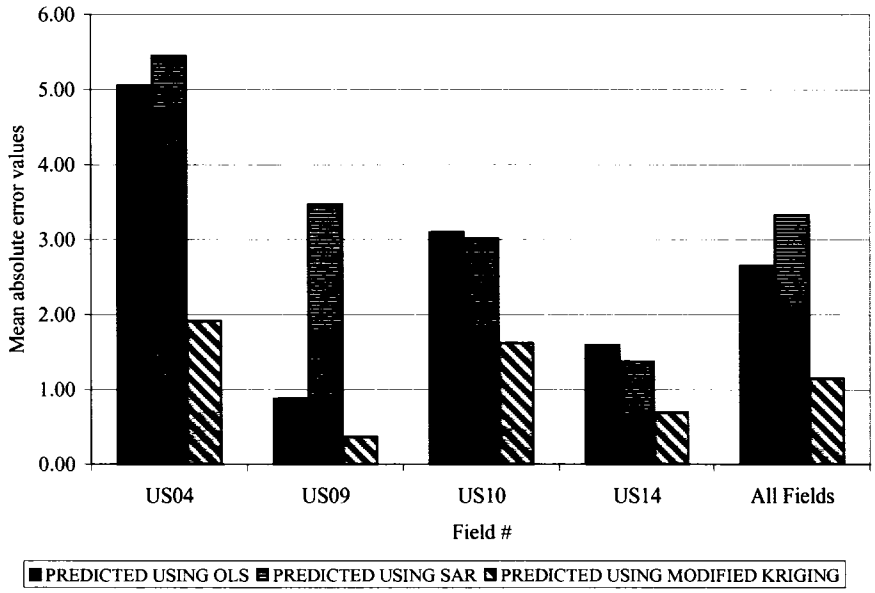


Figure C 7: Mean absolute error values of the OLS, SAR, and modified kriging models using 2004 data set of alfalfa fields in conjunction with the 2004 Landsat image.

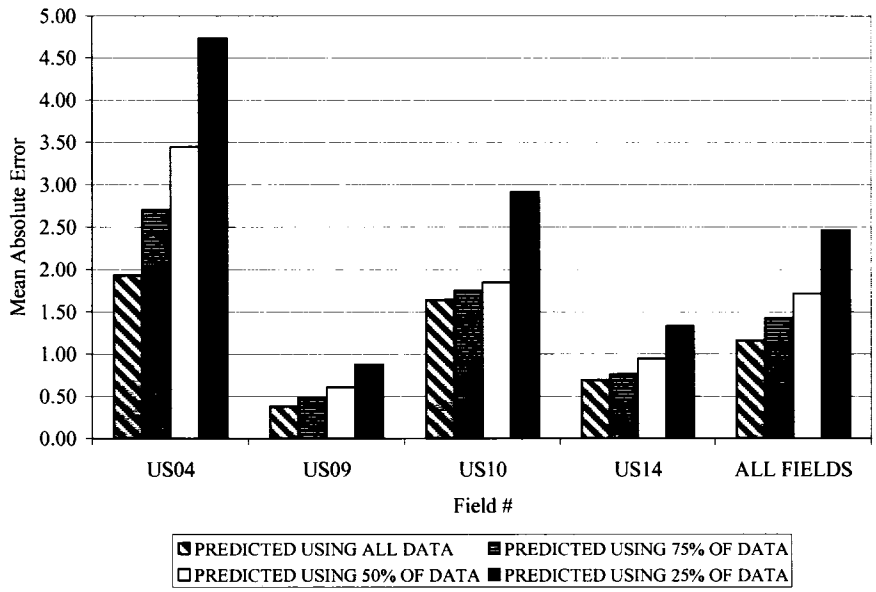


Figure C 8: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2004 data set of alfalfa fields in conjunction with the 2004 Landsat image.

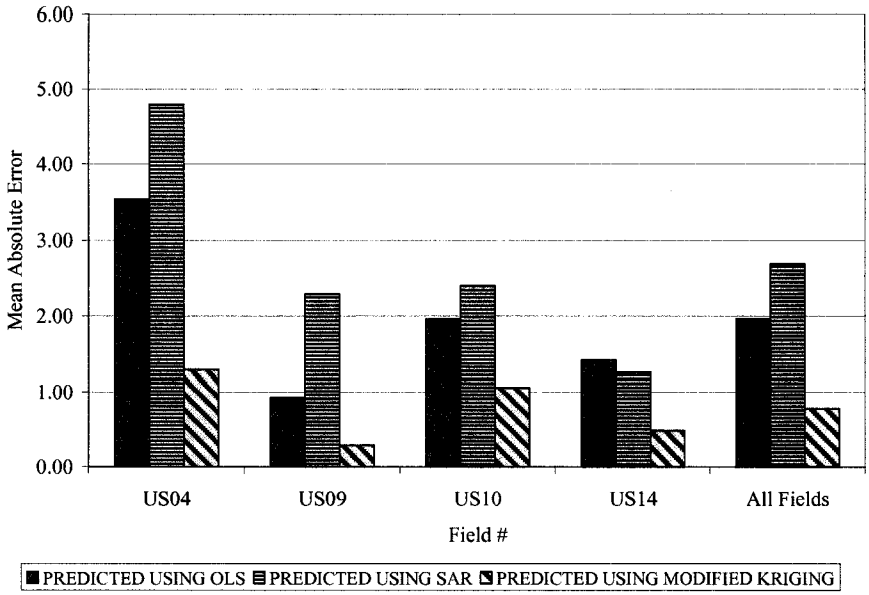


Figure C 9: Mean absolute error values of the OLS, SAR, and modified kriging models using 2004 data set of alfalfa fields in conjunction with the 2004 Ikonos image.

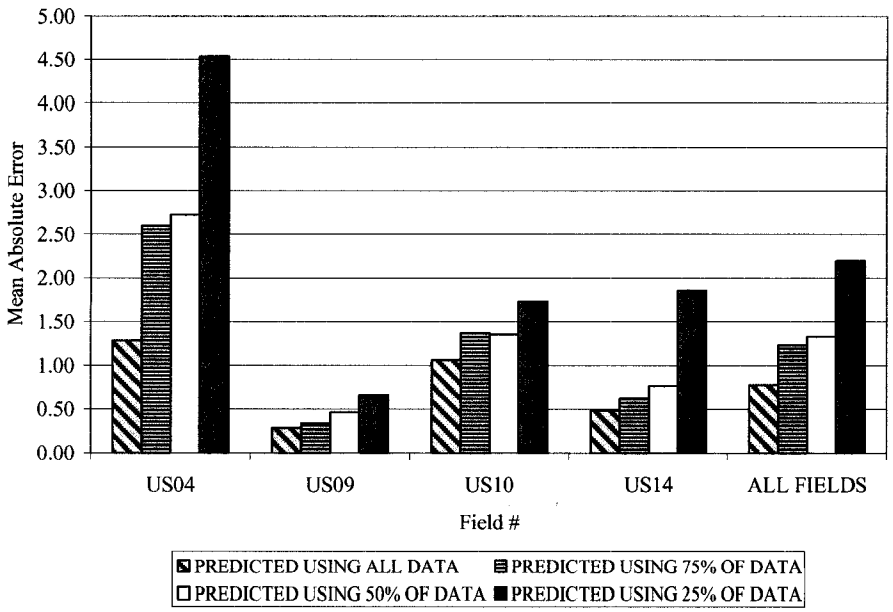


Figure C 10: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2004 data set of alfalfa fields in conjunction with the 2004 Ikonos image.

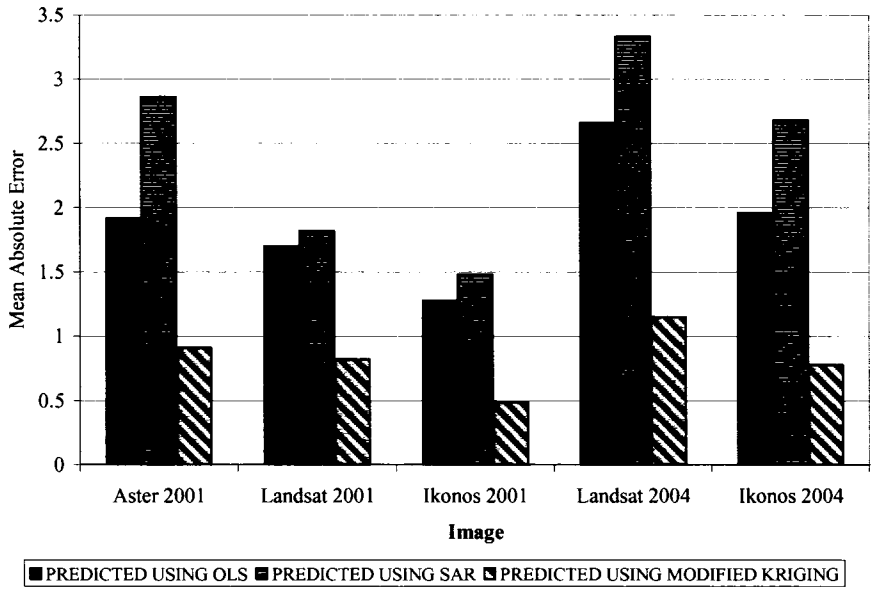


Figure C 11: Mean absolute error values of the OLS, SAR, and modified kriging models for Aster 2001, Landsat 2001, Ikonos 2001, Landsat 2001, and Ikonos 2001 images.

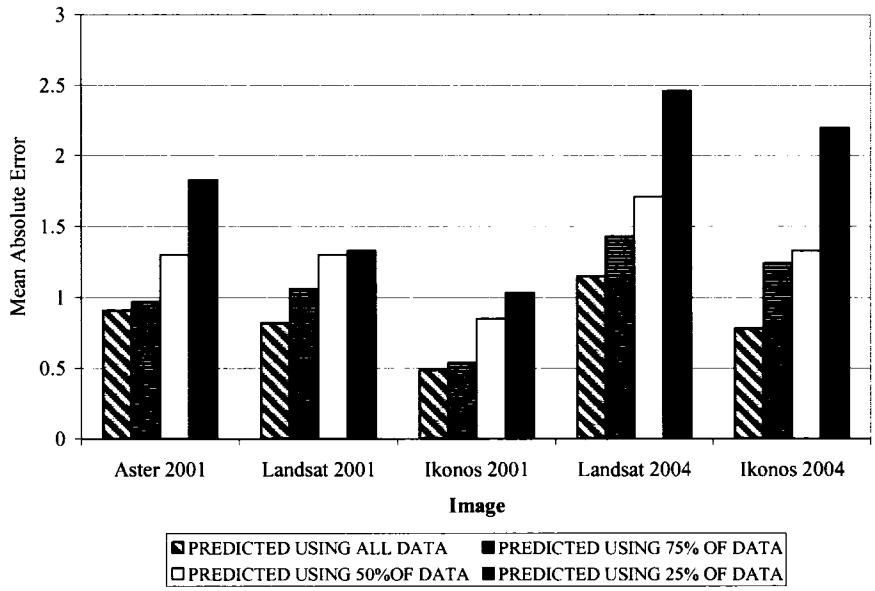


Figure C 12: Mean absolute error values of the all data set, 75% of data, 50% of data, and 25% of data using modified kriging models with the 2004 data set of alfalfa fields in conjunction with Aster 2001, Landsat 2001, Ikonos 2001, Landsat 2001, and Ikonos 2001 im