Ecology and Development Series No. 32, 2005

Akmal Akramkhanov

The spatial distribution of soil salinity: Detection and prediction

Cuvillier Verlag Göttingen

ABSTRACT

Inefficient irrigation and the excessive use of water on agricultural lands in the Aral Sea Basin over several decades have led to highly saline soils. Salinity appraisal in the Aral Sea Basin however, is still dependent upon traditional soil surveys with subsequent laboratory analyses for the total dissolved solids (TDS). This thesis has three specific objectives, namely, to identify techniques that enable rapid estimation of salinity, to characterize the spatial distribution of soil salinity and to estimate the spatial distribution of soil salinity based on readily or cheaply obtainable environmental parameters.

Soil salinity was measured by four electrical conductivity (EC) devices (2XP, 2P, 4P, and CM-138) on a regular grid covering an area of approximately 3 km by 4 km. Six nested samplings within selected grids were conducted to account for small-scale variation. The farm-scale (~15 km^2) results were used to upscale soil salinity to a district area (~400 km^2). Apart from widely used terrain indices and those acquired from remote sensing, distance to drains and long-term groundwater observation data were used to account for local parameters possibly influencing soil salinity. Standard statistical procedures were applied for data description, correlation between variables, analysis of variance, and regression. Characterization of the spatial distribution of soil salinity and interpolation of point data were carried out using geostatistics. Soil salinity estimation based on environmental attributes was carried out using a neural network model, as this offers enhanced generalization compared to other models. Analyses were integrated into a GIS for visualization and presentation of the results.

Techniques for rapid determination of soil salinity based on electrical conductivity were assessed and proved to be satisfactory in all cases. Measurement based on soil paste (EC_p) was highly accurate ($R^2=0.76$), whereas EC_a measurements at point scale in the field were of low accuracy ($R^2<0.5$). However, field assessment of soil salinity was considerably enhanced by the use of CM-138, because large areas can be quickly assessed, which in spite of lower accuracy, is desirable.

Topsoil (30 cm) salinity was highly variable even at short distances (40 m) compared to average soil salinity at 0.75 m and 1.5 m depth measured by the CM-138. Overall distribution of soil salinity was influenced by soil texture and topography, while at the local scale terrain attributes such as curvature, plan and profile curvatures, and solar radiation were the most influential factors. Factors obtained by remote sensing had significant correlation coefficients ($r=0.2$) with both the salinity of topsoil and salinity measured by the CM-138. Distance to drains is an important factor, especially for the bulk soil salinity (measured by the CM-138) of the profile. Correlation between distance to drains and salinity of the topsoil was low, which might be due to higher spatial variation of the topsoil salinity. Groundwater table depth and salinity had marked correlations with soil salinity; however, the direction of the influence could not be explained. The inclusion of these controlling variables in modeling is fundamental, and efforts must be directed towards obtaining reliable and accurate databases in order to derive them.

With an environmental correlation model that was built for the farm scale, soil salinity was estimated using environmental parameters in a neural network approach and shows a high correlation coefficient between estimated and measured soil salinity of 0.83. The accuracy of the prediction of soil salinity was satisfactory taking into account

that the measurement scales of soil salinity and environmental data derived from different estimations with unknown but certainly varying accuracy.

The use of environmental attributes and soil salinity relationships to upscale the spatial distribution of soil salinity from farm to district scale resulted in the estimation of essentially similar mean soil salinity values (0.94 dS m^{-1} vs. 1.04 dS m^{-1}). However, visual comparison of the maps suggests that the estimated map had soil salinity that was overly uniform in distribution, which is thought to be caused by inaccuracy of environmental data (including scale problems) or overgeneralization by the neural network model.

The upscaling proved to be satisfactory, but further research is needed before the model can be applied. Considering that the neural network model is an empirical model, further training of the model is required at locations where conditions are different to those on the farm where the model was generated. Furthermore, the model depends on the strength of the relationship between environmental variables and soil salinity. Therefore, the environmental variables must be either available for the study area in high spatial resolution or easily measurable.

Die räumliche Verbreitung des Bodensalzgehaltes: Erkennung und Vorhersage

KURZFASSUNG

Schon seit mehreren Jahrzehnten kommt auf den landwirtschaftlich genutzten Flächen des Aralseebeckens mangelhafte Bewässerungstechnik bei gleichzeitig übermäßigen Wassereinsatz zum Einsatz. Die Umweltkonsequenzen dieser Entwicklung sind durch hohe Salzgehalte verseuchte Böden. Eine Abschätzung des Salzgehalts im Aralseebecken erfolgt jedoch immer noch mit Hilfe von traditionellen Bodenuntersuchungen mit darauf folgenden Laboranalysen zur Bestimmung der Gesamtmenge gelöster Feststoffe (total dissolved solids; TDS). Diese Doktorarbeit hat drei spezifische Ziele: (i) Identifizierung der Methoden, die eine schnelle Abschätzung des Salzgehalts im Boden erlauben, (ii) Charakterisierung der räumlichen Verteilung des Salzgehalts im Boden und (iii) Abschätzung der räumlichen Verteilung des Salzgehalts im Boden mit Hilfe von vorhandenen oder leicht erhältlichen Umweltparametern.

Der Bodensalzgehalt wurde mit vier Geräten (2XP, 2P, 4P und CM-138), welche die elektrische Leitfähigkeit (EC) messen, in einem regelmäßigen Gitter, das ein Gebiet von 3 km x 4 km umfasst, bestimmt. Es wurden sechs verschachtelte Probenahmen (nested samples) innerhalb ausgewählter Gitter vorgenommen, um Abweichungen auf kleinem Raum zu berücksichtigen. Die auf Farmebene (~15 km^2) erzielten Ergebnisse wurden verwendet, um den Bodensalzgehalt auf eine Bezirksebene hochzurechnen (~400 km^2). Abgesehen von weit verbreiteten Bodenindizes und Fernerkundungsdaten wurden die Entfernungen der Messpunkte zu Drainagekanälen und Langzeit-Grundwasserdaten verwendet, um Parameter zu berücksichtigen, die möglicherweise einen lokalen Einfluss haben. Für die Datenbeschreibung, die Korrelation zwischen Variablen, die Varianzanalyse und die Regression wurden statistische Standardmethoden verwendet. Die Charakterisierung der räumlichen Verbreitung von Bodensalzgehalt und die Interpolierung von Punktdaten wurden mit Hilfe der Geostatistik vorgenommen. Die Schätzung des Bodensalzgehalts basierend auf Umweltparametern wurde mit Hilfe eines Neuralen Netzwerk-Modells durchgeführt, weil es im Vergleich zu anderen Modellen eine bessere Generalisierung erlaubt. Die Analysen wurden zur Visualisierung und Ergebnispräsentation in einem GIS integriert.

Schnelle Bodensalzgehalt-Bestimmungsmethoden auf der Grundlage elektrischer Leitfähigkeit wurden in allem Fällen zufrieden stellend bewertet. Die Salinitätsmessung in wässriger Bodenpaste (EC_p) war sehr präzise ($R^2=0.76$), doch EC_a Punktmessungen im Feld waren von geringer Genauigkeit ($R^2<0.5$). Allerdings war die Abschätzung des Bodensalzgehalts auf dem Feld beträchtlich verbessert, wenn das CM-138 verwendet wurde, weil große Flächen schnell geschätzt werden können, was trotz schlechter Genauigkeit wünschenswert ist.

Der Salzgehalt des Oberbodens (bis 30 cm) variierte sehr (sogar bei kurzen Entfernungen (40 m)) im Gegensatz zum mit dem CM-138 gemessenen durchschnittlichen Bodensalzgehalt in 0.75 m and 1.5 m Tiefe. Die Verteilung des Gesamtbodensalzgehalts wurde von Bodenbeschaffenheit und Topographie beeinflusst,

während auf der lokalen Ebene Terraineigenschaften wie *curvature*, *plan* und *profile curvatures* und Sonnenstrahlung die einflussreichsten Faktoren waren. Auf Fernerkundung basierende Faktoren korrelierten signifikant sowohl mit dem Salzgehalt des Oberbodens als auch mit dem Salzgehalt, der mit dem CM-138 gemessen wurde. Ebenso ist der Abstand zu Drainagekanälen ein wichtiger Faktor, insbesondere für den Gesamt-Bodensalzgehalt im Profil (bei Messungen mit dem CM-138). Der Salzgehalt war im Oberboden geringer, eventuell aufgrund der größeren räumlichen Verteilung des Oberbodensalzgehalts. Die Tiefe des Grundwasserspiegels und der Salzgehalt zeigten starke Korrelationen mit dem Bodensalzgehalt, doch die Ursache der Beeinflussung konnte nicht erklärt werden. Die Miteinbeziehung dieser Variablen, die einen starken Einfluss auf den Bodensalzgehalt haben, ist sehr wichtig und es deshalb sollten zuverlässige und präzise Datenbanken aufgebaut werden, damit diese Variablen erworben werden können.

Mit einem Umwelt-Korrelationsmodell, das für die Farmebene gemacht wurde, wurde der Bodensalzgehalt unter Verwendung von Umweltparametern mit Hilfe eines neuralen Netzwerks abschätzt; der Korrelationskoeffizient zwischen geschätztem und gemessenem Bodensalzgehalt war mit 0.83 hoch. Die Vorhersage des Salzgehalts mit diesem Modell war zufrieden stellend, insbesondere wenn man berücksichtigt, dass die Übereinstimmung in den Messskalen der Bodensalzgehalte und der Umweltfaktoren aufgrund der unterschiedlichen Erfassungsmethoden und unbekannter, aber sicherlich variierender Erfassungsgenauigkeiten stark variierte.

Die Verwendung von Relationen zwischen Umweltparametern und Bodensalzgehalt, um die räumliche Verteilung des Bodensalzgehalts von Farmebene auf Bezirkebene hochzurechnen, ergab im Wesentlichen übereinstimmende Abschätzungen der durchschnittlichen Bodensalzgehaltwerte (0.94 dS m^{-1} vs. 1.04 dS m^{-1}). Doch deutet ein visueller Vergleich der Landkarten darauf hin, dass der Bodensalzgehalt in der geschätzten Landkarte in seiner Verteilung zu einheitlich war. Dieses Problem wurde vermutlich durch Ungenauigkeiten in der Verteilung (einschl. Probleme bei der Skalierung) oder durch eine zu weit gehende Verallgemeinerung durch das Neuralnetzwerk-Modell verursacht.

Das Hochrechnen der Bodensalinität aus Umweltparametern hat sich als akzeptabel erwiesen, aber weitere Forschungen sind nötig, bevor das Modell praktisch angewendet werden kann. Unter Berücksichtigung, dass das neurale Netzwerk ein empirisches Modell ist, ist weiteres Training des Modells an anderen Standorten und unter anderen (sich von denen der ersten Farm unterscheidenden) Bedingungen notwendig. Das Modell hängt von der Stärke der Beziehung zwischen Umweltvariablen und Bodensalzgehalt ab. Deshalb müssen die Umweltvariablen für das Forschungsgebiet in sehr guter räumlicher Auflösung verfügbar oder einfach zu messen sein.

TABLE OF CONTENTS

1 GENERAL INTRODUCTION

1.1 Introduction

Inefficient irrigation and the excessive use of water on agricultural lands in the Aral Sea Basin over several decades have led to highly saline soils. The history of agriculture shows that irrigated agriculture cannot survive indefinitely without adequate salt balance and drainage (Tanji, 1990). An often quoted example of excessive irrigation and resulting salinization dating back to Ancient Mesopotamia is briefly explained by Hillel (2000). Similar problems are currently faced in all Central Asia, where extensive land development for irrigated agriculture with the subsequent recession of the Aral Sea is blamed for the current widespread salinization and ecological deterioration. This problem affects roughly 75% of the irrigated land in the Aral Sea basin (van Dijk et al., 1999).

Despite the ample and repeated criticism of the wasteful irrigation practices (Micklin, 1988) there has been little progress in reversing the trend of soil degradation. Detailed overviews of the environmental problems and recommendations for overcoming these problems are extensively documented in numerous reports of international organizations such as UNESCO, GEF, UNEP, as well as in individual studies. For example, Micklin (1988) attributes the causes of reduced inflow into the Aral Sea since 1960 to both climatic and anthropogenic factors. Aladin (1998) stresses that isolation from other large water basins and highly arid continental climatic conditions are the most important features distinguishing similar areas with that of the Aral Sea. However, anthropogenic factors have had particularly disastrous effects, especially the drastically high water use rates of 10,000-12,000 m^3 ha^{-1} of irrigated land and 2,500 m^3 $capita^{-1}$, compared to 5,500 m^3 ha^{-1} in Israel and 1,700 m^3 $capita^{-1}$ as the world average (Saifulin et al., 1998). The current understanding is not to refill the sea per se, but at least to stabilize the worsening environmental and social conditions of the most strongly affected adjacent territories and populations.

Since the two major rivers feeding the Aral Sea, Amu Darya and Sir Darya, pass through several countries, the affected environment covers a much wider area than that of the Aral Sea that is at the tail end of these rivers. One of the adversely affected areas in the Amudarya delta is the Khorezm region (Uzbekistan), which has received

little attention since the onset of the Aral Sea disaster. It is one of the oldest agricultural areas in the world.

Currently, the region heavily depends on river water for mainly cotton cultivation, while rice and wheat used to be the dominant crops in the past. The Khorezm region covers 455,000 ha of which 260,000 ha are irrigated (map of Ministry of Agriculture and Water Resources, 2000).

During three years of drought, from 1999 to 2001, the consequences of inefficient irrigation were hard felt by the local population in Khorezm. Furthermore, water supply problems have been revived through recent geopolitical developments in the Central Asian region with Afghanistan possibly wanting to lay claims to its share of the Amu Darya river water (Glantz, 2002). Therefore, a lack of water in the future should be anticipated and the 1999-2001 drought not just be viewed as a single event.

Located in the low deltaic plains, Khorezm is subject to high loads of salt particles in the river water (Abdullaev, 2002). Reports estimate that saline soils in the Khorezm region constitute more than 80% of the irrigated soils, which range from slightly to highly saline. According to Kust (1997), salinity change and profile distribution in this area is mainly governed by fluctuation of the groundwater table, duration of dry season, and location of the soils within the lower mesorelief elements within the river deltas. The general trend as described in soil surveys by Zhollybekov (1995) indicates that soil salinity has risen in the river delta due to the receding sea level.

It is a well known fact that high soil salinity impedes crop growth. Depending on the salinity level, crop yields are suppressed by 10-20% even by weakly saline soils (Kaurichev, 1989). For example, the maximum tolerance of cotton and wheat is to a chloride content of 0.03 g $100g^{-1}$ (ibidem). Various monitoring programs have been set up by governments to prevent and manage affected areas. These traditional soil salinity estimates are based on laboratory analysis of soil samples. However, this method requires considerable time, expense and effort. Therefore there is a need for new and more practical methods and tools to determine and predict soil salinity (i.e., quick, cheaper, on-the-spot determination, and GIS application) to improve decision making processes. Methods should target primary benefactors such as the Ministry of

Agriculture and the branches directly involved in land-use planning and water allocation.

With the need for monitoring in combination with the technological progress of salinity assessment based on electrical conductivity, more practical assessment methods have evolved. Specifically, a model by Rhoades et al. (1989) received great attention in this regard (Corwin and Lesch, 2003; Hall et al., submitted; Rhoades et al., 1999). It requires two additional parameters, clay and water content, which could be easily estimated in the field.

In part, facilitated by the availability of the quick appraisal methods, the intricate processes of soil salinity occurrence in irrigated arid regions are well understood but difficult to measure. As in other similarly difficult to measure processes, especially in soil formation, they are measured by their response. However, it can be argued that while soil processes to manifest may take thousands of years to come to equilibrium, soil salinity, being a mobile and dynamic attribute in the soil, reaches a state of equilibrium at a much faster pace (e.g., Park and Vlek, 2002). Soil salinity surveys at the landscape level in the developing world still remain the major source of information on salinity distribution, although these have many limitations, for example, conventional soil maps neither delineate all of the field's inherent variability nor represent specific soil attribute variations (Moore et al., 1993). The recent development of quantitative methods based on geostatistics and incorporation of environmental variables partly stem from practical constraints of conventional soil survey methods, which can be criticized as being too qualitative and too focused on soil management and land-use planning (McBratney et al., 2000).

1.2 Research objectives

Based on the understanding of the processes of soil salinity development and the technological progress, an intensive investigation of various aspects of the soil salinity in the setting of irrigated land has been designed. The overall research objective of the study was to develop a spatially distributed model to estimate soil salinity distribution at the farm and district level with an emphasis on readily available and cheaply obtainable measurements, so that real-time decision making can be supported both on-farm (by farmers) and off-farm (by policy makers) in Khorezm, Uzbekistan.

To achieve the overall objective, the following specific objectives were set:

i) Identify a technique for quick determination of soil salinity based on electrical conductivity devices, and comparatively assess suitability and sensitivity of the different methods;
ii) Characterize the spatial variability of soil salinity at both farm and district levels in the study region;
iii) Develop a soil salinity prediction model taking into account environmental factors (a method that would allow predicting soil salinity without actually measuring it);
iv) Predict the spatial distribution of soil salinity in Khorezm by combining remote sensing with the available soil database.

1.3 Scope of the study

Soil salinity is one of the major side effects in the agro-environment of former large-scale irrigation projects that lacked adequate water and currently threatens agriculture in the Khorezm region. The long-term German-Uzbek multidisciplinary research program on sustainable land and water use, of which the present study forms a part, was launched in 2002 with the main aim of understanding the structure of the agro-ecosystems and developing cost-effective landscape management options in this type of constrained environment. When discussing landscape management options, the use of maps is indispensable; however available maps are often obsolete. Moreover, the utility of existing conventionally surveyed maps is challenged by the development of modern mapping techniques (the shortcomings of the traditional soil survey methods are described elsewhere; (McKenzie et al., 2000; Moore et al., 1993; Tomer and James, 2004; Zhu et al., 2001). Therefore, the mapping of natural resources was a fundamental building block for the program, and hence, the present study engages not only in identifying adequate methods for field salinity measurement and assessment, but also in modeling of soil salinity, based on environmental predictors, for upscaling data from intensive field surveys onto a larger area.

1.4 Limitations of the study

Similar to many other successful attempts (i.e. Leij et al., 2004; i.e. McKenzie and Ryan, 1999; Moore et al., 1993; Tomer and James, 2004) to develop analogues of conventional surveys by terrain attributes, this study builds on readily available or easily obtainable data to account for the spatial variation of soil salinity. Therefore, the number of parameters is restricted by data availability. The accuracy of these datasets can be questioned, as they have not been specifically developed for this study, and all have intrinsic strengths and flaws (such as number of observations, representativity, layout of sampling locations, etc.), but they are regarded to be representative of the landscape and contain information valuable for salinity analysis and management. The focus here was whether it would be possible to develop models using the readily available existing data sets with their intrinsic flaws. If so, it would help decision making under the conditions encountered by farmers or decision makers. In order to achieve this, an artificial neural network (ANN) model was used. Historically, much of the inspiration in the field of ANNs came from the desire to produce artificial systems capable of sophisticated, perhaps "intelligent", computations similar to those that the human brain routinely performs. ANNs "learn" from examples, as children learn to distinguish dogs from cats based on examples of dogs and cats. If trained carefully, ANNs may exhibit some capability for generalization beyond the training data, i.e., to produce approximately correct results for new cases that were not used for training (Sarle, 1997).

1.5 Overall research procedures and thesis outline

Figure 1.1 shows the overall research procedures followed in this study. The thesis consists of 5 chapters, each with a brief introduction into the topic to be dealt with and the discussion of the results. This general introduction (Chapter 1) is followed by the comparison of electrical conductivity devices (Chapter 2) for a quick appraisal of the soil salinity. Chapter 3 characterizes the variation of soil salinity and other attributes and explores the influencing environmental factors at the farm scale. The farm scale results were then used to build a neural network model in Chapter 4. The same environmental attributes which were considered at farm level were obtained at the district scale and fed into the neural network model to estimate soil salinity at the

district scale. Chapter 4 compares the estimated and measured maps. Chapter 5 concludes with a general discussion of the major findings of this research.

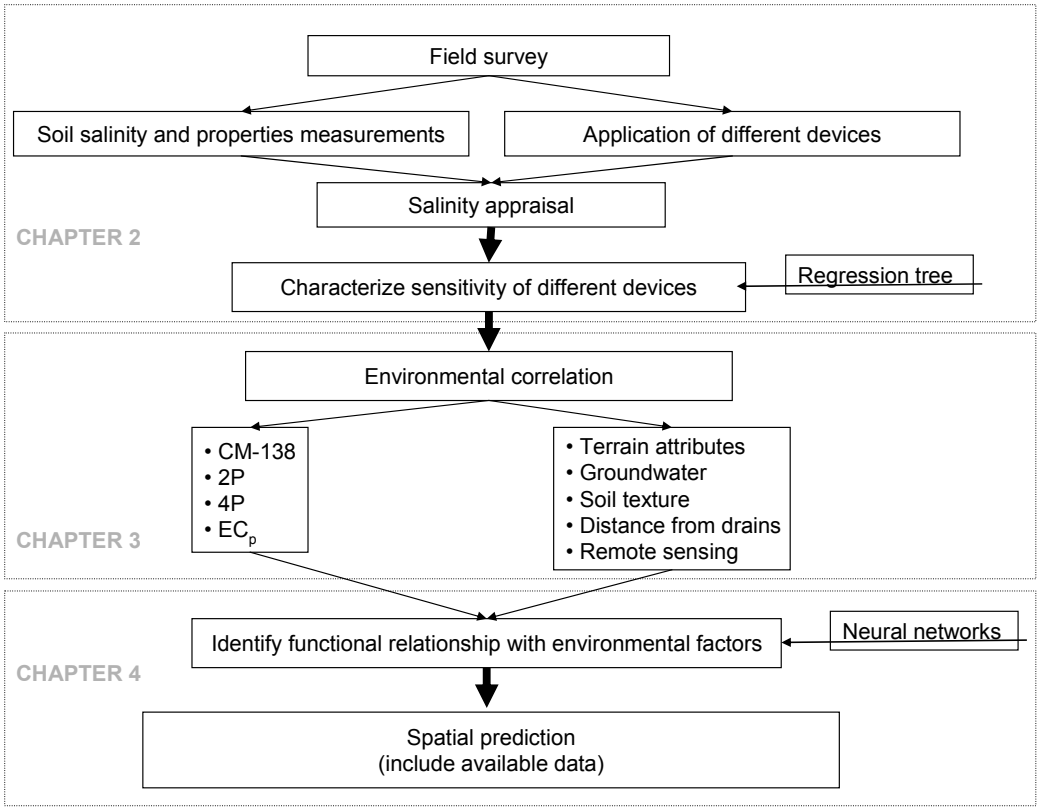

Figure 1.1 Overall research procedures followed in the study

2 COMPARISON AND SENSITIVITY OF MEASUREMENT TECHNIQUES FOR SOIL SALINITY DETECTION

2.1 Introduction

It is a well accepted fact that land use and management practices in arid regions are hampered by soil salinity. Over the last few decades, the area in the Amu Darya river delta in Uzbekistan has experienced soil salinization due to intensive farming. Timely interventions such as leaching have become ever prominent with growing deterioration of the environmental conditions caused by the over exploitation of the Amu Darya and the desiccation of the Aral Sea. Salinity appraisal in the region, however, still relies upon traditional soil surveys with subsequent laboratory analyses for the total dissolved solids (TDS).

While TDS is an important estimate of the soil salinity, it requires considerable time and resources. Furthermore, plants in saline soil are generally responsive to the concentration of the soil solution (Richards, 1954). The electrical conductivity (EC) of the solution is highly correlated with total salt concentration. The conductivity of the saturation extract (EC_e) is recommended as a general method for appraising soil salinity in relation to plant growth (Richards, 1954).

Although both approaches, TDS and EC, provide good estimates of the salinity, there are variations within the devices used and methods of analyses. Measurement of the apparent soil electrical conductivity (EC_a) can be done using electrical resistivity (Corwin and Hendrickx, 2002; Rhoades and Ingvalson, 1971; Rhoades and van Schilfgaarde, 1976), time domain reflectometry (TDR) (Dalton et al., 1984; Wraith, 2002), or electromagnetic induction (EM) (Hendrickx et al., 2002; McNeill, 1980). The latter is currently becoming one of the most frequently used techniques for characterizing the spatial variability of soil salinity.

Derivation of solute concentration from EC_a is a two-step process (Hendrickx et al., 2002). First, the electrical conductivity of the soil water (EC_w) is derived from the EC_a using an empirical regression equation or a physically based model. Next, the EC_w is converted into the solute concentration, which depends on its ionic composition. There are several models (Mualem and Friedman, 1991; Rhoades et al., 1989), which are based on the general principle that EC_a depends on several factors such as soil porosity and permeability (Archie, 1942), clay content and degree of pore saturation

(Rhoades et al., 1976). A detailed review of various models and developments is given by Hendrickx et al. (2002).

Studies on the direct comparison of EC measured with the various devices with estimates based on conventional laboratory methods are scarce, particularly in countries of the Commonwealth of Independent States (CIS) where EC_e is not a widely accepted practice. Additionally, disparity between definitions of soil textural fractions by the Kachinsky classification (Kachinsky, 1958) adopted in CIS countries, and definitions used by the Food and Agriculture Organization (FAO) might hinder the use of models where clay content is an important factor. Moreover, both soil textural class definitions do not discriminate between which salt constituent is present or dominant in the soil solution. This information is essential if the knowledge of the particular solute concentration is needed for determination of salinity type, toxicity or soil sodicity.

The main objective of this chapter is to identify and compare quick and practical determination techniques for soil salinity appraisal. Additionally, the study explores the sensitivity of each device to the individual salt constituent, using regression trees, an advanced data analysis technique (Breiman et al., 1984).

2.2 Materials and methods

Electrical conductivity of the soil is measured in different ways, the most referenced is the electrical conductivity of the saturation extract, EC_e. Electrical conductivity of the soil paste is defined further in the text as EC_p. The apparent electrical conductivity of the soil, which is a measure of the bulk electrical conductivity of the soil, is termed as EC_a.

2.2.1 Instruments

Handheld conductometer (2XP)

CIS organizations responsible for monitoring soil and groundwater salinity increasingly use electrical conductivity of the soil paste (EC_p) as an alternative to the total dissolved solids analysed in the laboratory. Rhoades et al. (1989) points out that determining EC_p as an estimate of EC_e has been in use since early 1900 in the USA.

A locally made conductometer, X-Express (Agromeliotaraqqiyot, Tashkent, Uzbekistan) (assembled by A. Chernishev, *personal communication*, 2002), tested for a

variety of soils in Uzbekistan at the Central Asian Irrigation Research Institute (SANIIRI) was used for the study. It has been proven to provide accurate measurements of EC_p, which correlated well with EC_e on a silt loam soil in Uzbekistan (Shirokova et al., 2000).

The model X-Express (2XP) has two electrodes (Figure 2.1). It measures the conductivity of the soil paste in a small glass where dry and manually ground soil (30 g) is mixed with distilled water at a 1:1 ratio. The temperature of the soil paste is measured with the same instrument. Standardization of the measured EC_p values to the reference temperature of 25°C is done using the formula provided by the USDA-Soil Salinity Laboratory.

$$EC_{25} = EC_{measured} / (1+0.02\ (T-25)) \qquad [1]$$

where:

EC_{25} - the conductivity [dS m^{-1}] of the solution at 25°C

$EC_{measured}$ - the measured conductivity [dS m^{-1}] of a solution at sample temperature

T - sample temperature (°C)

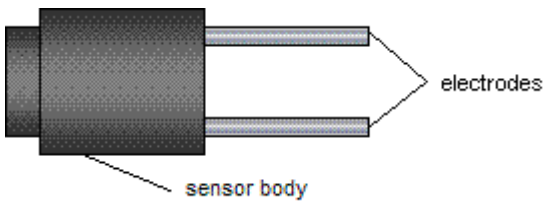

Figure 2.1 Sensor of the X-Express conductometer (Source: Chernishev, 2004)

Two-electrode conductivity probe (2P)

A two-electrode conductivity probe, Progress 1T (2P) (Agromeliotaraqqiyot, Tahskent, Uzbekistan) (assembled by A. Chernishev, *personal communication*, 2002), for field soil salinity studies measures EC_a and was tested earlier at the Central Asian Irrigation Research Institute (SANIIRI). The probe (Figure 2.2) consists of two 7-cm electrodes spaced 2 cm apart. The active part of the probe is 16 cm and measures the electrical conductivity of an approximately 20 cm layer, similar to the previous probe.

Measurements can be made at depths down to 80 cm and the probe is easy to insert into the soil due to the small 1 cm diameter of the probe. The older model did not have a temperature correction function integrated into the probe, therefore separate readings with a temperature probe were made during the measurements to calculate EC at a reference temperature of 25°C.

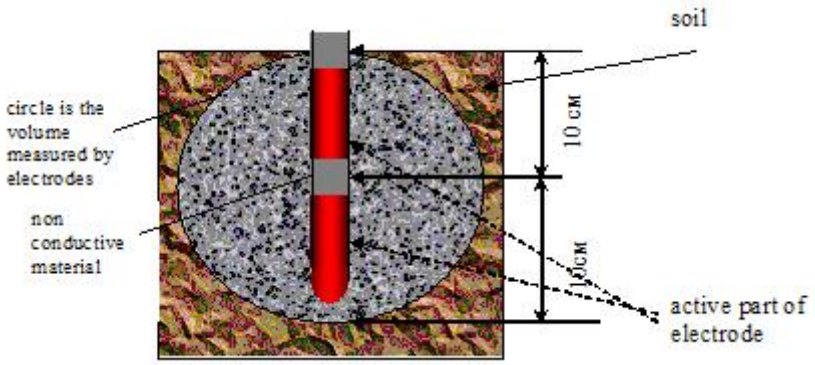

Figure 2.2 Schematic presentation of the two-electrode probe Progress 1T

Four-electrode conductivity probe (4P)

A four-electrode conductivity probe (4P) was developed by Rhoades and van Schilfgaarde (1976) based on the Wenner method. The probe has four electrodes (Figure 2.3). After an electrical current is induced between the two outer electrodes, the voltage drop between the two inner electrodes is measured. It offers EC_a measurements of smaller localized soil regions and more detailed assessment at given intervals through the soil profile, and thus is particularly helpful when salinity is highly variable between layers. However, several disadvantages exist. These include soil removal before probe insertion, which could be quite difficult in compacted soils, and the need for several measurements for the bulk salinity of a layer exceeding the sensor length. Additionally, soil moisture is a constraint factor in implementation of any type of sensors for field measurements. In this study, a commercially available probe (Eijkelkamp Agrisearch Equipment, Netherlands) was used for the measurements. The probe measures 110 cm and takes measurements of the 20 cm range interval corresponding to approximately 80 cm^3 soil volume.

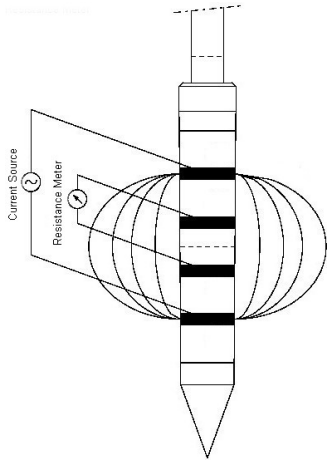

Figure 2.3 Sensor of the 4P conductivity probe (electrodes colored black)

Several electrical conductivity estimates were made in depths of 0-20, 10-30, 20-40, and 30-50 cm. The soil was removed with a gauge auger included in the set to the desired depth. The probe was inserted and the resistivity measured followed by measurement of a temperature correction factor. The EC_a then was calculated using the following equation provided by the manufacturer of the probe:

$$EC_a = k * f_t / R_t \qquad [2]$$

where:

EC_a – soil electrical conductivity [dS m^{-1}] at 25°C

k – cell constant of the probe 17.5 cm^{-1}

f_t – temperature correction factor

R_t – measured resistivity in ohms

Electromagnetic conductivity meter (CM-138)

Electromagnetic conductivity meter measures EC_a over a given depth (1.5 m in vertical position and 0.75 m in horizontal position). Of the variety of sensors, the electromagnetic induction (EM) technique is widely employed for salinity studies. The EM-38 of Geonics Ltd. (Canada) developed by Rhoades and Corwin (1981) is well documented (Corwin and Rhoades, 1982; Lesch et al., 1992; Robinson et al., 2004). Cheap and reliable prototypes of the instrument have become available, such as the CM-138 conductivity meter (GF-Instruments, Czech Republic). It has a dipole center

distance of 1 m, operates at a frequency of 14.406 kHz, with a maximum effective depth of 1.5 m in vertical positions and half that in horizontal positions. The CM-138 measures apparent electrical conductivity. The schematic work principle provided in the manual is presented in Figure 2.4.

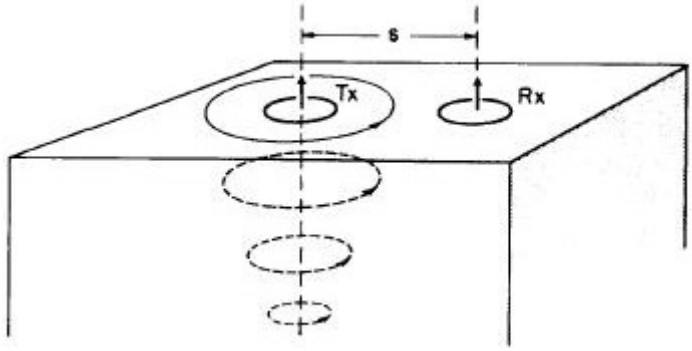

Figure 2.4 Induced current flow (homogeneous halfspace) (Source: McNeill, 1980)

The theory behind the electromagnetic technique can be found in Rhoades and Corwin (1981) and McNeill (1980). There are several advantages of the CM-138, an important one being rapid and undisturbed measurements. Contactless soil salinity measurements do not have to deal with poor sensor-soil contact in dry soils. Hence, the CM-138 can work in a variety of soil conditions, dry and wet, and is able to provide a rough vertical separation of salt distribution in the soil profile. With the latest developments, including the ability to mount the device on a mobile platform (Corwin and Lesch, 2003; Sudduth et al., 2003), it allows detailed mapping of the field. However, all these advantages come at the expense of a high price compared to the other salinity sensors. Electromagnetic devices are also sensitive to a wide range of metal objects. Additionally, latest studies by Robinson et al. (2004) report observations where high temperatures impaired EC readings. The latter is a technical shortcoming, which could be modified, rather than a principle drawback of the technique.

In an effort to measure salinity over shallower depths, several CM-138 readings held at different heights above ground were taken at each location. Measurements in the vertical mode were made at 0, 20, 40, 60, 80, and 100 cm above the soil surface (abbreviated CMv, CMv20, CMv40, CMv60, CMv80, CMv100, respectively), and in the horizontal mode only once on the soil surface (CMh). This

approach was first tried by Corwin and Rhoades (1982) and might be a means of capturing soil-surface salinity. The automatic readings were recorded by the Hewlett Packard palmtop (HP-200LX) integrated into the control unit. Readings of CM-138 are automatically temperature compensated.

2.2.2 Site description

The experiment was conducted on the research farm (41°36'N, 60°31'E) of the Urgench State University, south-west of the city of Khiva (Figure 2.5). The area is located in the transition zone where alluvial soils in the north merge with desert sand of the south. Soils are mainly formed by the activity of the Amu Darya river and to a lesser extent by irrigated cultivation (Nurmanov, 1966). The last soil survey of the area conducted by the Soil Research Institute in 1997 defines soils of the area as meadow alluvial soils, with a light to medium loamy texture and underlying silt and sands. The middle part of the soil profile normally has thin loamy layers. FAO (2003) attributes these meadow soils on alluvium and sands to gleyic and calcaric Arenosols. Myagkov (1995) reports that the upper layer of the soils in this region consists of small to fractional alluvial deposits with a thickness varying from 0.5 to 1.5 m and an infiltration rate from 0.5 to 1.0 m day^{-1}.

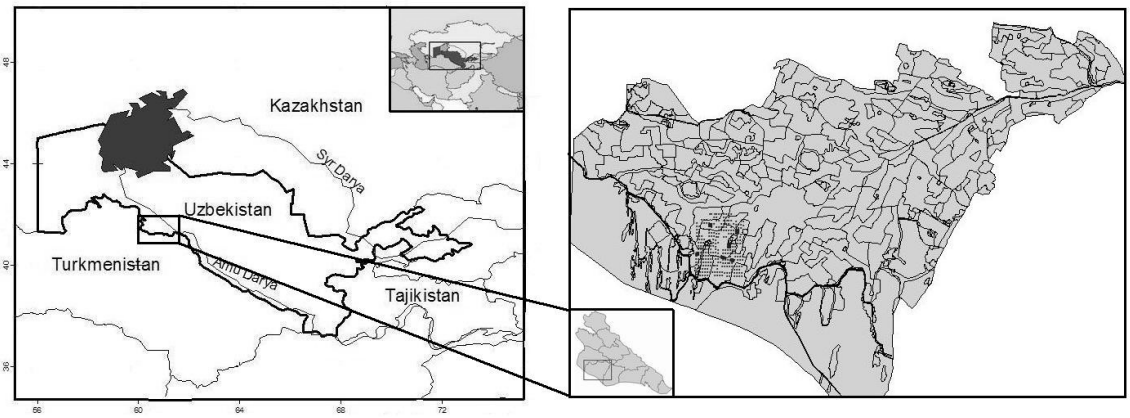

Figure 2.5 Research site location and the remote sensing image of the sampling area

Four soil pit excavations done in the area in June 2004 show that mean annual groundwater level can vary from as shallow as 100 cm in sandy fields to 180 cm in the loamy (I. Forkutsa, *personal communication*, June 2002). The sampling area (3 m x 4 km) includes soils varying from loamy to sandy. It has a relatively flat topography, with elevations ranging from 89 to 110 m above sea level. The southern

part of the sampling area borders on the main drainage water collector (known as Ozerniy collector), which carries drained water from the fields out of the region to the Sariqamish depression. Most of the area is cultivated land. However, bare or abandoned land was also included in the study to possibly increase the variability of soil salinity. The main crops grown in the area are cotton (*Gossypium hirsutum* L.), wheat (*Triticum aestivum* L.) and rice (*Oryza sativa* L.).

The climate is arid with an annual precipitation of about 100 mm (ranging from 35 mm to 170 mm during dry and wet years, respectively), about 70% of it occurring in the winter and spring. The area is irrigated by canals. Chub and Myagkov (1999) report that the TDS content of the river water at lower reaches of the Amu Darya can increase to up to 2 g l^{-1} during the low flow in the summer period, with a minimum of 0.7 g l^{-1} (data of river monitoring post; M. Ibrakhimov, *personal communication*, 2004). Mean daily temperatures vary from minus 15 to 10°C in winter (January average -8°C), and from 28-45°C in summer (July average 32°C).

2.2.3 Field survey

Field measurements were conducted from June to August 2002 in an area of approximately 1200 ha. Core sampling and EC measurements were done over a systematic 150 m by 200 m square grid designed for a parallel study. Some fields were sampled at a finer 40 m by 40 m grid (~3 ha) to study the effects of distance from water bodies, micro-topography, and to identify short range variation in successive geostatistical analyses. At each grid-node, soil core samples from 0-30 cm depth were taken, in duplicate, with a split tube sampler with an inner diameter of 53 mm. One sample was used for the analysis of the gravimetric water content and bulk density in the laboratory at UrSU. The second sample was air dried and analyzed by the Soil Research Institute (SRI) for organic matter (determined by the Turin method), saturation extract analysis (1:5 ratio) for total dissolved solids (TDS), soluble salts HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+} (by titration), Na^+ (by the difference between cations and anions), pH (by potentiometer) (Anonymous, 1977), and soil texture (pipette method) (Anonymous, 1963). Additionally, electrical conductivity (EC_p) of the air-dried soil sample was measured in the soil paste (1:1) by the 2XP handheld conductivity meter.

Electrical conductivity of the bulk soil (EC_a) in the field was measured using the three above-described devices: (i) CM-138, (ii) 4P and (iii) 2P. In order to check if 2P and 4P instruments measure similarly in the same solution, a bucket experiment with two solution types was carried out. Soil paste of low, medium, and high salinity was compared with distilled water with different amounts of salt diluted in it. Plastic buckets were used for this purpose, and electrodes were submerged into the solution to a depth of 25 cm. Readings at each salinity level and in the each bucket were taken in 5 replications.

Soil samples and measurements were, where appropriate, taken from the top of the ridge. To minimize interference from other devices, the CM-138 readings were taken first or within 2 m^2 around the CM-138 to speed up sampling, followed by soil core sampling. EC measurements were made immediately next to the core sampling spot.

Volumetric soil moisture content of the 0-30 cm depth was measured with two frequency domain moisture sensors ThetaProbe type ML2x (Delta-T Devices Ltd., UK). The sensor consists of a cylindrical head, containing an oscillator circuit, with four pointed rods (6.2 cm long) protruding from one end that are inserted into the soil. It was inserted vertically to obtain topsoil moisture content and horizontally along a 30 cm deep trench to obtain average readings for the soil moisture of the layer. In total, 6-7 readings were made per location and their average value was used for the analyses.

2.2.4 Data analysis

First, the validity of the EC_p measured with the 2XP probe was calculated by comparing those values with a calculated ECp using the soil parameters outlined in Rhoades et al. (1989) model. The Rhoades model considers the electrical current conductance through a soil, with three elements acting in parallel, (1) conductance through alternating layers of soil particles and interstitial soil solution (a solid-liquid series-coupled element, referred to as pathway 1 in Figure 2.6); (2) conductance through the interstitial soil solution (a liquid element, referred to as pathway 2 in Figure 2.6) (3) conductance through or along the surfaces of the soil particles (primarily associated with exchangeable cations) in direct contact with one another (a solid element referred to as pathway 3 in Figure 2.6)

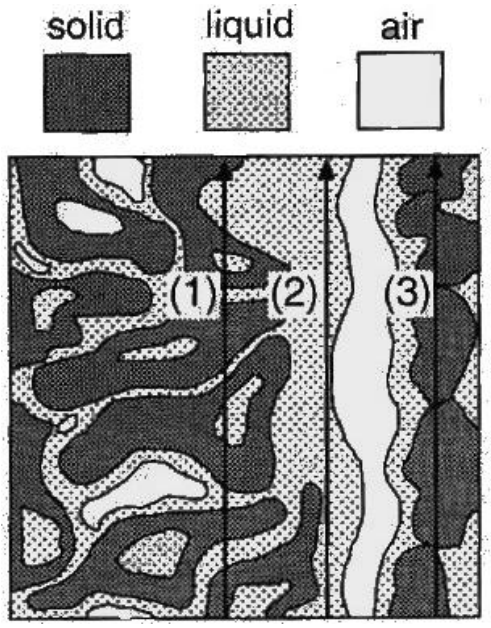

Figure 2.6 Schematic representation and model of the three electrical conductance pathways through soil: soil-liquid (1), liquid (2), and solid (3) (Source: Rhoades et al., 1989)

The model calculates EC_p using (1) the volumetric content of soil (θ_s), (2) the volumetric content of soil water in the series-coupled pathways (θ_{ws}), essentially the water in the fine pores, or so-called immobile water, (3) the volumetric content of soil water in the so-called continuous pathway, essentially the water in the large pores, or so-called mobile water (θ_{wc}), (4) the volumetric content of soil water (θ_w), (5) the average electrical conductivity of the soil particles (EC_s), (6-8) electrical conductivities of the soil water components EC_{ws}, EC_{wc}, and EC_w, (9) saturation percentage (SP), and (10) percent clay content of the soil (%clay). We measured θ_w, clay content, and TDS. Sensitivity analyses and practical use of these models was demonstrated by Rhoades et al. (1989; 1990) together with indication of which parameters can be estimated accurately or which of them should be measured.

$$EC_p = \left[\frac{(\theta_s + \theta_{ws})^2 EC_e EC_s}{(\theta_s) EC_e + (\theta_{ws}) EC_s} \right] + (\theta_w - \theta_{ws}) EC_e \qquad [3]$$

where:

$$EC_e\ (dS\ m^{-1}) = TDS\ (ppm) / 640 \quad [4]$$

$$(\theta_w - \theta_{ws}) = 0.0237(SP)^{0.6657} \quad [5]$$

$$EC_s = 0.019(SP) - 0.434 \quad [6]$$

$$\theta_s = 1 - \theta_w \quad [7]$$

$$\theta_{ws} = \theta_w - 0.0237(SP)^{0.6657} \quad [8]$$

$$SP = 0.76(\%clay) + 27.25 \quad [9]$$

More details of the equations and their meanings are given in Rhoades et al. (1989).

Similarly, to check the validity of the 2P, 4P, and CMv and CMh (devices measure EC_a) an equation provided by Rhoades was used to estimate EC_a (Rhoades et al., 1989), which assumes $EC_w = EC_{ws} = EC_{wc}$. The model calculates EC_a using (1) the volumetric content of soil (θ_s), (2) the volumetric content of soil water in the series-coupled pathways (θ_{ws}), essentially the water in the fine pores or so-called immobile water, (3) the volumetric content of soil water (θ_w), (5) the average electrical conductivity of the soil particles (EC_s), (6, 7) electrical conductivities of the soil-water components EC_{ws} and EC_{wc}, (8) saturation percentage (SP), (9) percent clay content of the soil (%clay), (10) bulk density of the soil (ρ_b), and (11) average density of the soil particles (ρ_s) assumed to be 2.65 g cm^{-3}.

$$EC_a = \left[\frac{(\theta_s + \theta_{ws})^2 EC_{ws} EC_s}{(\theta_s) EC_{ws} + (\theta_{ws}) EC_s} \right] + (\theta_w - \theta_{ws}) EC_{wc} \quad [10]$$

where:

$$EC_w = EC_{ws} = EC_{wc} = EC_e \left(\frac{\rho_b}{\theta_w} \cdot \frac{SP}{100} \right) \quad [11]$$

$$\theta_{ws} = 0.639\theta_w + 0.011 \quad [12]$$

$$EC_s = 0.023(\%clay) - 0.021 \quad [13]$$

$$\theta_s = \rho_b / \rho_s \quad [14]$$

Sensitivity of the equipment to individual ions was analyzed using the classification and regression tree CART 5.0 (Salford Systems, USA) software. The regression trees were successfully applied by others in producing high resolution soil maps and predicting soil properties (McBratney et al., 2000; McKenzie and Ryan, 1999) and in ecology (Martius et al., 2004). Regression tree analysis allows better-explained differentiation by a splitting variable, which splits a dataset into increasingly homogeneous subsets.

The decision of a CART on which variable and what value to use to split the dataset is based on how homogenous the resulting subsets are. CART uses two-way splits. One parent node is split into only two child nodes using, for example, a threshold value for the continuous data (i.e., <5 goes into one bin and >5 goes into another bin). Subsequent divisions of the dataset are carried out until the resulting bins have a minimum number of cases (1 in our analysis).

A tree with one node per observation is termed an overfitted tree and will perform poorly on a new dataset. To avoid such overfitting, a *v*-fold (*v* is equal to 10 in this analysis) cross validation (CV) procedure is used. This CV procedure randomly divides the dataset into *v* different subsets and constructs *v* regression trees, each time leaving one subset for testing. The CV error is then computed averaging the errors of the subsets which were left out during regression tree building. Additional information regarding the procedures and properties of estimates obtained from CART are discussed in Breiman et al. (1984).

2.3 Results

2.3.1 Descriptive statistics

Table 2.1 presents a statistical summary of the data used to compare measured EC_p with calculated EC_p. Due to the settlements located within the sampled area, equipment performance, and soil conditions, the total number of samples for each measured parameter varied. To have an equal number of measured EC_p, TDS and other variables used in the Rhoades model, the total number of cases was reduced to 264 cases. The readings for the EC_p of the soil paste ranged from 0.31 to 16.83 dS m^{-1}, with a mean of 2.2 dS m^{-1}. A soil with $EC_e > 4$ dS m^{-1} is generally considered to be saline (Richards, 1954), while at 6-7 dS m^{-1} yields of cotton and wheat, the major crops in the area, are

already reduced by 20% (Maas and Hoffman, 1977). TDS readings ranged from 0.07 to 3.28 g $100g^{-1}$, with a mean of 0.35 g $100g^{-1}$. Depending on salinity type, Kaurichev (1989) reports soils are considered to become saline at TDS > 0.15 g $100g^{-1}$ and for plants, Richards (1954) reports that some are adversely affected at 0.1 g $100g^{-1}$.

Table 2.1 Statistical summary of selected variables

Variable	Median	Mean	Std. Dev	Min	Max	CV	Skewness	N
EC_p (dS m^{-1})	1.75	2.2	1.77	0.31	16.83	80.45	3.24	264
TDS (g $100g^{-1}$)	0.27	0.35	0.33	0.07	3.28	94.29	4.75	264
(calc) ECe (TDS640) (dS m^{-1})	4.14	5.41	5.09	1.02	51.25	94.09	4.75	264
CLAY (<0.001) (%)	11.55	11.61	3.08	4.6	21.3	26.53	-0.05	264
CLAY (<0.002) (%)	15.4	15.32	4.84	4.59	31.48	31.59	0.15	264
CLAY (<0.01) (%)	33.8	33.92	9.24	10.6	58.4	27.24	-0.03	264

The soil texture analyses represented here as clay content at three particle sizes show that the average 'physical' clay content (particle size <0.01 mm) is expectedly the highest (34%) followed by clay content expressed in the international classification system (15%) and clay content by the Kachinsky system (12%).

Table 2.2 summarizes data for the three EC_a measuring devices. The dataset was reduced provide an equal number of observations for each device. This was due to the difficulty encountered when pushing particularly the 4P into the soil. It is thought that measurements of the 10-30 cm and 20-40 cm layers were possible only because of the wet soil condition. Hanson and Grattan (1990) experienced similar problems with this 28 mm diameter 4P, noting that it was extremely difficult to insert it into the soil, even with a pilot hole, and equally difficult to extract. Thus, to avoid bias due to the soil conditions (wetness, compaction, etc.), which affected one device while with another device it was still possible to take measurements (with difficulty though), it was decided to equalize the number of observations for all devices. Nevertheless, the remaining samples are well distributed within the sampling area and reflect different textures and, to lesser extent, crop types (Table 2.3).

Table 2.2 Descriptive statistics for three EC_a devices

EC_a variable measured by*	Median	Mean	Std. Deviation	Min	Max	CV	Skewness	N
	----------dS m^{-1}----------					%		
2P (0-20 cm)	0.34	0.52	0.46	0.08	2.37	88.33	1.55	71
2P (10-30 cm)	0.43	0.68	0.56	0.07	2.94	81.37	1.68	71
2P (20-40 cm)	0.53	0.80	0.77	0.11	5.41	96.06	3.43	71
4P (0-20 cm)	0.26	0.50	0.46	0.09	1.83	92.83	1.35	71
4P (10-30 cm)	0.33	0.61	0.65	0.10	4.07	106.93	2.86	71
4P (20-40 cm)	0.41	0.69	0.71	0.13	4.50	101.71	2.86	71
CMv	0.71	0.76	0.25	0.47	1.85	33.18	1.44	71
CMv20	0.60	0.66	0.22	0.40	1.37	32.86	1.26	71
CMv40	0.53	0.59	0.17	0.35	1.14	29.84	1.23	71
CMv60	0.46	0.52	0.14	0.29	0.95	27.84	1.15	71
CMv80	0.45	0.49	0.11	0.33	0.84	23.20	1.07	71
CMv100	0.41	0.44	0.09	0.32	0.69	20.17	1.21	71
CMh	0.65	0.73	0.24	0.44	1.66	33.21	1.60	71

**2P (0-20 cm)=EC_a of 20 cm soil layer; 4P (0-20 cm)=EC_a of 20 cm soil layer*
CMv and CMh=EC_a in vertical and horizontal positions; CMv20=EC_a measured by CM-138 at 20 cm above ground

Table 2.3 Samples located within different texture-vegetation groups

Soil type	Crop						Total
	Alfalfa	Bare field	Cotton	Maize	Mixed	Wheat	
Loam		1				1	2
Loamy sand			7				7
Sand			24				24
Sandy loam	1	1	4				6
Silt loam		1	24	3	1	3	32
Total	1	3	59	3	1	4	71

Measured values varied widely for all sensors. Interestingly, average electromagnetic signal readings (measured by CM-138) show low coefficients of variation while probe CVs (measured by 2P and 4P) are 3 times higher. Low CV for CM-138 were also observed by Hassan et al. (1983) and Hanson and Grattan (1990), who pointed out the dependence of variation on sampling volume. The sampling volume of CM-138 is about 1 m^3 compared to a volume of 80 cm^3 measured by the 4P device.

2.3.2 Comparison of instruments

Similarity of 2P and 4P instruments

The bucket experiment allowed the comparison of two instruments in two types of solutions. Readings made in soil paste and the plot of 2P versus 4P is presented in

Figure 2.7. The assumption that two instruments measure similarly is supported by the strong linear relationship and the slope in the regression equation, which equals 1. Identical measurements in the distilled water solution with varying levels of salinity also confirm the assumption (Figure 2.8). The relationship in water solution becomes even stronger as salinity levels increase.

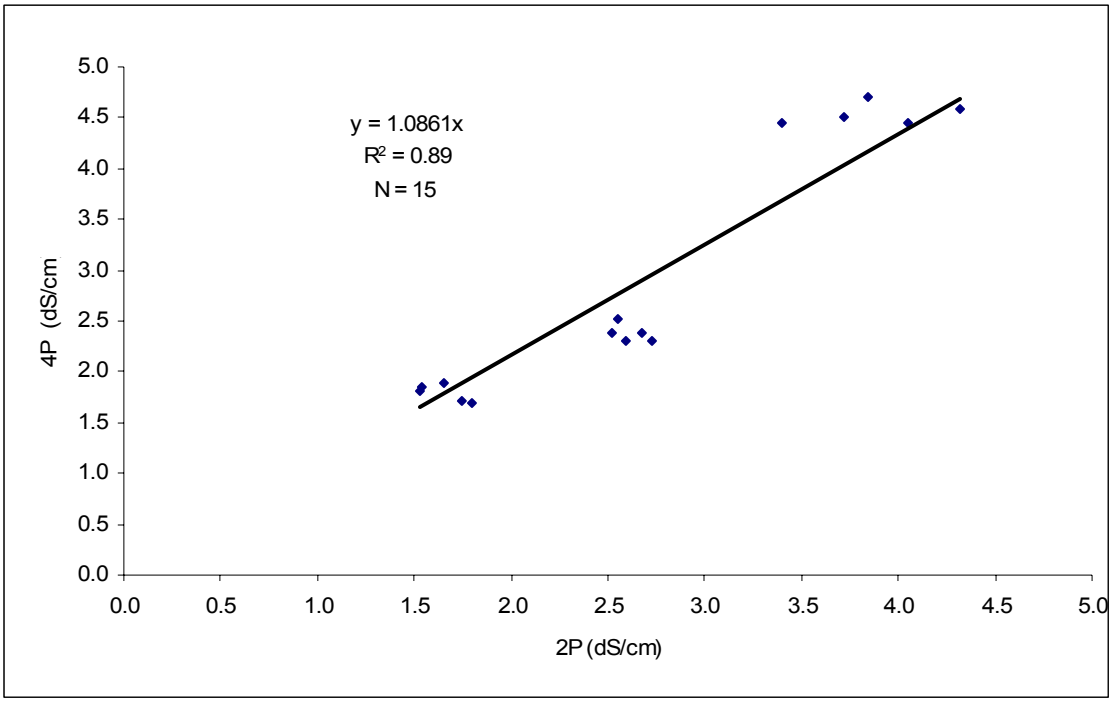

Figure 2.7 Relationship between 2P and 4P instruments as measured in soil paste

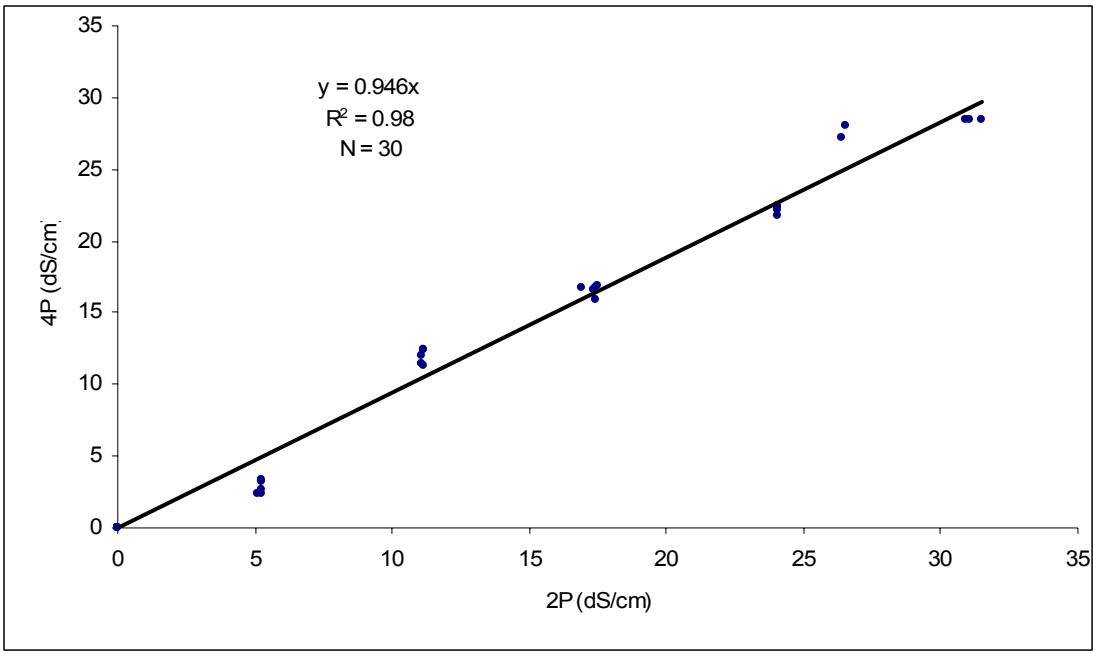

Figure 2.8 Relationship between 2P and 4P instruments as measured in distilled water solution with subsequent addition of salts

Results demonstrate that under ideal conditions, the 2P instrument is as sensitive to salinity as the commercially available and extensively tested 4P instrument. Given that there is a strong linear 1:1 relationship, the 2P readings in the field could be equated to the second replicate of the 4P instrument measurement. In this sense, the two readings of EC probes per sampling location could serve as an indication of soil salinity variability at this fine scale.

However, the 2P and 4P readings at low salinity levels seem to deviate from each other more than at higher salinity levels in aquatic solution. In this low salinity range, salinity assessment is most important because of crop sensitivity and sharp yield decreases. There is consequently less interest in high salinity ranges when one device under- or overestimates readings, because high accuracy at those levels is not essential for crops. Nevertheless, the slight variation of 2P at lower salinity levels in soil paste and good match to 4P is well demonstrated with this bucket experiment and justifies equating these devices.

EC_e and EC_p measured with 2XP

Equation [3] was used to calculate EC_p from known EC_e (from Eq. 4) and other measured values. Results of the correlation of measured EC_p and calculated EC_p using

clay content as particle size of <0.001, <0.002, and <0.01 mm, yielded a correlation coefficient (*r*) of 0.875, 0.872, and 0.867, respectively. Since one objective was to examine the use of the EC_p to estimate EC_e, the possibility that different estimation procedures result in significant differences was explored. However, different procedures for estimating the parameters of Eq. 3 and the expression of the clay content in three different particle sizes show approximately similar results, reflected in correlation coefficients of around 0.87. No discernible difference between clay content expressions could thus be detected, probably caused by the low clay content of the samples and similarity of the frequency distribution of the soils, irrespective of the definition of "clay" (Figure 2.9).

Although their skewness is low, the histograms of the clay content in different definitions show that their distribution is bimodal (Figure 2.9). Thus, the changed definition of clay fraction does not appear to change the classification of the soil and thus may not have a great effect on the correlation between measured and calculated EC_p and EC_a, respectively. For further analysis, the clay content (<0.001 mm) in the locally accepted classification was used, which does not require conversion into an international system.

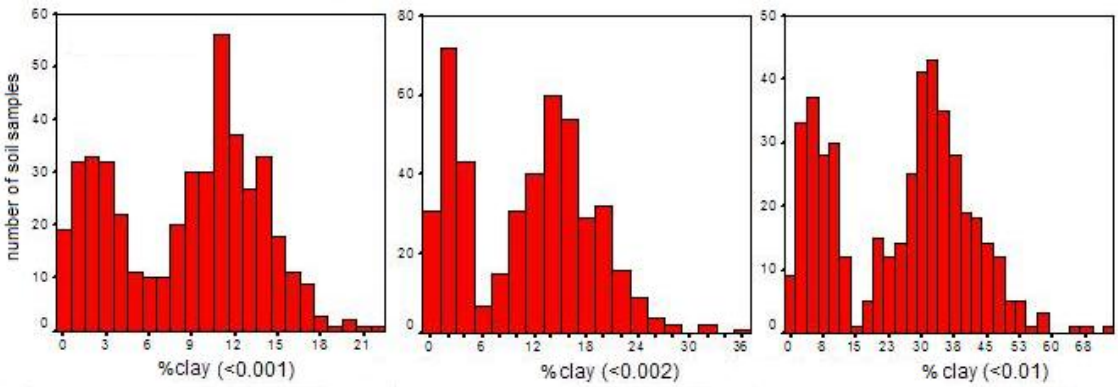

Figure 2.9 Histograms of clay percentage of soils sampled in this study, using different clay particle sizes (<0.001, <0.002, <0.01)

The regression equation between EC_p measured and calculated using the local definition for clay (<0.001) thus yields:

$$EC_p \text{ measured} = 1.152\ (EC_p \text{ calculated}) + 0.152 \qquad [15]$$

and demonstrates the fit between measured and calculated EC_p, with reasonably good R^2 (0.77). However, the R^2 is lower than that reported by other authors (Rhoades et al., 1989).

Since local authorities more frequently work with TDS, which is related to EC_e (Eq. 4), measured EC_p was also correlated with EC_e, which yields:

$$EC_e \text{ measured} = 2.52\ (EC_p \text{ measured}) - 0.159 \qquad [16]$$

The estimate of the EC_e from measured EC_p also has reasonably high R^2 (0.76) and low intercept. The slope of 2.52 is close to the 2.2 reported in Landon (1984), but lower than the 3.5 established for some soils in Uzbekistan (Shirokova et al., 2000). This could mean that for this location the coefficient of conversion is slightly different to that of other places in Uzbekistan and reflects site-specific features.

On the other hand, the lower R^2 than reported elsewhere is perhaps due to subsampling errors, when a large amount of soil (600-900 grams) is collected from the 30 cm depth, and a small subsample is taken from it for the analyses of TDS. Larger subsample sizes are reported to have less variability but at the same time showed a slower recovery rate of applied Cl (Hassan et al., 1983). Rhoades et al. (1990) also stress that sample variability can be due to differences in volumes and locations of soil used to measure salinity. They judged the error involved to be appreciable and to result in low R^2 values in the instrumental/model comparison. Further errors may have been introduced during analyses for TDS and EC_p, which were conducted separately in different laboratories.

Considering that TDS itself is highly variable and represents low salt concentrations, and that EC_p was able to explain 76% of the variability of the TDS, it can be concluded that EC_p can be used to estimate salt concentration in soils.

EC_e from EC_a measured with 2P, 4P, and CM-138

Correlation coefficients between apparent soil electric conductivity measured with the 2P, 4P, and the CM-138 and calculated EC_a using Eq. 10 are given in Table 2.4. Linear regression had to be done using transformed data due to the non-normal distribution of

the data. Natural logarithms were used for all except the CM-138 data, for which reciprocal transformation was used.

Table 2.4 Correlation table of measured EC_a using different devices versus calculated EC_a

EC_a variable measured by*	Correlation coefficients of calculated EC_a					
	EC_e (TDS/640)			EC_e (2.52EC_p-0.152)		
	Clay			Clay		
	<0.001	<0.002	<0.01	<0.001	<0.002	<0.01
2P (0-20 cm)	0.41	0.47	0.55	0.60	0.62	0.66
2P (10-30 cm)	0.55	0.61	0.69	0.73	0.75	0.78
2P (20-40 cm)	0.51	0.57	0.65	0.74	0.76	0.80
4P (0-20 cm)	0.61	0.66	0.72	0.74	0.75	0.76
4P (10-30 cm)	0.54	0.60	0.67	0.74	0.76	0.79
4P (20-40 cm)	0.53	0.59	0.67	0.73	0.75	0.79
CMv	0.54	0.61	0.69	0.67	0.70	0.73
CMv20	0.53	0.61	0.69	0.69	0.72	0.76
CMv40	0.53	0.61	0.69	0.70	0.73	0.76
CMv60	0.52	0.59	0.67	0.66	0.69	0.72
CMv80	0.46	0.53	0.61	0.62	0.65	0.69
CMv100	0.51	0.56	0.61	0.55	0.58	0.61
CMh	0.53	0.61	0.70	0.74	0.77	0.81

**2P (0-20 cm)=EC_a of 20 cm soil layer; 4P (0-20 cm)=EC_a of 20 cm soil layer; CMv and CMh=EC_a in vertical and horizontal positions; CMv20=EC_a measured by CM-138 at 20 cm above ground Correlations were significant at the 0.01 level (2-tailed)*

Calculation of EC_a using three distinct approaches and different clay representations show varying correlation coefficients. For all devices, the best results are obtained with a clay content represented as locally termed 'physical' clay as defined by the Kachinsky method (particle size <0.01). The low correlation coefficient for the 2P in the topsoil is probably the result of poor electrode-soil contact due to the dry and loose topsoil. The correlation improved in lower layers and is similar to correlation coefficient obtained for 4P.

For the highest correlation coefficients for each device and a clay content defined as 'physical clay' (<0.01) the following regression equations were constructed. The remaining measurements were not used to build regression equations, because they are highly correlated with each other and would have resulted in similar equations. The dependent variable was EC_a measured by different devices and the independent variable EC_a calculated using Eq.10.

$\ln EC_a$ measured by 2P (10-30 cm)	= 0.8 ($\ln EC_a$ calculated) - 0.96	[17]
$\ln EC_a$ measured by 4P (0-20 cm)	= 0.91 ($\ln EC_a$ calculated) - 1.4	[18]
$\ln EC_a$ measured by CM-138 (horizontal)	= 1.63 – 0.41 ($1/EC_a$ calculated)	[19]

The accuracy of the devices was somewhat lower (R^2 0.44, 0.48, and 0.47 for the 2P, 4P, and CMh, respectively) compared to previous studies (Rhoades et al., 1990). This result should be treated as indicative, since the 2P and 4P readings of the 20 cm layer are compared with the EC_a of bulk topsoil (30 cm). Therefore, 2 or 3 incremental readings by these probes should improve the accuracy. The same applies to the CM-138 readings, i.e., the device shows reasonable accuracy with only one horizontal reading and good potential for determining the depth-weighted salinity of the layers of interest. There are already established techniques for calibrating the instrument (Corwin and Rhoades, 1982; Johnston et al., 1997; Triantafilis et al., 2000), but because they are usually site specific and cannot be readily implemented for other areas or for upscaling, they will not be covered in this paper. Correlation of EC_a to TDS, which is related to EC_e (Eq. 4) yielded very low accuracy and is not discussed further.

2.3.3 Sensitivity of devices to individual salt ions

It is well recognized that different salt ions contribute to electrical conductivity in the soil. In this study, the salinization type was predominantly chloride-sulphate according to the classification used by local laboratories (classification described in Kaurichev, 1989, p. 486; (Table 2.5). The relation between the electrical conductivity and the salt content of various solutions is reported by Richards (1954). He observed that the EC curves for the chloride salts and Na_2SO_4 almost coincide, but $MgSO_4$, $CaSO_4$, and $NaHCO_3$ have lower conductivities than the other salts at equivalent concentrations. Chloride is usually the predominant anion in salt-affected soils and is non-reactive with other components (Robbins and Wiegand, 1990).

Table 2.5 Summary statistics of ions (in meq)

	Mean	Standard deviation	Min	Max	CV%	Skewness	N
Anions	4.88	2.98	1.09	16.03	61.10	1.44	71
Cations (Ca+Mg)	2.74	1.74	0.99	9.78	63.74	2.11	71
HCO3	0.40	0.10	0.20	0.61	24.92	-0.01	71
Cl	1.27	1.26	0.28	7.79	99.51	2.94	71
SO4	3.21	2.21	0.37	9.96	69.01	1.12	71
Ca	1.66	1.15	0.50	6.74	69.10	2.14	71
Mg	1.08	0.82	0.08	4.27	76.46	1.32	71
Na	2.15	1.88	0.02	7.37	87.49	1.11	71

The question is, what ions are the most dominant for determining the signal of the various devices, and are these consistent across devices. Considering that certain ions contribute more to the EC of the devices, sensitivity analyses were carried out using a classification and regression tree (CART). The software CART 5.0 generates a summary report that lists the variable importance used as a splitting variable when constructing the tree. The most important variable has a ranking of 100, and the remaining variables are ranked in decreasing order of importance. Assuming that the variation range of the contribution of individual ions to total EC is not substantial, the contribution of ions to EC_a and EC_p can be ranked.

The response variable in this regression tree was EC_a measured by 2P, 4P, and CM-138, EC_p measured by 2XP, and the explanatory variables used were individual ions (HCO_3^-, Cl^-, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+). A regression tree was built for each depth (for 2P and 4P) and height (for CM-138) increments, and variable importance data were obtained from the summary reports. The results for the 2P and 4P devices at different depths were averaged (Figure 2.10). Only one horizontal CM-138 reading is presented in the graph, because its response curve is closer to the surface where the analysis of ions is obtained, while the other CM-138 measurements were similar to the CMh measurements.

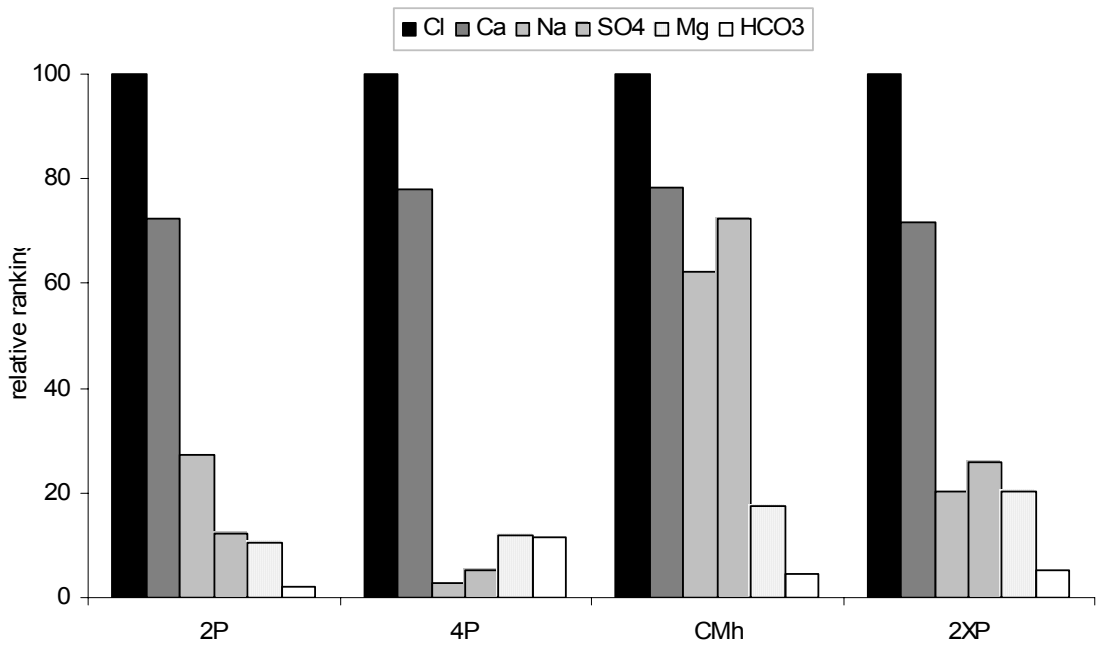

Figure 2.10 Ranked contribution of salt ions (expressed in meq) to the EC_a and EC_p of devices as determined by regression tree analysis (CART 5.0)

Chloride and Ca are the most sensitive ions for all the devices. The other ions vary between devices but, in general, ions rank similarly for 2P and CM-138, whereas for 4P the rankings are different. However, the overall sensitivity of all devices to Cl suggests that the measuring principles are similar for all 3 instruments. The devices can therefore be used interchangeably The likely reasons for the dissimilarities in sensitivity to ions between devices will be presented in the discussion.

2.4 Discussion

There are no easy or straightforward analyses for measuring soil salinity, because of temporal and spatial variability prevalent in-situ. The electrical conductivity instruments used in the study to measure soil salinity performed reasonably well given the number of influencing factors such as varying soil texture, ranging from sandy to loamy. Approximations such as representation of clay content in different systems or as EC_e calculated from TDS could have introduced some errors. Nevertheless, as a first step, the use of EC_p measured by 2XP to replace laborious and time consuming EC_e analyses proved to be successful.

Of the three EC_a devices, the 2P showed the lowest accuracy. There are several reasons why 2P could have performed less well, one being technical. The 2-

electrode conductivity instrument is prone to polarization, contamination and cable resistance errors (Thermo, 2004). The manual explains that polarization errors occur at the boundary layers between the measuring electrode and the ion-conducting measuring medium. Contamination due to deposits on the electrode surface has a similar effect, i.e., the conductivity reading is lower than the actual value. Cable resistance adds to the measured sample conductance. All of these shortcomings are overcome by 4-electrode probes.

Despite these disadvantages, slope and intercept values of 2P and 4P (Eq. 17 and 18) are close, which is not the case for the CM-138. Again, there is a technical reason for this. In the horizontal mode, the CM-138 measures the bulk of 75 cm, and the laboratory analyses are not based upon a similar depth of measurement. Similar depth-related differences and degrees of soil profile layering between sites were observed by Sudduth et al. (2003).

A single reading by CM-138 offers additional information about the salt in the soil profile. Generally, high CM-138 readings mean high salt storage within the profile and, depending on management of the field and climate, salt can move towards the surface very rapidly as observed by many researchers (Hendrickx et al., 1992; Norman, 1988). The authors verify the well-known fact that the salinization process in arid areas is relatively fast, and that irrigation and soil management in flat irrigated lands determine the spatial variability of salinity to much larger extent than prevalent soil characteristics.

Apart from these physical hindrances in EC measurements (those mentioned earlier include soil-electrode contact problems, cable resistance, polarization, contamination, and response curves), there are theoretical problems that are hard to account for. Although the EC of individual ions is known, their conductance varies with the kind of the soil (Li, 1997). As the author explains, the EC of ions in a colloidal system is determined by the distribution of these ions between the electric double layer and the free solution and their distribution within the double layer. These distributions are dependent mainly on the surface charge density of the soil colloid.

Another effect of the electric conductance of the soils mentioned by Li (1997) is the frequency of the applied current especially in the presence of electrolytes. He also notes that the effect of anions on conductivity dispersion is greater than that of cations.

Although the author mainly discusses the effect on variable charge soils, the principles equally apply for constant charge soils, which can be seen from the sensitivity analyses, where all Cl anions ranked the first for all devices.

2.5 Conclusions

Soil salinity assessment using electrical conductivity provides a quick and inexpensive alternative to laboratory-based analyses. In this study, two commercial and two locally assembled EC sensing devices were compared. The main emphasis was on practicality and device accuracy over diverse field conditions without differentiating soil types.

The locally assembled 2XP for measurement of electrical conductivity in the soil paste (EC_p) was checked against the calculated EC_p using the Rhoades model and showed 77% accuracy. The 2XP also estimated TDS (which is related to EC_e) as commonly used by local authorities with 76% accuracy. Based on this analysis, it can be concluded that the 2XP can replace laboratory measurements of TDS or EC_e with high confidence.

The locally assembled two-probe EC_a meter (2P) measured more accurately over a wide range of soil salinity values than a commercial 4-electrode probe (4P). This demonstrates that the locally assembled 2P can be used to accurately estimate electrical conductivity.

The EC_a values measured by the 2P, 4P, and CM-138 devices were generally less accurate than EC_a calculated using the Rhoades model ($R^2 < 0.5$). On the other hand, the equal sensitivity of all devices to Cl and Ca demonstrates that device validity is acceptable.

The direct estimation of TDS from EC_a (measured by 2P, 4P, or CM-138) however, was not satisfactory. Differences in the measured volume of TDS and EC_a and the use of TDS conversion instead of real EC_e measurement are believed to be the main factors complicating the direct conversion from EC_a to TDS.

3 ENVIRONMENTAL FACTORS INFLUENCING SPATIAL DISTRIBUTION OF SOIL SALINITY ON FLAT IRRIGATED TERRAIN

3.1 Introduction

Salinization processes on a landscape level have already been well studied. Rhoades et al. (1999) summarizes that salinization is caused by the effects and interactions of varying edaphic factors (soil permeability, water table depth, salinity of perched groundwater, topography, soil parent material, geohydrology), by management-induced factors (irrigation, drainage, tillage, cropping practices), as well as by climate-related factors (rainfall, amount and distribution, temperature, relative humidity, wind). More specific, Valenza et al. (2000) concluded that in semi-arid regions; the extent of surface saline soils is due to (i) existence of ancient saline soils, (ii) irrigation, and (iii) intensive evapotranspiration.

Salinity is complicated by the common and substantial variability in soils and their attributes. Pariente (2001) observed that areas with low precipitation (less than 200 mm rain) showed higher temporal heterogeneity, rate of change, potential of change and differences between soil layers than those with higher precipitation (over 200 mm). Collection of information is often complicated due to spatial variability and rapid changes in soil salinity. Conventional appraisal procedures rely on soil sampling and laboratory analyses and do not provide sufficient spatial information of salinity distribution and natural or management-related causes. As Guitjens and Hanson (1990) correctly note, engineers consider spatial and temporal variability to be explicit parts of salinity analysis, yet they rarely apply the statistical techniques used to characterize variability as routinely as they use more traditional methods. Statistical models can help to determine and evaluate spatial variability.

The development of new statistical methods now allows the use of data sets generated or sampled on different spatial scales. The use of interpolation techniques to predict values at unsampled locations adds more opportunities to combine factors which were previously hard to incorporate. There is a large emphasis on this, since there are economic and logistical reasons for including the ancillary factors influencing soil variability, especially if the ancillary factors are more readily and cheaply available (McBratney et al., 2000).

For the purpose of characterizing spatial variability of soil salinity, the concept of soil-landscape relationship can be applied. The theoretical details of this concept can be found in the reviews by Scull et al. (2003) and McBratney et al. (2000). Initially, the approach was conceived by Milne (1935), who observed a repeated change of soil types from crest to hollow throughout the vast area of East Africa, where the landscape is undulating or hummocky. This "catenary" soil development, as it has been coined by the author, has been successfully used to predict soil attributes (McBratney et al., 2000; Moore et al., 1993). Application of the concept to characterize more dynamic soil attributes such as exchangeable cations and soil moisture was also successful (Burt and Park, 1999; Park and van de Giesen, 2004). Similar concepts with additional environmental variables were used for salinity studies (Evans and Caccetta, 2000; Odeh et al., 1998; Searle and Baillie, 1997).

In most of those studies, the assumption was that the parent material was uniform. Furthermore, most of them were made in landscapes that had not been altered by irrigation projects, and they often covered catchments. In irrigation schemes, human-induced factors (manipulation of drainage for rising groundwater level for subirrigation) in some areas and the condition of the drainage network in combination with a relatively flat terrain are important in the redistribution of soil salinity. There is a lack of studies in locations under irrigation or areas that have been extensively modified by humans. An understanding of salinity distribution in such areas has important implications for the build-up of salts in soils. Capturing these human-induced factors is difficult, but geostatistical and geospatial techniques may need to be tested for this purpose.

The main objective of this study is to characterize the spatial variability of soil salinity at the farm level and to determine the spatial dependence of soil salinity on environmental factors. To achieve the objective, the conceptual framework described below is implemented. First, the geostatistical analyses are used for characterization of the spatial distribution of soil salinity and related attributes. Second, environmental attributes are linked to soil salinity to discern similarities in spatial distribution and to correlate them.

3.2 Conceptual framework

The approach taken for this study assumes that local terrain serves as a simplified surrogate integrating the numerous landscape processes that influence the total amount of soil salinity. For the arid and semi-arid areas, Salama et al. (1999) related the spatial distribution of saline land and water to the hydro-geomorphology (e.g., topography and hydro-stratigraphy). The conceptual model of physical and chemical processes responsible for the development of saline soils developed by those authors is given in Figure 3.1. Jurinak (1990) also emphasizes the importance of water, because it serves as the principal vehicle for salts. Salinity is closely linked to lowlands or depressions where water drains and accumulates. Restricted drainage due to topography, which results in a high groundwater table, is a factor that usually contributes to salinization (Richards, 1954). Therefore, salinization is regulated by the evaporation and/or transpiration of water rather than by surface runoff and drainage (Jurinak, 1990). Moreover, it has been long recognized that topography indirectly influences soil properties and processes (Ruhe, 1956).

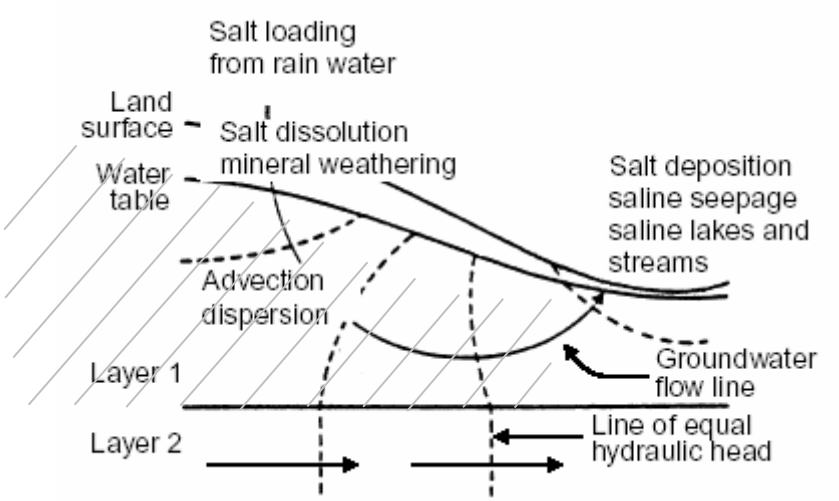

Figure 3.1 Conceptual model of soil and water salinization (Source: Salama et al., 1999)

The method for this study is based upon Moore et al. (1993) and has been demonstrated by Gessler et al. (1996) to be applicable for many comprehensive environmental models. A proper representation of landscape processes serving as environmental factors has long been under discussion (e.g., topographic indices). Additionally, other local environmental factors, such as the water network, soil texture, cropping pattern, groundwater level and salinity, are thought to further improve

estimates of soil salinity distribution. The paper aims to concisely describe parameters to be used in the model in relation to the soil salinity in the study area in Khorezm.

Many studies have attempted to investigate the functional relationship of certain soil properties with selected topographical parameters. The topographic index is the numerical representation of landscape features (i.e., slope, flatness etc.). It is assumed that areas possessing the same topographic indices behave similarly, regardless of their location in the landscape. While mainly developed for hydrological modeling, topographic indices are well suited for soil salinity studies. For example, recent studies by Florinsky et al. (2000) integrated the concept of accumulation, transition and dissipation zones together with digital terrain model to map small-scale salinity risk from the large scale maps.

Seelig and Richardson (1994) note that lateral transient water flow and evapotranspiration from the shallow water table redistribute and concentrate salts at distinct locations in all landform positions. The slope indices can be used to characterize this rate of the flow.

When considering large-scale topography, there is a differentiation between macro-, meso-, and micro-depressions (Kovda, 1946 cited in Florinsky et al., 2000). The author explains that salt accumulates in closed and partly drained macro-depressions, which are natural accumulators of substances moved by gravity from macro-crests and slopes. In meso-depressions, saline soils occur on meso-slopes, and in micro-depressions the build-up of salts typically occurs on micro-crests marked by low moisture and high evaporation. To account for these differences, curvature indices can be used. They indicate acceleration or deceleration of the surface water flow. Also, plan curvature determines if the flow is convergent or divergent. Combined occurrence of the plan and profile curvature would indicate a topographical depression.

The drainage of salt-bearing waters away from the higher lands of the basin may raise the groundwater level in the lower lying lands (Richards, 1954). This phenomenon can be represented by the concept of flow accumulation at certain points. The higher land is often termed the upslope contributing area.

The above mentioned topography indices are so called primary indices. As described in most text books, primary terrain parameters are those that can be derived directly from the DEM using local filter operation. In order to capture local terrain

features other indices are used. They are often the derivatives of the primary indices. For instance, wetness the index (considered a compound topographic index) is a ratio between catchment area and slope, which primarily reflects accumulation processes and could have implications for soil salinity development through the accumulation of water.

Observations of the locality show that salinization is sometimes a problem near water networks. Surrounding areas have higher moisture levels or the groundwater table is lifted, and therefore such areas have a higher evaporation potential. This could also be investigated by a stream power index, which bundles crossing streams into one and assigns a higher rating to the resulting stream.

Another factor of the local environment is land management. If an area is intensively used for agriculture, human interference via tilling and cropping of the soil is anticipated to have a strong influence. Various studies (Hendrickx et al., 1992; Norman, 1988) show that irrigation and soil management in flat irrigated lands determine the spatial variability of salinity to a much larger extent than prevalent soil characteristics. Thus it is critical to include land management in models.

Land management can be represented by many variables, so which variables should be used? Since the purpose of this study is to detect or predict soil salinity using readily available data or variables that are easy and cheap to measure only those cropping variables were selected that are easy to obtain. Three main crops in the Khorezm region can be distinguished by their different water requirements: Rice and cotton require high and moderate water input, respectively, while wheat is a less water-intensive crop.

Many authors associate intensity of soil organic matter (SOM) accumulation in arid environments with crop residue presence and moist conditions (Karavanova et al., 2001; Tursunov and Abdullaev, 1987; Wysocki et al., 2000). Both are found in places where groundwater is close to the surface, i.e., in hydromorphic soils such as in Khorezm. Hence, in this environment, the SOM content is important for salinity studies.

Primary and secondary topographic indices are not only important, they are also easy to calculate with modern computers and their use in soil salinity prediction is thus facilitated. Park and Vlek (2002) observed that results improved when more variables were used for prediction of selected soil attributes (i.e., total exchangeable

bases, soil pH). Therefore, the analysis in this study was first performed using all available indices. Later, those that did not contribute to the prediction accuracy after correlation analyses were removed.

As groundwater in the region is often characterized by a high salt content, evaporation leads to soil salinization due to precipitation of salts in the soil profile. A key factor in controlling the amount of evaporation is the depth of the water table below the surface. Typically, the contribution of groundwater to evaporation is minimal when the groundwater table is below 1.5-3.0 m (Salama et al., 1999). The concept of critical depth in arid regions was widespread in the former Soviet Union; however, it can only give a relative estimation of the degree of salinity. For example, Goossens et al. (1996) observed that if the depth of the groundwater table is less than the critical depth, almost any value of soil salinity can be observed.

Remotely sensed data have great potential for providing spatial estimates of soil salinity. Also, in an arid environment, differences in the soil salinity and hydromorphic level provide the most significant differences in their spectral reflectance and an increase in salt content led to an increase in the spectral reflectance of the saline soil (Karavanova, 2001). Remote sensing is a valuable source of ancillary information for soil prediction at the catchment and regional scale (McBratney et al., 2000).

Having said all the above, there is a constraint to keep in mind. The study area consists of flat alluvial plains with low potential energy for driving water flow (Abdullaev, 2002). Wysocki et al. (2000) notes that historically, generic terms (e.g., broad interstream divide, rise, flat, etc.) or terms developed for other settings have been applied to flat plains with unsatisfactory results. On the other hand, the flat topography enhances drainage effects on soil salinity.

Consequently, the 11 terrain parameters that are normally used in terrain modeling were calculated:

- Wetness index (wt)
- Upslope contributing area (ua)
- Aspect (as)
- Solar radiation (solar)
- Slope gradient (sl)
- Slope length-gradient factor (ls)
- Divergence and convergence (dc)
- Profile curvature (profc)
- Plan curvature (planc)
- Curvature (curv)
- Terrain component index (tci)

A detailed explanation of the parameters is given by Wilson and Gallant (2000), and some are briefly described here. The upslope area represents the amount of runoff that would flow through each cell in the digital elevation model (DEM) assuming homogeneous soil and rainfall conditions. The slope gradient affects overall rate of movement down the slope, while the slope aspect defines the direction of flow. Flow tends to accelerate when the profile curvature is positive and to decelerate when it is negative and therefore influences erosion and deposition. The plan curvature determines convergence and divergence of flow. For both, profile and plan curvature, a positive curvature indicates that the surface is convex, and vice versa. Similar to the assumption made by Florinsky et al. (2000), gravity-driven overland and intra-soil transport can be interpreted in terms of divergence or convergence and deceleration or acceleration of flows to account for salt transport to the area.

3.3 Materials and methods

3.3.1 Farm survey sampling

The survey was conducted on two farms: the research farm (41°36'N, 60°31'E) of the Urgench State University and the Pahlavon Mahmud private shareholder farm (shirkat), south-west of the city of Khiva (Figure 3.2). The farms were selected because their soils provided a sufficient range of values for several characteristics (e.g., texture, topography, crop type) that are reported to influence soil salinity. The sampling campaign was carried out from June to August 2002. Core sampling and electrical conductivity (EC) measurements were done systematically in a 150 m by 200 m rectangular grid. They were complemented by a two-way nested sampling with a finer 40 m by 40 m grid in order to study the effects of distance from water bodies and of micro-topography and to identify short-range variation in successive geostatistical analyses. Nested fields were randomly selected within the grid area totaling 6 nests.

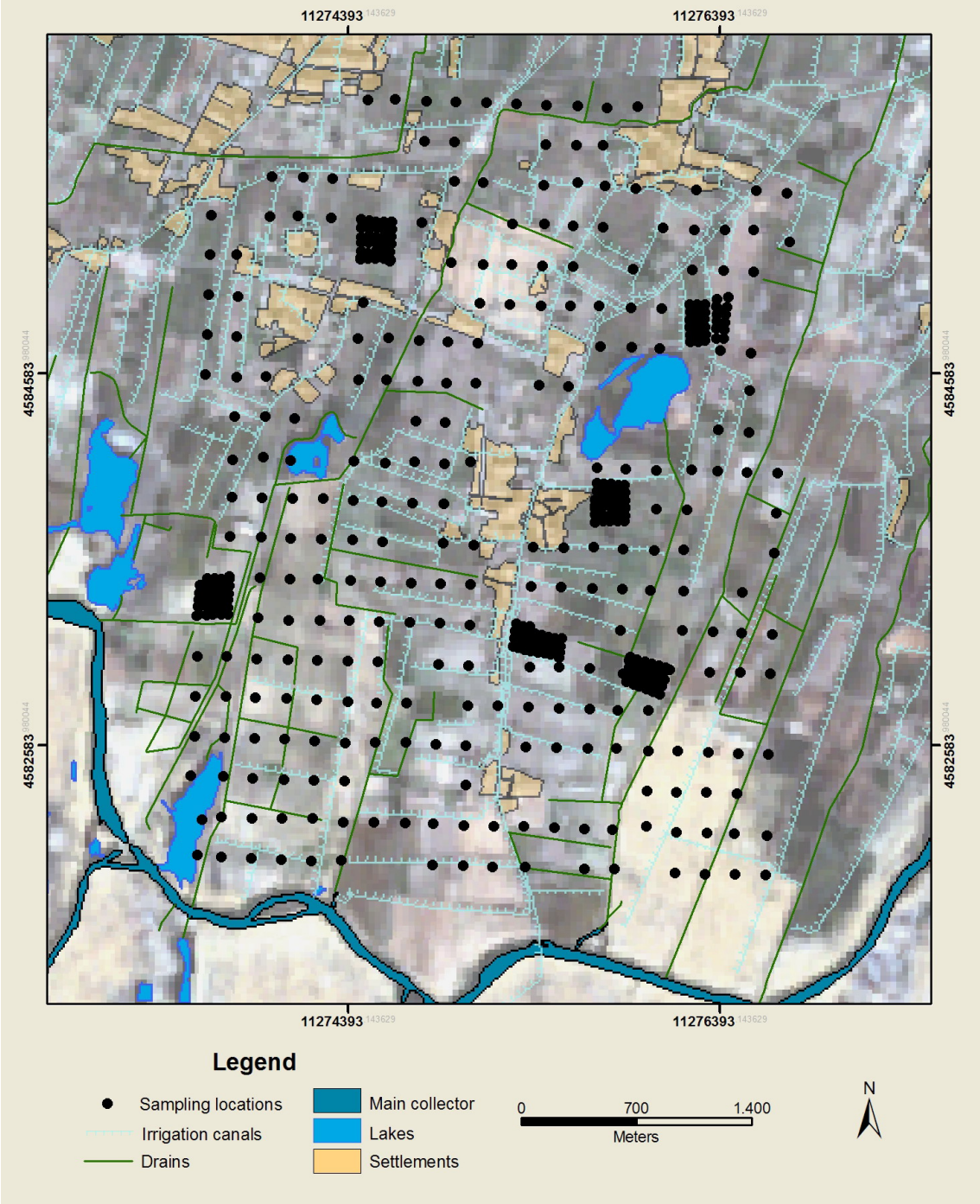

Figure 3.2 Farm scale sampling layout with irrigation and drain network, lakes, main collector, and settlements all overlaid on Landsat image (July 12, 2002)

The locations of the grid nodes were laid out prior to the sampling campaign and uploaded to a Global Positioning System (GPS) receiver (GPS 12, Garmin International Inc., USA). This design provided 580 grid nodes out of which 452 were sampled; the rest was excluded because of house settlements or unsuitable soil

conditions. At each grid node, soil core samples from 0-30 cm depth were taken in duplicate with a split tube sampler with an inner diameter of 53 mm. One sample was used for analysis of gravimetric water content and bulk density in the laboratory of UrSU. The second set of samples was analyzed by the Soil Research Institute (SRI, Tashkent) for TDS and ion analysis (see section 2.2.3).

As discussed in the previous section (2.3.2), the different classification methods for clay and the resulting minor differences in particle size distribution did not influence the calculation of EC. Unless mentioned otherwise, for the ease of transferability of the results, clay was assumed to consist of particles with a size of less than 0.001 mm (Russian classification scheme). TDS was taken as the variable that approximates soil salinity measured in the laboratory, as it correlated strongly with Na and Cl contents.

Data on groundwater table depth (GWT) and its salinity (GWS) were obtained from the observation wells installed in the area by the Hydrological Melioration Expedition of the Khorezm Department of Land and Water Resources. The extensive network of groundwater observation wells in the Khorezm region is described and analyzed in greater detail by Ibrakhimov (2004). The average value of groundwater table depth and salinity collected in July for the period from 1990 to 2002 for 45 groundwater observation wells were averaged and used for interpolation. It was assumed that long-term average data represent the spatial distribution of groundwater table depth and salinity better than single-year observations. Due to a very small number of wells in the areas surrounding the farms (16 wells) and to avoid edge an effect in interpolated maps, a larger number of wells (45 wells) covering a larger area were used. Also, to improve the interpolation and due to the low density of existing observation wells (45 per 90 km^2), data from additional wells were acquired. In July and August 2002, measurements were obtained for 8 observation wells in conservation agriculture trial fields (O. Egamberdiev, personal communication, 2002) and 6 from irrigation efficiency study fields (I. Forkutsa, personal communication, 2002).

The electrical conductivity of the bulk soil (EC_a) was measured using the CM-138 (described in section 2.2.1). Salt leaching is a common practice in agriculture in arid regions and aims at removing excess salts from the root zone. Topsoil is thus important for salinity appraisal. Various devices are available for measuring salinity and

mapping its distribution. However, recent studies in the region show that improper leaching coupled with frequent water shortage periods may increase the risk of salinity build-up from the salt in the subsoil. Therefore, while realizing that topsoil salinity assessment is important, care and attention should be given to salt accumulation in the subsoil, which may take some time to show up on the soil surface. The electromagnetic equipment is very useful in this regard, as it detects salinity not only in the topsoil but also to a depth of 1.5 m. The final data set included X and Y coordinates taken by GPS in the World Geodetic System 1984 (WGS84) geographical reference system.

3.3.2 Environmental factors

Elevation data from a 1:10,000 topographic map were obtained to create a Digital Elevation Model (DEM). Because a DEM is of primary importance to derive terrain indices, elevation data were obtained from points and contour lines covering an area of approximately 27 km^2. The accuracy of the generated DEM and other interpolated parameters is assessed by using cross validation and the root mean square standardized prediction error (RMSSE), which should be close to 1.

Terrain indices were calculated from a 30 m x 30 m raster-based DEM. A grid size of 30 m was selected, as it has proven to be the most suitable for soil-landscape analyses (Park et al., 2004). The following terrain indices were calculated using the software DiGeM 2.0 (Olaf Conrad, Gottingen, Germany) with methods given in references in parentheses: aspect (as), slope (sl), profile curvature (profc), plan curvature (planc), curvature (curv) (Zevenbergen and Thorne, 1987), divergence/convergence (dc) indices, flow direction (fd), flow accumulation (upslope contributing area, ua), wetness (wt), and erosivity (based on universal soil loss equation, ls) (Moore et al., 1993).

The factor ls is the slope length-gradient factor and represents a ratio of soil loss under given conditions to that at a site with the "standard" slope steepness of 9% and slope length of 22.1 m (Wilson and Gallant, 2000). The steeper and longer the slope, the higher is the risk of erosion.

Additionally, curvature (curv7) and the terrain characterization indices (tci) were calculated according to Park et al. (2001):

$$tci = curv7 * log_{10}(ua) \quad [17]$$

An agricultural map with scale 1:10,000 was used to obtain the water network infrastructure that consisted of irrigation and drainage canals within the sampling area. The layers were digitized and the nearest distance from sampling point to each layer was obtained using ArcView 3.2 (ESRI Inc., USA).

Remote sensing parameters were obtained from a Landsat 7 satellite image acquired on the July 12, 2002. A subset of the image was analyzed in Erdas Imagine 8.2 (Leica Geosystems GIS & Mapping, USA) to calculate vegetation- and soil-related indices. Vegetation indices were analyzed carefully, as Hick and Russell (1990) showed that degraded vegetation cover did not necessarily indicate saline areas, because volunteer (i.e., halophyte) vegetation could develop well despite increasing salinity. Although maximum discrimination is usually achieved during maximum vegetation growth, an attempt was made to correlate salinity data with the normalized difference vegetative indices (NDVI=(band4-band3)/(band4+band3)) as well as the transformed normalized difference vegetative indices (TNDVI = sqrt ((band4 - band3) /(band4 + band3) + 0.5)). Bands used for the computation of these indices are widely accepted in salinity studies (Hick and Russell, 1990).

Additional indices calculated from these bands are: soil-adjusted vegetation index (SAVI=((band4-band3)/(band4+band3+0.5))*(1+0.5)) and ratio vegetation index (RVI=band4/band3) known to delineate reduced reflectance due to salinity (Wang et al., 2002), and soil band ratio (RS57=band5/band7), which is sensitive to clay minerals and is the combination offering the maximum likelihood to identify saline soils (Goossens et al., 1998; Shepherd and Walsh, 2001). Altogether, remote sensing provided 12 variables (including raw band readings; bands 1-5, 7, and 8) that were included into the analysis.

3.3.3 Geostatistical analysis

The geostatistical analyses were performed in ArcMap 8.3 (ESRI Inc., USA) using the Geostatistical Analyst 8.3 extension. Kriging interpolation assumes that the measured property satisfies the condition of second order stationarity, and this, in fact, is regarded as a major limitation of the technique (McBratney et al., 2000; Yates and Warrick, 2002). Second order stationarity implies that the mean, variance, and semivariogram are

not dependent upon sampling location (Mulla, 1988). Apart from trends in the data set, local directional effects could also occur, which means that in certain directions similarity may be greater than in other directions. This is called anisotropy, and although most variables exhibited some signs of it, the decision whether to include anisotropy into the model was made based on the accuracy of the generated maps.

The data set was tested for normal distribution and the evidence of trends; if trends were detected, it was attempted to model the trend with a second-order polynomial function. Interpolation of EC_a measured by the CM-138 was done using raw data, as Triantafilis et al. (2001) reasoned that fewer errors were incorporated into the kriging system than when EC_a data were converted into EC_e estimates.

Spatial variability of elevation, TDS, Cl, CMv, CMh, clay, humus, groundwater table and salinity were evaluated with the aid of semivariograms, which were constructed according to standard geostatistical methods (Warrick et al., 1986). Variograms explain how strongly a variable value at one spot resembles a value at another as a function of the distance between the two spots. Semivariograms in all cases are fitted with a spherical model unless otherwise mentioned. Ordinary kriging was applied as an interpolation method, as it minimized the influence of outliers on prediction performance as has been widely observed (Odeh et al., 1994; Saito and Goovaerts, 2000; Triantafilis et al., 2001). A default number of neighbors (5 or at least 2 for each angular sector) was used for kriging.

3.3.4 Data analysis and visualization

The calculated topographic indices maps were imported into ArcView 3.2 (ESRI Inc., USA) where their values were extracted and added as a new variable into the table for each sampling point. Similarly, distances to the water bodies were extracted from the grid maps generated in ArcView 3.2. Further correlation analyses were performed in SPSS 11.0 (SPPS Inc., USA).

Soil salinity in agricultural land may vary not only with location but also with land-cover type or soil type. While salinity change solely due to location can be assessed by geostatistical analysis, the contribution of the land-cover type and soil type can be distinguished by analysis of variance (ANOVA). ANOVA, which is mainly used for experimental hypothesis testing, can also provide a way to assess the sources of

variation and is widely used in landscape analysis (Burt and Park, 1999; Park and Burt, 2002; Yemefack et al., 2005). The total variance of soil salinity may be subdivided into three components, where two components are explained by land and soil type and the remainder represents the unaccounted error.

Since the number of factors calculated or obtained from the DEM and remote sensing can be large, most significant variables have to be identified. There could be some within each group that correlate well with others. For this purpose, an ordination technique implemented by Park and Burt (2002) was used to identify underlying mechanisms for complex terrain variables and indices calculated from remote sensing data. A principal components analysis (PCA) was also carried out. It basically involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components.

Because factors are obtained from two different data sources, i.e., terrain and remote sensing, factors coming from each of these sources are subjected to principal components analysis. The number of components to be extracted was decided based on eigenvalues and scree plot interpretation. The eigenvalues obtained from the PCA are equal to the variance explained by each of the principal components in decreasing order of importance. It follows that any component with an eigenvalue of at least 1 explains more of the variance than any original variable. A rule-of-thumb is, therefore, to select that number of principal components having an eigenvalue of at least 1. The direct oblimin rotation was used, because independence among extracted principal components could not be assumed. More details of PCA are given in Park and Burt (2002). The variables were not standardized, because the PCA results were similar with both transformed and standardized values.

Also, a description of the spatial relationship of soil salinity and other parameters was attempted via scientific visualization. 3D maps were created with soil salinity data represented as the height and environmental parameters as color.

3.4 Results

3.4.1 DEM construction

A DEM for the study area of 27 km^2 (approximately 5.2 km by 5.2 km) was constructed from very low vacillating elevation values indicated by the low range between

minimum and maximum values (Table 3.1). Trend analysis shows that in this area topography exhibited a strong trend in the east-west and north-south directions (Figure 3.3). Most of the higher elevation points are in the north-east (Khiva city area) and the lower elevation was down in a south-west direction. For the visual comprehension of the area the Landsat image was draped over the DEM landscape Figure 3.4). The red, green, and blue bands are represented by 3, 2 and 1, respectively. This is the natural color composition and is primarily used for display purposes. It also delineates shallow waters well. It can be seen that the area is mainly subdivided into agricultural fields (areas with straight boundaries) with or without crops (on July 12, 2002). The extent of the area devoted into cultivation is high (approximately 60-70%).

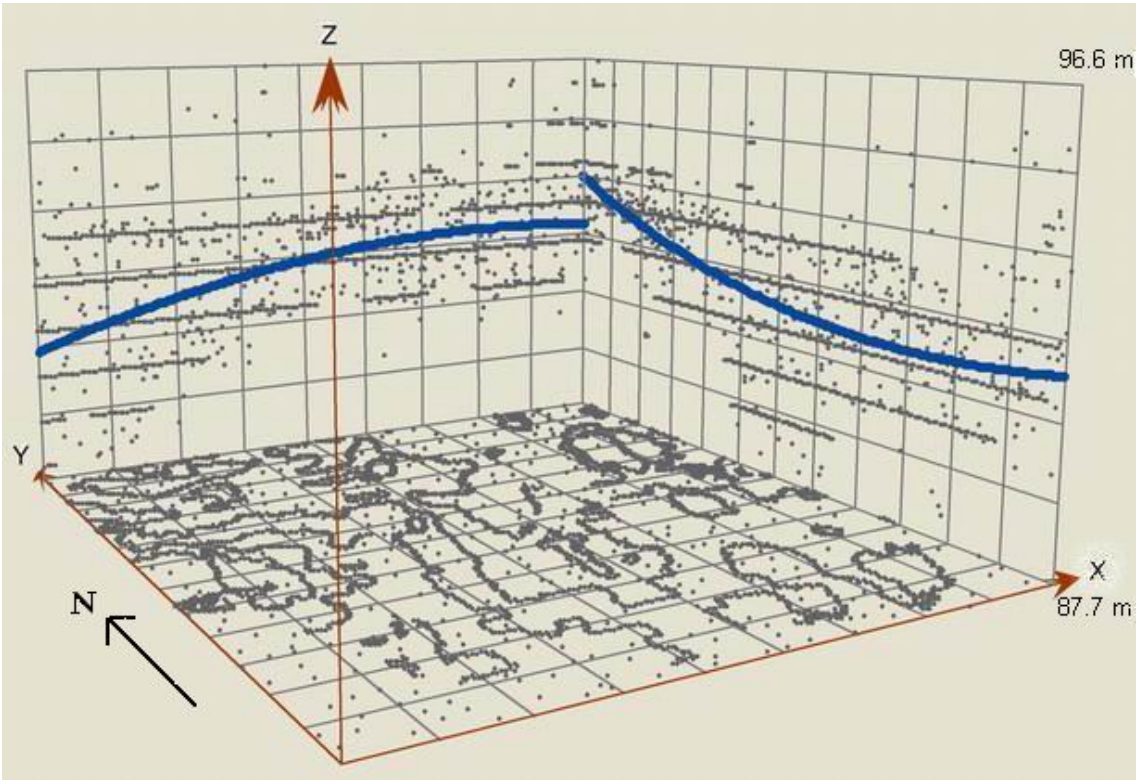

Figure 3.3 Trend analyses of elevation data points at farm scale

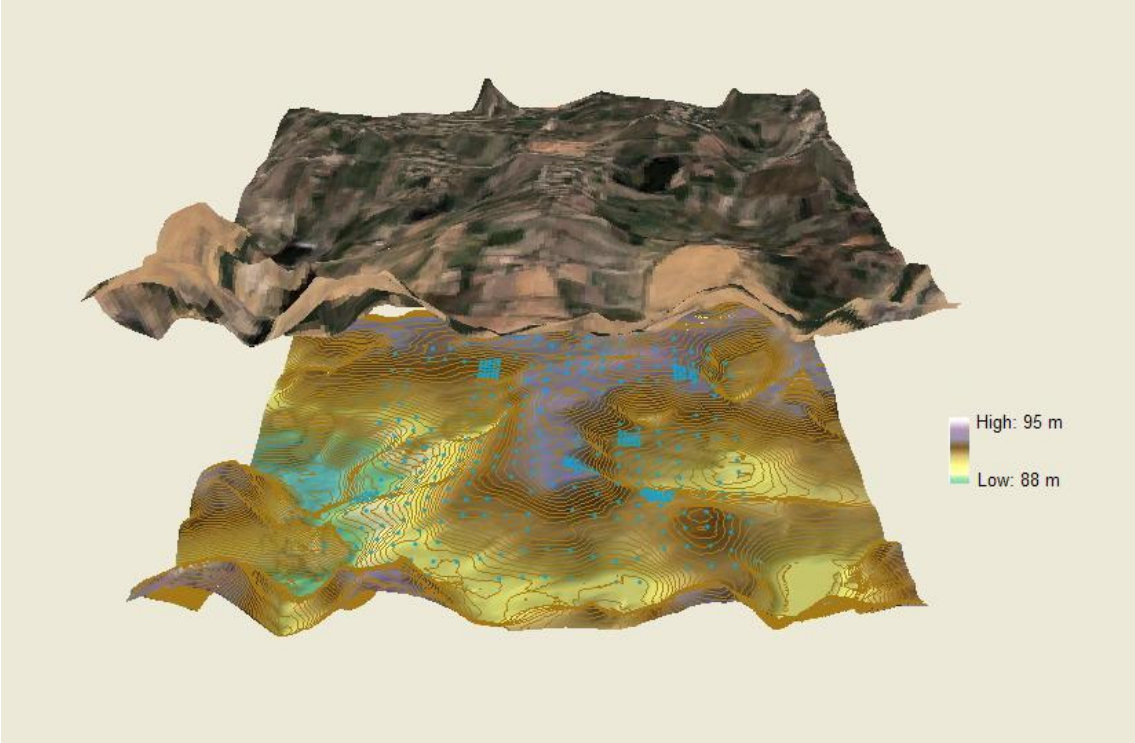

Figure 3.4 Landsat image draped over DEM of the study farm

Based on the summarized data in Table 3.2, the kriging quality of the interpolation is considered good, indicated by a mean standardized error (MSE) close to 0. The RMSSE for elevation data in Table 3.2 is 0.74, which shows that variance was overestimated (RMSSE<1). Perhaps overestimation of the variance is caused by the interspersed hummocks in the southern and north-eastern parts of the area. The sampling area has a reasonably gradual change in elevation, and it can be assumed that the calculated indices will not be affected by the overestimation of the variability of the predicted values.

Table 3.1 Summary statistics of variables used for interpolation

	Units	Mean	Median	Std. Dev.	CV	Minimum	Maximum	Skewness	Log Skewness	N
Elevation	m	92	92	1.3	1	88	97	-0.40	-0.44	4122
CMv (CM-138 vertical)	dS m^{-1}	0.50	0.50	0.33	66	0.01	1.92	0.66	-0.94	445
CMh (CM-138 horizontal)	dS m^{-1}	0.72	0.69	0.22	31	0.35	1.66	1.21	0.16	222
TDS (total dissolved solids)	g $100g^{-1}$	0.32	0.22	0.35	109	0.06	3.46	4.68	0.58	448
CL (chloride)	g $100g^{-1}$	0.04	0.03	0.06	150	0	0.73	6.27	0.49	448
Humus	g $100g^{-1}$ C	0.67	0.70	0.31	46	0.04	1.97	0.29	-1.18	448
Clay	%	8	10	5	60	0	22	-0.12	-1.73	448
GWT (groundwater table depth)	cm	121	118	31	25	81	262	2.25	1.01	57
GWS (groundwater salinity)	dS m^{-1}	2.9	2.4	1.3	45	1.3	6.7	1.62	0.87	59

Table 3.2 Semivariogram parameters and cross validation results for interpolated variables

Variable	Units	Major range (m)	Anisotropy	Nugget effect	Partial sill	Lag size	Number of lags	Mean error	RMSE	Average standard error	Mean standardized Error (MSE)	RMSSE	N
Elevation*	m	1778	282	0.00	0.00	150	12	-0.0006	0.36	0.44	0.0006	0.74	4122
TDS*	g $100g^{-1}$	571	358	0.24	0.18	110	8	-0.0092	0.22	0.21	-0.0580	1.22	442
Cl*	g $100g^{-1}$	1501	4	0.42	0.16	170	11	-0.0014	0.03	0.03	-0.0636	1.04	442
CMv	dS m^{-1}	448	-	0.00	0.05	80	10	0.0027	0.18	0.15	0.0099	1.22	445
CMh*	dS m^{-1}	563	-	0.01	0.05	60	13	-0.0022	0.17	0.13	-0.0239	1.23	222
Clay	%	1214	-	0.76	12.32	150	11	-0.0002	2.41	1.60	0.0022	1.40	441
Humus	g $100g^{-1}$	1500	-	0.02	0.04	150	11	-0.0008	0.19	0.18	0.0002	1.07	448
GWT*	cm	1004	89	0.00	0.03	110	10	-0.0956	20.48	20.02	-0.0198	1.02	57
GWS*	dS m^{-1}	1541	279	0.05	0.14	130	12	0.1432	0.88	1.20	0.0742	0.81	59

* *interpolation was done with log_{10} transformed values*

PCA of terrain attributes and remote sensing data

As mentioned earlier, 12 terrain indices (including elevation in meters) were calculated from a constructed DEM and 12 variables from a remote sensing image. The principal components analysis (PCA) extracted six principal components out of these 24 variables that had eigenvalues greater than one and explained 85 % of the total variance. Attributes that were not normally distributed were log transformed before analysis. The first two components, PC 1 and PC 2, together explained half (38 % and 17 %) of the total variance with the rest contributing 11, 8, 6, and 5 %. The attributes were sorted and regrouped after rotation based on the factor loadings of each component (Table 3.3).

A clear difference can be seen between the principal components, because those factors originating from remote sensing are grouped separately with highest loadings in PC 1 and PC 3. Interestingly, their oblique rotation loadings, which refer to regression coefficients (not to be confused with correlation coefficient[1]), are greater than 1. Joreskog (1999) warns that this might suggest that there is a high degree of multicollinearity in the data. All the indices calculated from the band combinations appeared in PC 1, with some multicollinearity. In PC 3, multicollinearity seems to be less expressed with Band 4 having the highest value.

The principal components 2, 4, 5, and 6 are all combinations of terrain indices. Component 2 consists of variables that represent curvatures and TCI, which is also derived from curvature. These curvatures describe the landscape form and were grouped by the PCA into one component. Similarly, variables ls and slope are grouped together in component 4, which indicates the process they represent is distinct from others. They basically represent erosion potential based on the ls factor, which is in turn calculated from the slope variable.

Thus, the extracted components can be roughly divided into (i) curvature or shape group (PC 2), (ii) erosivity group (PC 4), which contains the erosivity factor and

[1] The misunderstanding probably stems from classical exploratory factor analysis where factor loadings are correlations if a correlation matrix is analyzed and the factors are standardized and *uncorrelated* (*orthogonal*). However, if the factors are *correlated* (*oblique*), the factor loadings are *regression coefficients* and not *correlations* and as such they can be larger than one in magnitude. Just remember that a standardized coefficient of 1.04, 1.40, or even 2.80 does not necessarily imply that something is wrong, although, it might suggest that there is a high degree of multicollinearity in the data. Joreskog, K.G. 1999. How large can a standardized coefficient be? [Online]. Available by Scientific Software International http://www.ssicentral.com/lisrel/column2.htm (verified December 21, 2004).

slope gradient used to calculate it, (iii) radiation group (PC 5), and (iv) water group (PC 6), which was mainly represented by upslope and wetness indices. Correlation of these extracted principal components with salinity will be presented and discussed later during the description of the controlling factors of soil salinity (section 3.4.3).

The grouping of the relevant variables into the same components (e.g., plan and profile curvatures, curvature, and TCI in principal component 2) confirms that the processes or attributes they represent are reflected in similar groupings. Further inference can be made since 1) the extracted components represent similar processes, and 2) 6 components accounted for 85% of the variance of the dataset. Based on this, a simple model can be built, which takes into account either components scores calculated during PCA or one variable per component that has maximum loading and hence can be regarded as representative of that component.

Table 3.3 Results of factor correlation matrix after direct oblimin rotation of principal components

Group	Variable	Mean	Std. Deviation	Minimum	Maximum	Skewness	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	Communality
PC 1 Remote sensing (RS) group	SAVI	0.13	0.18	-0.23	0.64	0.46	-1.02	0.00	0.15	-0.03	0.00	-0.01	0.99
	NDVI	0.08	0.12	-0.15	0.43	0.46	-1.02	0.00	0.15	-0.03	0.00	-0.01	0.99
	RATIO43*	1.23	0.33	0.73	2.48	1.08	-1.02	0.01	0.15	-0.03	0.00	-0.01	0.99
	TNDVI*	0.76	0.08	0.59	0.96	0.27	-1.02	0.00	0.15	-0.03	0.01	0.00	0.98
	CLAYRS*	1.36	0.18	1.04	2	0.63	-0.91	0.02	-0.03	0.02	-0.02	0.03	0.84
	BAND3*	76.75	18.80	36	136	0.45	0.80	0.00	0.41	0.01	0.00	0.01	0.99
	BAND1	70.10	7.71	49	88	0.00	0.77	0.00	0.42	0.02	-0.03	0.00	0.94
	BAND2*	67.62	10.40	40	98	0.23	0.73	0.01	0.50	0.01	0.01	0.01	0.97
	BAND7*	59.93	21.02	19	133	0.67	0.68	-0.01	0.54	-0.04	0.01	-0.01	0.95
PC 2 Shape group	PROFC	-0.0000007	0.0000268	-0.0001	0.00014	0.66	0.00	0.94	0.01	0.00	-0.03	0.16	0.81
	PLANC	0.0000003	0.0000239	-0.00008	0.0002	2.37	-0.08	0.76	-0.04	0.03	-0.01	0.02	0.58
	CURV7	0.00	0.01	-0.02	0.02	-0.25	0.06	0.67	0.00	-0.02	0.01	-0.40	0.79
	TCI	0.00	0.02	-0.12	0.07	-0.99	0.07	0.65	0.02	0.00	0.00	-0.41	0.77
PC 3 RS group	BAND4	89.30	11.41	36	124	-0.46	-0.44	0.01	1.01	-0.01	0.02	0.00	0.98
	BAND8	83.44	11.20	34	117	-0.05	0.24	-0.01	0.88	0.00	0.02	0.00	0.95
	BAND5*	78.60	20.29	26	142	0.52	0.47	0.00	0.72	-0.04	0.00	0.00	0.93
PC 4 Erosivity	LS*	0.03	0.03	0	0.25	2.25	0.01	0.00	0.07	0.95	0.10	0.18	0.91
	SL_RAD	0.0017168	0.0013300	0.000011	0.0094	1.21	0.08	0.03	-0.03	0.93	0.00	-0.03	0.89
PC 5 Radiation	AS_RAD	3.24	1.67	0.04	6.28	-0.04	0.04	0.06	0.02	0.08	0.84	0.00	0.71
	SOLAR	7.14	0.00	7.12	7.15	-0.49	-0.11	-0.12	-0.08	0.08	0.76	-0.13	0.61
	ALT*	91.99	0.97	89.85	93.85	-0.52	-0.13	-0.04	-0.12	0.23	-0.47	-0.29	0.46
PC 6 Water group	UA*	3.66	0.62	2.46	6.72	0.82	-0.01	-0.02	-0.01	0.12	0.07	0.90	0.84
	DC	-0.33	12.51	-55.31	64.08	0.15	-0.01	0.08	0.01	-0.13	0.07	-0.80	0.67
	WT*	12.45	1.95	8.9216	21.7597	1.40	-0.02	-0.05	-0.03	-0.48	0.02	0.75	0.90
Eigenvalue							9.03	4.03	2.76	2.02	1.41	1.19	
Variance explained							37.63	16.79	11.49	8.41	5.87	4.98	
Cumulative percentage							37.63	54.43	65.92	74.33	80.20	85.17	

** variables were log_{10} transformed for PC analysis*

3.4.2 Characterization of spatial distribution

Salinity measured by TDS, chloride and CM-138

The statistical summary of the data used for interpolation is given in Table 3.1. The average TDS value indicates that the soils are either not or slightly saline (Figure 3.5a) according to classification described in Kaurichev (1989). The salinity type is mainly chloride-sulphate (Figure 3.5b), which is typical for soils in this region (Figure 3.5c) (Popov et al., 1992; Tursunov and Abdullaev, 1987). As shown in section 2.3.3, chloride was the main ion responsible for EC; therefore it was also included in the analysis of spatial distribution here. TDS and Cl data were not normally distributed as can be seen in the large range between minimum and maximum levels, a large difference between median and mean, and a high coefficient of variation (CV). This phenomenon is normal for salinity data, and Wu et al. (2002) reported that it can be caused by a number of very high values. This was also true for the data in this study, as a logarithmic transformation (see Log Skewness column in Table 3.1) resulted in a normal distribution of those variables that had shown a skewness greater than 1.

On the other hand, salinity as measured by the CM-138 exhibited a lower coefficient of variation. This is very likely due to the larger volume of soil measured by the device, which takes readings in the bulk soil down to 1.5 m depth, whereas TDS and Cl were measured in a smaller soil sample taken from the top 30 cm layer only. However, there is a 2-fold difference in the CV between CM-138 measurements taken in the vertical (CMv) and those taken in the horizontal mode (CMh). It should be noted that the CMh data refer to readings to a depth of 0.75 m taken only in the northern part of the sampling area. When the CV of CMv readings was calculated only considering the data of this area (n = 222), the variation was similar to that of the CMh readings. This implies that vertical mode readings are not necessarily more variable than those taken in the horizontal mode.

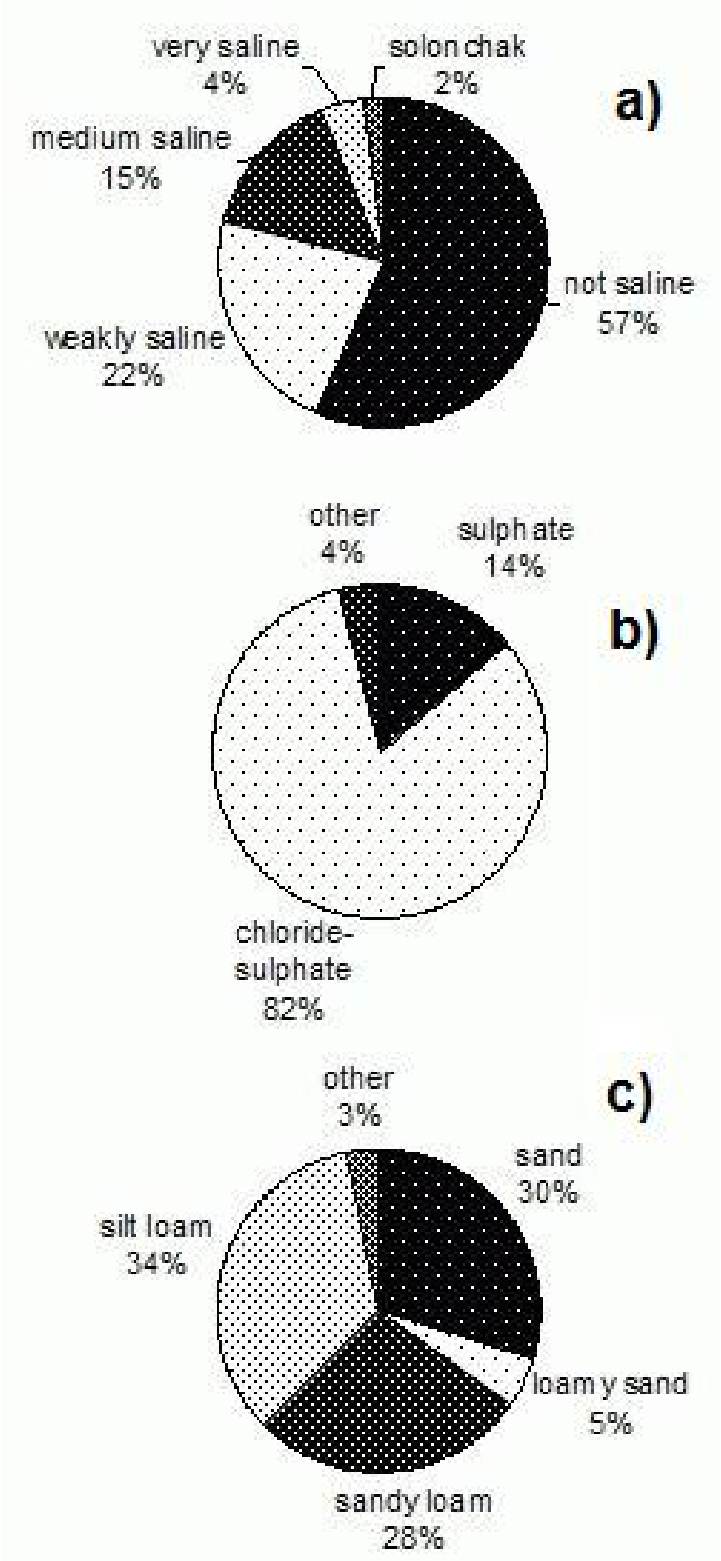

Figure 3.5 Proportions of the samples by
a) degree of salinity
b) type of salinity
c) soil texture (clay<0.002 mm)

A trend analysis for each variable revealed trends in two directions. The strongest trend ran in a north-south direction, while the weaker trend went from east to west, roughly resembling the trend of the elevation points (Figure 3.3). Since it did not satisfy the assumption of stationarity, further analysis was done with the trend removed. Trend removal is a precondition for carrying out interpolation by kriging. The semivariogram will then model the spatial autocorrelation among data points without having to consider the trend in data. The trend is automatically added back to the calculations before the final surface is produced (Johnston et al., 2001).

The study of the semivariograms, which show the squared differences of the values between each pair of points at different distances (Johnston et al., 2001), reveals that there is a structure in the spatial distribution of the data points. The semivariograms of TDS, Cl, CMv, and CMh are presented in Figure 3.6. Values of semivariance were fitted to the spherical model, which agreed well with the empirical semivariance. The spherical model for the CM-138 was fitted without directional anisotropy, whereas anisotropy was included for TDS and Cl data. The summary of the variogram parameters that were used to construct the empirical semivariogram and the corresponding spherical model is given in Table 3.2.

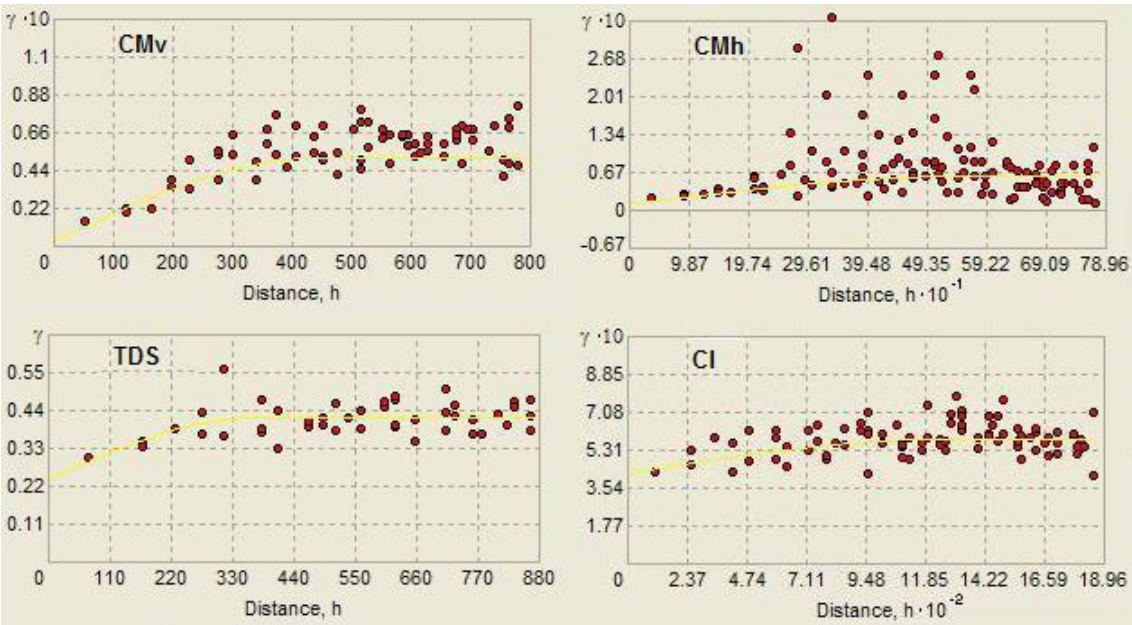

Figure 3.6 Semivariogram of CM-138 vertical mode reading measured at the soil surface

The major range, i.e., the lateral difference within which the soil salinity is autocorrelated, varied from 448 to 1501 m. The spatial pattern is rather similar when based on CM-138 and TDS measurements, however, the range for Cl is much larger. That means that the scale of spatial correlation for Cl varies over 1501 m, and that, since its distribution exhibited anisotropy, the direction (which is given in degrees clockwise from the north) varies from north to south. Similar anisotropy is exhibited by TDS.

The variograms display that TDS and Cl had a considerable nugget effect. Theoretically, as the distance between samples decreases, the variance should also decrease, converging at 0 variance at 0 m between samples. However, TDS and chloride appear to have variability even at short distances. The nugget effect is a measurement error and/or microscale variation (variation at spatial scales too fine to detect) (Johnston et al., 2001). Although the regular grid design was complemented by a nested sampling with denser grid sizes to account for this type of small-scale variations, apparently the grid size with a 40 m spacing between the sample points was not fine enough. To check if the nested sampling had any effect, an empirical variogram was built without the nested grid samples. This increased the nugget variance and sill considerably for both TDS and Cl, from 0.24 to 0.42 and from 0.42 to 0.52, respectively. This indicated that the spatial variability is still significant even within the 40 m scale and even finer grid spacing to account for this effect would be required.

The interpolated maps for CMv, CMh, TDS, and Cl are given in Figure 3.7. Kriging errors presented in Table 3.2 indicate that Cl was well interpolated with a RMSSE close to 1. However, the mean standardized errors for TDS and Cl are higher than those for CM-138 data, which is likely to be due to the nugget effect. The legend values in the TDS and Cl maps display that extreme values were considerably smoothened by the kriging interpolation. The generated Cl map has a gradually changing structure. In contrast, the TDS, CMv, and CMh maps all have RMSSE values greater than 1, which indicates that the variability of predicted values is underestimated.

The most salt-affected areas were found in the north of the study region. Salinity measured by CMv shows two distinct areas of high and low readings. Several hot spots probably make the study map look less smooth compared to the Cl map. The main reason for the salinity distribution patterns on this farm can be hypothesized to be elevation or soil type, since topographic depressions have low CMv readings, whereas

high readings were found in areas with where heavier. These differences may be related to a changing topography or lithology or other factors. A description and correlation of all considered factors are discussed in the later sections.

Evident spatial distribution patterns were singled out as potential causes for the observed topsoil salinity. In general, the TDS and Cl contents (Figure 3.7) virtually reciprocated the elevation contours (Figure 3.4) with only one mismatch in the southern part of the study area. Also, higher elevations generally showed a higher content of TDS and Cl, but in the southern part the salt content was elevated although the topography was depressed. Closer examination of that particular area (Figure 3.8) showed that the mismatch was probably caused by a clayey soil texture and together with the low elevation. Salinity peaks here might be associated with high evaporation rates and the accumulation of the salt in the uppermost topsoil, since CMv readings did not pick this salinity area up noticeably.

An important aspect in characterizing the spatial distribution of soil salinity was to understand the spatial structure. Summarizing, topsoil salinity (TDS and Cl) content showed a high variability both in data sets (indicated by coefficient of variation) and at short distance range (nugget effect revealed by variograms). On the other hand, CM-138 measurements exhibited small variability in the data sets and no short distance range variability. Therefore, the use of CM-138 measurements to correlate with environmental factors offers a more stable structure.

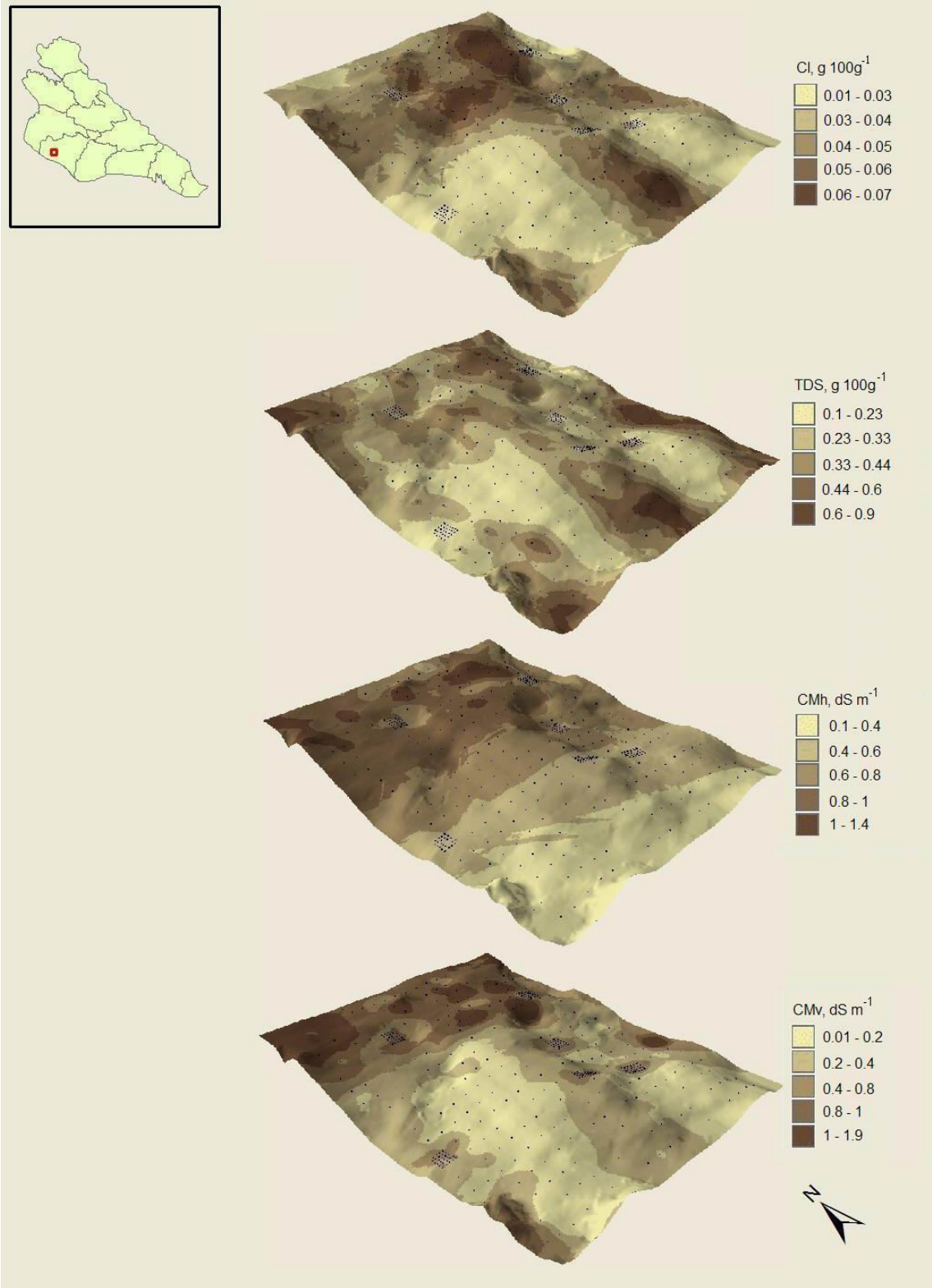

Figure 3.7 Interpolated maps of CM-138 in vertical (CMv) and horizontal (CMh) positions, total dissolved solids (TDS) and chloride (Cl) content draped over elevation of the study area

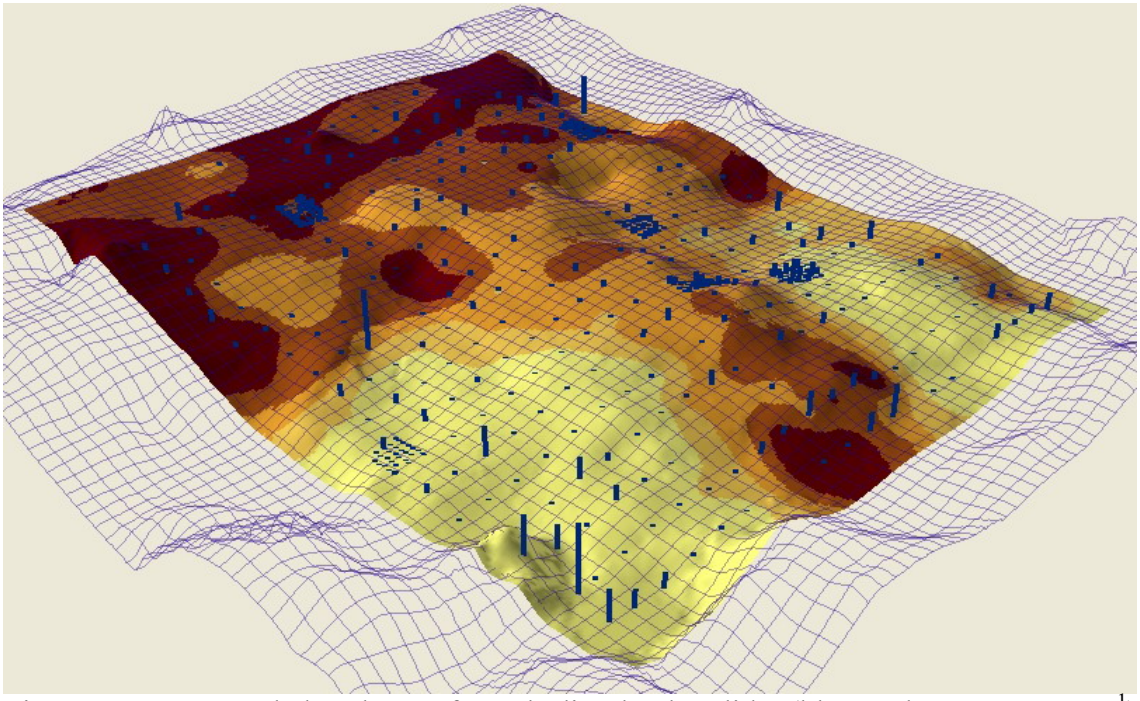

Figure 3.8 Extruded values of total dissolved solids (blue columns; g $100g^{-1}$) differentiated by clay content (darker areas = more clay; %) both overlaid on elevation. Extreme values of TDS (extruded columns) coincide with local depressions within both lighter areas (sandy) and darker areas (clayey)

Spatial distribution of measured soil parameters

The statistical summary of measured properties used as ancillary data such as clay, humus content, groundwater depth and salinity is included in Table 3.1. The interpolated maps are shown in Figure 3.9. The CVs for these variables show a moderate to high magnitude of variability, with clay and humus contents above the typical ranges in published soil studies as summarized by Mulla and McBratney (2000). A relatively high clay content is typical for the area, where soils belong to the so-called meadow group of soils (according to the Russian soil nomenclature and described in detail in Tursunov and Abdullaev (1987)). The prevailing low clay content is explained by the interspersed sandy desert soils. Figure 3.5c presents the proportion of samples by soil texture according to the USDA classification. Soils are roughly divided between loamy and sandy texture, with silt and sandy loam texture comprising two thirds, the rest being sandy.

The average humus content is low. This is typical for only recently developed irrigated meadow soils (Tursunov and Abdullaev, 1987), which initially tend to

decrease in humus content. In the long run, irrigated cultivation generally leads to a build up to the original humus level, and centuries-old farming land can even have a higher humus content than the zonal soils in the area.

Both clay and humus content show low skewness. However the histogram of clay content shows a bimodal distribution (Chapter 2), because the soils in the survey area consist of loamy and sandy soils. Groundwater table depth and salinity were not normally distributed; however, their natural logarithmic transformation reduced the skewness considerably. This shows that groundwater table depth and salinity are erratic in distribution with locally extreme values compared to clay and humus content, which are quite uniformly distributed over the landscape.

The trend analyses revealed similar patterns for clay and humus content, and for groundwater table depth. They gradually decreased from the north to the south. The trend in the east-to-west direction was not steady and clear, but when both directions were combined, there was a clear trend running from the north-east to the south-west. The same directional trend was displayed by groundwater salinity, with the only difference that it gradually increased along the north-east to south-west gradient.

Similar trends for all measured soil properties and the elevation indicate that one process is responsible for these phenomena. In this case it is most likely that the Amu Darya river is the single most important factor, as it formed the existing relief and influenced the soil texture. Due to the heavier soils in the northern part of the study area, soil salinity is higher. The southern part is characterized by low soil salinity, as the prevailing sandy soils in this region basically do not show a marked cation adsorption capacity as is the case for clayey soils and also have a limited ability for capillary rise. As a consequence, however, groundwater salinity is high. There are two more possible reasons for this: the sandy soils in this region have a higher transmissivity and lower relief compared to the northern part of the study site, triggering the accumulation of salts coming from surrounding areas.

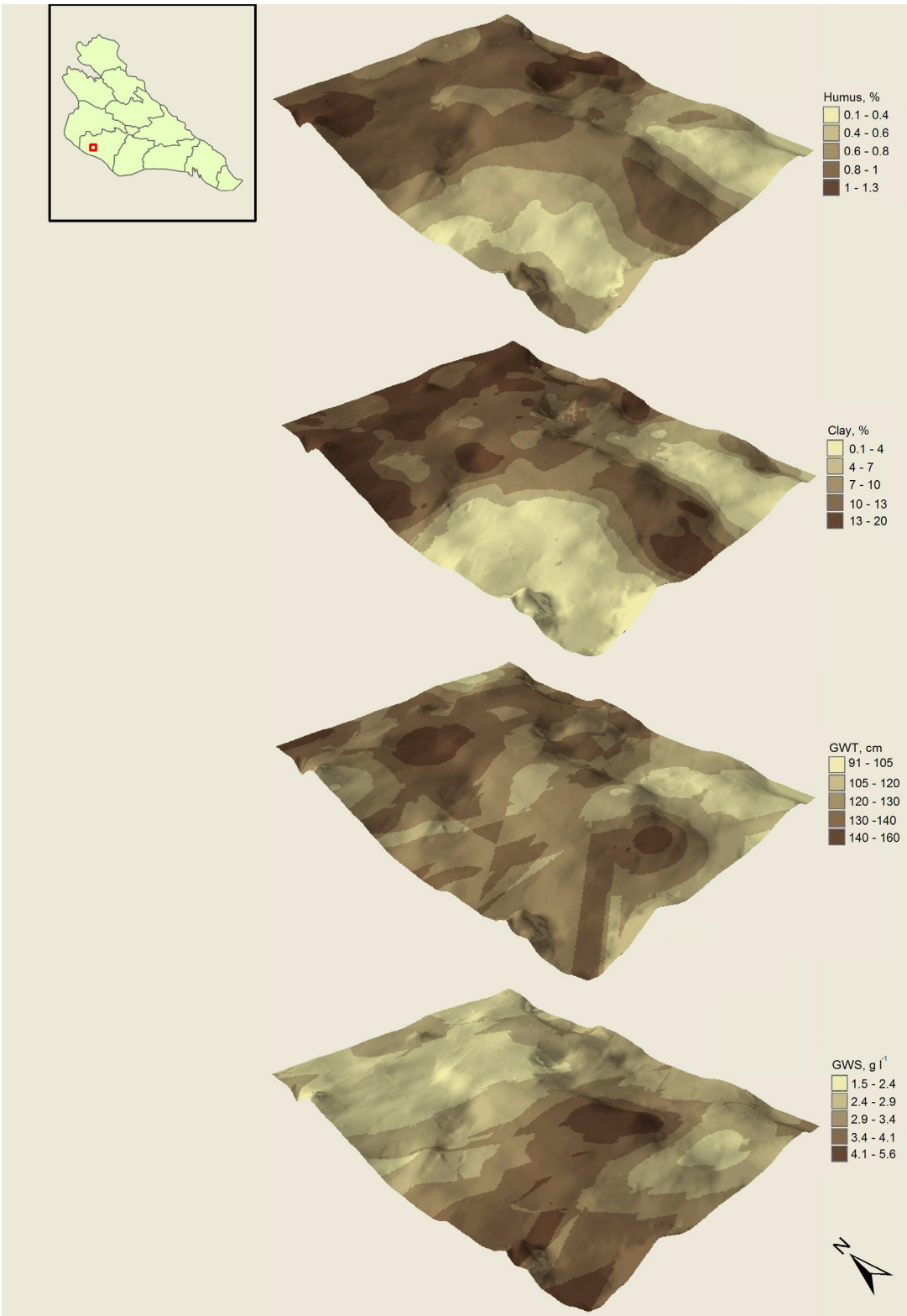

Figure 3.9 Interpolated 3D maps of humus, clay, groundwater table depth (GWT) and salinity (GWS) draped over elevation of the study area (Note: straight lines in the groundwater depth and salinity are caused by lack of data for the interpolation)

3.4.3 Environmental factors controlling the variability of soil salinity

Effect of land cover and soil texture

A two-way ANOVA (dependent variables: TDS, CMv, and CMh; independent variables: land cover and soil type) showed highly significant interaction between the two independent variables (Table 3.4) meaning that for CMv and TDS, variation by land cover differed among soil types. However, the interaction with CMh was not significant. This was, most likely caused by the low number of recordings (n=222), because when the number of cases for TDS and CMv was reduced to match the number of readings obtained by CMh, interaction for TDS and CMv was also no longer significant. The overall effect of interaction of land cover and soil type is somewhat inconsistent and equivocal.

Inconsistency probably stems from (i) the reduced number of samples used in the ANOVA, and that (ii) samples were collected from an area with uniform land cover. The only significant variable remaining (with both n > 440 and when n = 222) in both cases was soil type, which also had a higher F value compared to the land cover. As the significance level for the soil type is consistently highest, it can be concluded that soil salinity is different among soil types.

The goodness-of-fit (R^2) of the two-way ANOVA model is about two times lower for TDS than for the salinity measured by the CM-138. Because TDS is collected from the top 30 cm soil, its low R^2 value suggests that it might be less sensitive to land cover and soil type or that the management effect is more pronounced on surface soils.

Table 3.4 Results of two-way ANOVA of effect of land cover and soil type on soil salinity

Variable	Land cover (F value)	Soil texture (F value)	Interaction	R^2	N
TDS§	3.93**	6.93**	2.05**	0.25	448
TDS§	0.98	3.68**	1.02	0.27	221
CMv	1.94	11.22**	1.84*	0.45	445
CMv	2.31*	5.64**	1.63	0.60	222
CMh§	2.23*	6.37**	0.43	0.58	222

*** correlation is significant at the 0.01 level (2-tailed)*
** correlation is significant at the 0.05 level (2-tailed)*
§ log transformed

Effect of environmental factors

Table 3.5 shows the correlation coefficients between the scores of principal components, distance to drains, groundwater-table depth and groundwater salinity with soil salinity. In most cases, the correlations were low (<0.5), similar to findings by Park and Burt (2002) who explained that because of the complex distribution of soil properties higher correlations are possibly unlikely. In a previous section, soil texture was found to be a good explanatory variable of salinity. Nevertheless, the influence of the terrain-related PCs (2, 4, 5, and 6) was low. This is not surprising, because most 'information' from topography will already be included in the basic soil properties, and there may be considerable uncertainty associated with topographical attributes, as was already noted by Leij et al. (2004).

Interestingly, TDS or Cl content was not controlled by any principal components derived from terrain indices. This might again be attributed to the fact that topsoil salts are dynamic in nature, and land management practices (e.g., leaching) might have contributed greatly to spatial distribution. Moore et al. (1993) also emphasized that surface soil properties are most modified by land management and therefore features of lower horizons in the profile may show greater response to topographic attributes. The statement is supported by bulk soil salinity measurements (CM-138), where terrain indices had a low but significant influence. This supports the hypothesis that overall soil salinity is developed in response to the way water moves through the landscape.

Table 3.5 Pearson's correlation matrix between principal components, distance to drainage, groundwater table depth and salinity and soil salinity

Variable	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	DCOLL	GWT	GWS§
CMv	-0.28**	-0.07	-0.18**	-0.02	-0.21**	-0.14**	0.44**	0.35**	-0.42**
CMh§	-0.23**	-0.15*	-0.33**	0.16*	-0.11	-0.11	0.45**	0.36**	-0.70**
TDS§	-0.19**	-0.08	-0.28**	-0.04	-0.09	-0.02	0.10*	0.04	-0.13**
Cl§	-0.28**	-0.06	-0.33**	-0.08	-0.08	0.01	0.15**	0.13**	-0.18**

*** correlation is significant at the 0.01 level (2-tailed)*
** correlation is significant at the 0.05 level (2-tailed)*
§ log10 transformed

Principal components consisting of remotely sensed data (PC 1 and 3) correlated significantly with all salinity parameters due to the contribution of vegetation

cover to reflectance values and the subsequent vegetation indices calculated from band combinations.

Distance to the drains (dcoll) correlated slightly but significantly with TDS and Cl. Since the existing drainage network was built to lower the groundwater level, its functionality should be indirectly reflected by changes in soil salinity. Close to the drains, soil salinity should be lower and should increase with distance from the drain. Apparently drains have little influence on topsoil salinity, which might be explained by the high variability of salinity at small scales or the irregular layout of the drainage network. Another factor to be kept in mind is that the drainage network layout was obtained from old maps, and there could be minor changes to the irrigation and drainage canals. These would not be included in the map, since no up-to-date information was available.

Conversely, drain proximity significantly influences CM-138 readings as expected, which is similar to the effect of terrain attributes. The CM-138 shows the average amount of salt within a larger volume and is less affected by land management than topsoil salinity.

As anticipated, groundwater table depth (GWT) and salinity (GWS) had high correlation coefficients, but the direction of the influence is somewhat contradictory to that presumed. The positive sign of the correlation coefficient of groundwater table depth with salinity suggests that salinity was higher when groundwater was deeper. While this could be explained with certain soil textures, the analyses (correlation and scatter plots) separating soil texture were complicated. It was difficult to discriminate the effects dominating such behavior, and therefore it is difficult to make any conclusive statements.

A similar behavior was observed for groundwater salinity. The data indicated that soil salinity increased with decreasing groundwater salinity. Taking into account the possible errors in the data of the groundwater table depth obtained for the study area and in the salinity data, it was considered prudent to refrain from drawing conclusions from this dataset.

3.5 Discussion

3.5.1 Methodological approach

The grid sampling design with nested grids (vs. selective sampling) implemented in this study provided a good way of exploring the spatial structure of soil salinity distribution. The regular grid sampling captured both the large scale variation and the nugget effect. The semivariograms for the topsoil salinity indicate that a large degree of the variability is associated with the nugget effect, and that relatedness between spatially separated measurements is limited. Among many factors that could cause small-scale (within 1600 m^2) variation of the topsoil salinity, the uneven soil surface (because no leveling is done) and the frequent presence of subsurface impermeable layers could account for such variation. Land management (e.g., leaching) is another important factor to account for variability of topsoil salinity. Therefore, when interpolating with kriging modifications as suggested by Walter et al. (2001) (who suggests that ordinary kriging with a local variogram integrates the local spatial structure, whereas kriging with a whole-area variogram only considers the sampling intensity) may provide improved interpolation, because kriging is sensitive to small-scale variation, although it is less sensitive to outliers.

The environmental attributes considered as the possible controlling factors were selected with a view that they are representative of the area and can be easily extracted for the rest of the Khorezm region. The only variable that was discarded was ‘distance from main collector’ (the collector which carries drainage water out of the region), which was initially considered in the analyses. It had the highest (0.75) correlation coefficient with CMv; this was unexpected and it was decided to drop the variable from further analyses, because it is not representative of other areas of the Amu Darya delta.

The negligible effect of terrain attributes on soil salinity expressed by low correlation coefficients can be considered an artifact of the correlation tool used. It is a well-known principle that landform has a significant effect on salinity distribution, and for the study area was confirmed by own observations. Indeed, the ordinary correlation coefficient cannot account for the presence of both significant negative and positive correlations existing between two variables at different frequencies (Nielsen et al., 1983). Therefore it is apparent that standard correlation or analysis of variance methods

would yield very little information. The use of other methods, such as a sensitivity analysis by artificial neural network will be explored in section 4.4.1.

3.5.2 Interpretation of results

The characterization of the spatial distribution of the soil salinity by geostatistical analyses shows that topsoil salinity is highly variable, whereas bulk soil salinity to a depth of 1.5 m is less so. Therefore, when interpreting analyses results, more emphasis should be given to the CM-138 measurements, because they provide the average of the soil profile salt content and seem to be less affected by disturbances or other human activities.

The distribution pattern of parameters such as clay and humus content approximately reflects the topography. The strong influence of topography on the clay content can be visually confirmed by the generated maps (Figure 3.8). As soil texture largely influences salt content, it could be expected that terrain exerts a strong influence on salinity similar to soil texture or humus content. The lack of variance of soil salinity explained by topography suggests that there are some other factors not considered here or constraints inherent in local terrain. The environmental factors were selected from the literature to ensure they were included in the analyses. The constraints inherent in local terrain could be important and are mentioned here.

Several constraints should be taken into account when considering the results from this analysis. As mentioned before, terrain indices are mostly well pronounced for the topography following catenary development. The study area is mainly flat, and delineating it into landscape units according to a catena was not possible. However, some catenary distribution of the topsoil salinity could be discerned if isolated parts of the study area were considered individually. Figure 3.8 shows that the lowest points on sandy soils had high TDS values. In contrast, the same sandy soils on the slope (although slopes here are gentle they have a marked effect) had low values of TDS, while heavier soils on the lower slopes had higher TDS values. Nonetheless, confirmation of this micro-catenary effect would require a greater number of samples and study slopes.

Since the research site represents an area of recent land development where various transformations are still under way (e.g., cropping areas that were previously

unused, filling in depressions and converting them into fields), the soil properties are not in the steady state statistical approaches assume. Park and Burt (2002) also stress that the soil-forming factors at a given point may change through time, and this makes it more difficult to relate soil salinity to landform geometry.

Similarly groundwater table depth and salinity are affected by the interpolation technique. This probably increased the error of the groundwater data, whose reliability on this scale was doubtful to start with. The author is aware of the errors peculiar to the monitoring procedure of observation wells and accepts these. Also, the range of groundwater table depth and salinity is not very high and thus does not allow inferences to be made. Perhaps groundwater observation data taken for July only was not 'sufficient' to account for the processes that influence salinity. Therefore, no conclusive statements can be made on the direction of the correlation coefficients of soil salinity and groundwater depth and salinity.

The significance of the effect of soil texture on soil salinity is encouraging. However, as Park et al. (2004) stress, larger study areas generally show poorer environmental correlation due to the additional heterogeneity of the environmental factors. The use of a large number of variables obtained or calculated from remote sensing or topography would improve the ability of the model to predict soil salinity. However, the danger of multicollinearity exists, and a minor change in one variable could then have a great influence on model output. For further use, it would be better to find one factor per principal component that would represent the principal components extracted from these variables.

3.6 Conclusions

The spatial distribution of soil salinity and influencing factors in the landscape of the Khorezm region were investigated by linking environmental variables to soil salinity. Topsoil salinity was seen to be highly variable even at short distances (40 m) compared to average bulk soil salinity (0.75 and 1.5 m) measured by CM-138.

Soil texture had a significant influence on soil salinity and was largely dominated by topography. However, soil salinity was poorly correlated with terrain attributes. The likely reasons for these poor correlations of terrain attributes and soil salinity are flat topography, land management practices that were difficult to incorporate

in this study or the failure of the correlation tool (the reason is discussed in section 5.2) used to study soil salinity relationship with terrain attributes.

Factors obtained from remote sensing had significant correlation coefficients with both salinity of topsoil and measured by the CM-138. Since band signals and calculated indices are mainly an indication of vegetation, the correlation suggests that salinity affects crop growth significantly and can be used as a remote sensing indicator.

Distance to drains is an important factor especially for the bulk soil salinity of the profile. It was lower for the topsoil which might be due to higher spatial variation of the topsoil salinity due to land management practices (e.g., leaching).

Groundwater table depth and salinity show a high correlation with soil salinity; however, the direction of the influence could not be explained. The lack of groundwater observation wells in the study area, unreliability of the data, and interpolation errors should be taken into account.

The next chapter deals with the construction of a model to predict soil salinity and fuse factors that explain best the spatial distribution of the salinity.

4 UPSCALING THE SPATIAL DISTRIBUTION OF SOIL SALINITY BASED ON ENVIRONMENTAL FACTORS

4.1 Introduction

The demand in precision agriculture has increased the need for accounting for the variability of soil properties in more detailed scales (Petersen et al., 1995), at least in developed countries. Fulfilling this need has been hastened by the wider application of GIS in natural sciences, and Moore et al. (1993) note that one of the reasons that GIS technology has been readily adopted is because it allows spatial information to be displayed in integrative ways that are readily comprehensible and visual. Petersen et al. (1995) emphasize that quantification of spatial relationships of observable landscape attributes and soil spatial variability is a scientifically based and potentially cost-effective means of both mapping soils and understanding pedology at multiple scales. Examples of the progress in application of such studies in soil salinity mapping are summarized in Table 4.1.

4.1.1 Objective of the study

This chapter describes the construction of a soil salinity model at farm level and to upscale this to a higher level, here the district level of Khorezm. There are numerous studies on soil salinity; however, salinity studies where the landscape has been under cultivation for centuries and which hence is characterized more by human impact than by the natural environment are lacking. In fact, studies devoted to irrigated landscapes within drylands ("oasis") are very limited. As in any other empirical soil-landscape study, the basic assumption is that a constructed model in a representative area will be applicable to unsurveyed areas where similar environmental conditions exist. The overall aim is to make the best use of available data, such as topography, soil texture, groundwater observation wells dataset, drainage network, and remote sensing as an alternative to soil salinity surveys, or to improve the quality of traditional map generation procedures.

Table 4.1 Selected studies on soil salinity or electrical conductivity and relationship with terrain parameters

Author(s)	Environmental settings	Topographical variables	Statistical procedure(s); results
Odeh et al. (1991)	Forreston, Australia; phyllies, siltstones and shale; sheep grazing	Slope gradient; plan convexity; profile convexity; upslope distance; upslope area; square root of upslope area; aspects (sine, cosine); solum depth, depth to bedrock	Canonical correspondence analysis; principle components analysis, redundancy analysis
McKenzie and Austin (1993)	Macquarie Valley, NSW, Australia; irrigated agriculture and grazing	Slope; impeding layer; relief; landform classification; gilgai presence; upslope distance	General linear model (multiple regression); R^2 = 27% (10 cm depth); R^2 = 17% (70 cm) depth
Roberts et al. (1997)	Tout Park, New South Wales; cropping, grazing, and orchards; Cuballing, Western Australia; sheep and cattle grazing	Elevation (high), lowness, plan curvature, upness, upness equalized, low, concavity upness equalized	Fuzzy set theory
Searle and Baillie (1997)	Gympie and Ipswich, Queensland, Australia; broad range of land uses	Alluvium; basalt; basalt fringe; laterite; laterite fringe; dykes; faults; plan curvature; profile curvature; rainfall; geology; soil; vegetation; wetness indice	Linear additive method in GIS; R^2 = 0.53 – 0.86
Evans and Caccetta (2000)	Dumbleyung and Mt. Barker regions, WA, Australia	Average upslope height, average upslope slope, flow slope, height above nearest salt within watershed, height above nearest stream within watershed, upslope area, flow path length, percentage upslope cleared area, total upslope area	Decision tree model, expert knowledge; risk and non-risk prediction accuracy 78% and 80% respectively

4.2 Modeling procedure

Different tools and research approaches were applied in this study varying from conceptualizing the salinization pattern, relating it to the soil-landscape using statistics and GIS. Although the conceptual models account for most factors, the tools which deal with parameterization all have inherent errors or uncertainty. Complex models can be viewed as one way of reducing errors; in other words, the inclusion of a large number of explanatory variables is a way of reducing uncertainty of the data. However, Jansen (1998) warned that complexification of models beyond a certain degree can even increase errors.

On the other hand, being cautious of errors, the current state of technology can theoretically solve complex models by powerful computers and, with the generation of proper variables, could achieve accurate results. However, as McBratney et al. (2000) correctly note, uncertainty, imprecision and ambiguity are inevitable or inherent parts of natural systems such as soils. Heuvelink (1998) warns that it is illusory for a modeler to aim at creating an exact copy of the real world. The principle known as Ockham's razor (i.e., keep it simple) would caution to avoid complexity, and this was adhered to here in modeling soil salinity. Furthermore, some parameters, precise enough to make soil salinity modeling manageable, may be difficult to obtain, causing a trade-off between predictive precision and costs/effort of the measurement. In fact, the complexity of many models stems from overemphasized precision, which does not always mean greater truth (McBratney et al., 2000). Therefore it is common to anticipate only 70 % of the total variance of a soil attribute to be explained by the use of terrain models (Moore et al., 1993).

In this study, the spatially distributed model of salinity works at different scales using an approach characterized mainly by comprehensive models. The methods are collectively categorized into the emerging field termed *pedometrics* (McBratney et al., 2000). The definition of pedometrics is "the application of mathematical and statistical methods for the study of the distribution and genesis of soils". The schematic diagram of the techniques used in pedometrics is represented in Figure 4.1.

The scale issue should be mentioned here, because factors well correlated at the farm scale might not be well suited at larger scales. It has been stressed by many studies (Heuvelink and Pebesma, 1999) that factors obtained at lower resolution can be

obliterated at higher scales. That is one of the main reasons why the majority of models end up being scale-specific. Therefore, to avoid ending up with a scale-specific model, the variables that are available should be used. Heuvelink and Pebesma (1999) suggest to interpolate first and then calculate, and they further explain that information of the specific spatial correlation structure of the independent variables is lost when the dependent variable is calculated before interpolation. On the other hand, uncertainty at different scales can be overcome if independent variables are used as covariates when kriging, but this gets complicated for larger models.

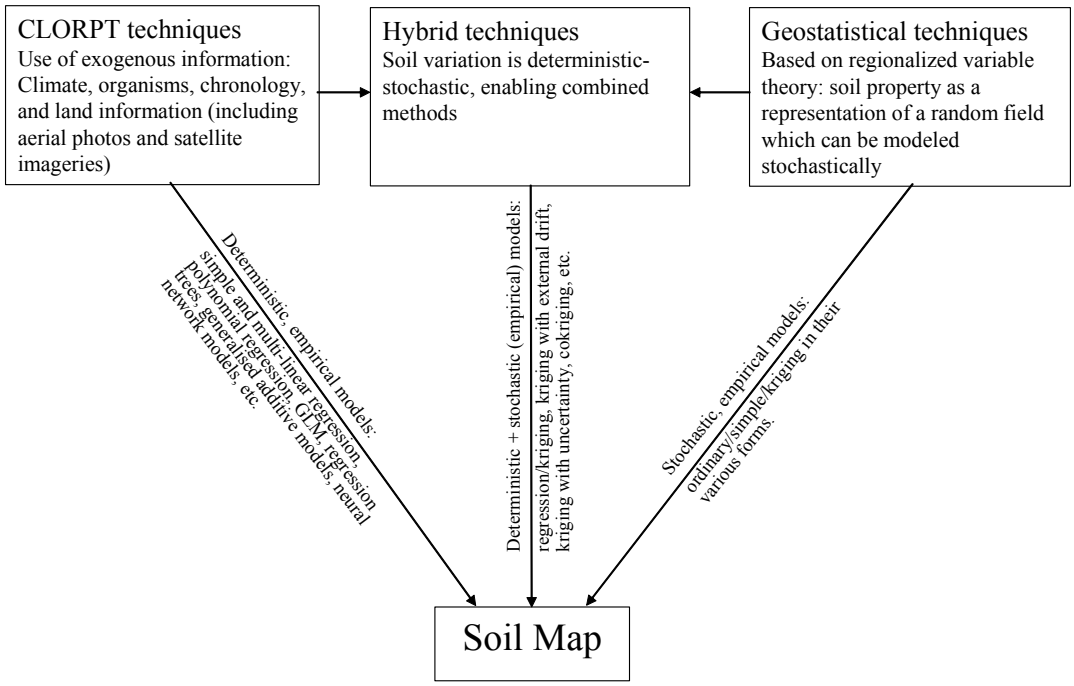

Figure 4.1 Generic pedometric techniques, their hybrid and styles with respect to mode of spatial prediction (Source: McBratney et al., 2000). (CLORPT stands for climate, organisms, relief, parent material, temperature)

Building on the premises mentioned above, an attempt is made here to build a model that is capable of explaining soil salinity distribution while keeping the model simple and at the same time practical enough to be suitable for extension services. the soil-landscape concept, implemented to characterize the spatial distribution of soil salinity, identified several correlating factors thought to govern its spatial distribution. In a next step, the salinity distribution was modeled using the identified factors at the farm scale and this model used to upscale the salinity distribution to the district level. The research chain used to achieve the final result can be illustrated using the diagram

of Hoosbeek and Bryant (1992) (Figure 4.2). They used two perpendicular axes to classify models (from quantitative to qualitative and from empirical to mechanistic) to obtain or describe the data used. Different scale hierarchical levels on the vertical axis are distinguished, ranging from molecular interaction to world scale, with plot scale designated in the center and denominated with "*i*". The details of the scale diagram and its use in some studies are given in (Bouma et al., 1998; Curmi et al., 1998; Hoosbeek and Bouma, 1998).

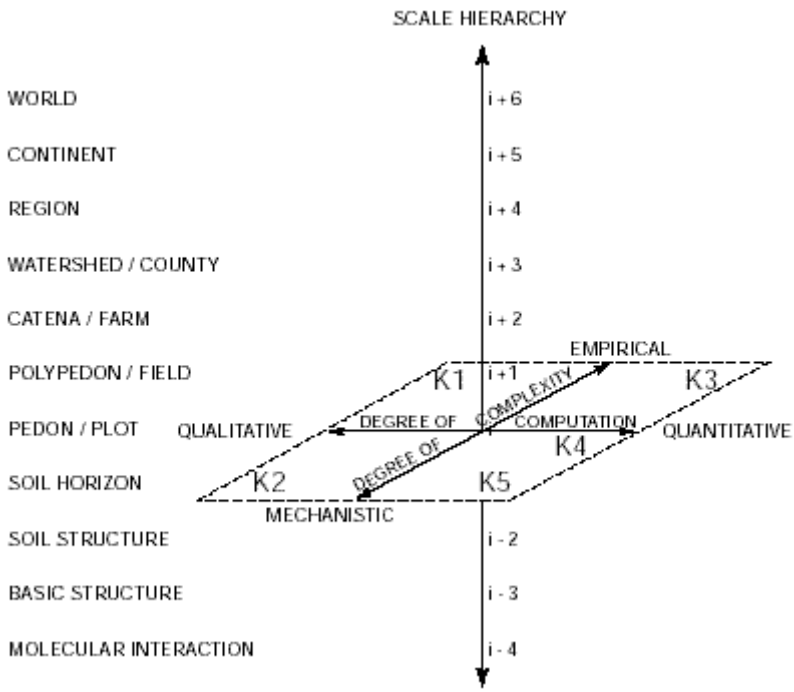

Figure 4.2 Classification of research procedures based on scale hierarchy, degrees of computation and complexity (Bouma and Hoosbeek, 1996; Hoosbeek and Bryant, 1992). Knowledge types: K1 = user expertise; K2 =expert knowledge; K3 = simple comprehensive models; K4 = complex comprehensive models; K5 = knowledge from very detailed specialistic models

4.2.1 Statistical analysis: artificial neural network approach

The complex nature of the environmental variables and assumptions in classic statistical analysis often make the use of statistical theories difficult. An alternative is the use of artificial neural networks (ANN), the computerized information processing model, the development of which was motivated by the way the brain works. It is able to modify its structure by learning from the data rather than being programmed, and is thus able to solve complicated relations (Principe et al., 2000). In salinity studies, ANN have been

used for a variety of purposes, including the prediction of a salt build-up in the crop root zone (Patel et al., 2002), dielectric constant-soil water relationships (Persson et al., 2002), and river water salinity forecasting (Bowden et al., 2002; Maier and Dandy, 1999).

In studies with a spatial domain, Eklund et al. (1998) applied a neural network algorithm for inductive inference from GIS layers to classify soil salinity risk potential. Within the soil-landscape dimension, Zhu (2000) showed that the neural network approach reveals much greater spatial detail and has higher quality than that derived from the conventional soil map. Similarly, (McBratney et al., 2000) gave an example of clay content estimation where neural networks performed among the best. In all these studies, as well as in the majority of other studies, feed-forward networks have been used. Feed-forward neural networks are composed of layers of neurons, in which the input layer of neurons is connected to the output layer of neurons. The training process is undertaken by changing the weights such that a desired input-output relationship is realised.

Multilayer perceptrons (MLPs) are layered feed-forward networks typically trained with static backpropagation (Principe et al., 2000). With backpropagation, the input data is repeatedly presented to the neural network. With each presentation, the output of the neural network is compared to the desired output and an error is computed. This error is then fed back (backpropagated) to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output (this process is known as "training"). Principe et al. (2000) emphasize that their main advantage is in being easy to use, and the key disadvantages are that they train slowly and require a large amount of training data (typically three times more training samples than network weights). However, MLPs with a sufficient number of hidden units can approximate any continuous function to a pre-specified accuracy; in other words, MLP networks are universal approximators (Cherkassky and Mulier, 1998, p. 259).

In this study, ANNs are used as an alternative to regression techniques for prediction of the soil salinity distribution. More than four hundred cases of field-collected data are used for model development. Further details of the model building are given in the methodology section.

4.3 Materials and methods

4.3.1 District description

Site descriptions for the farm scale study are given in section 2.2.2. The district level mainly consists of the part of the Khiva district where the previous study area is located. Khiva district is one of 11 districts of the Khorezm region and is situated in the south-west of the region (Figure 4.3). The area for upscaling excludes parts of the district in the very south, which are occupied by desert sands and instead includes the area in the north where data were available. Thus, the total area at the district scale used in this study is approximately 407 km^2. A range of attributes present at farm scale are also represented in the larger region, with the difference that rice is grown in many areas of the district, whereas at the study farm, rice was not grown during the sampling year.

The topography of the district is similar to that of the farm, i.e., flat, with elevation points normally distributed ranging from 85 to 110 m above sea level (mean 93 m and standard deviation 2.9). The landscape is dissected by an extensive network of drains and collectors (Figure 4.4).

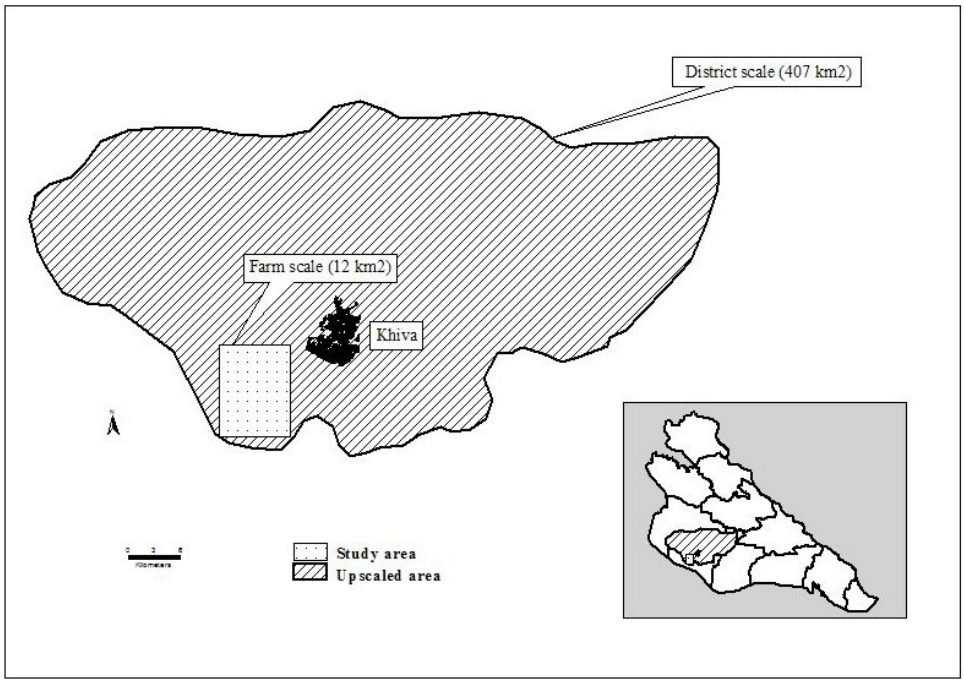

Figure 4.3 Location of study area for model development and district area for upscaling

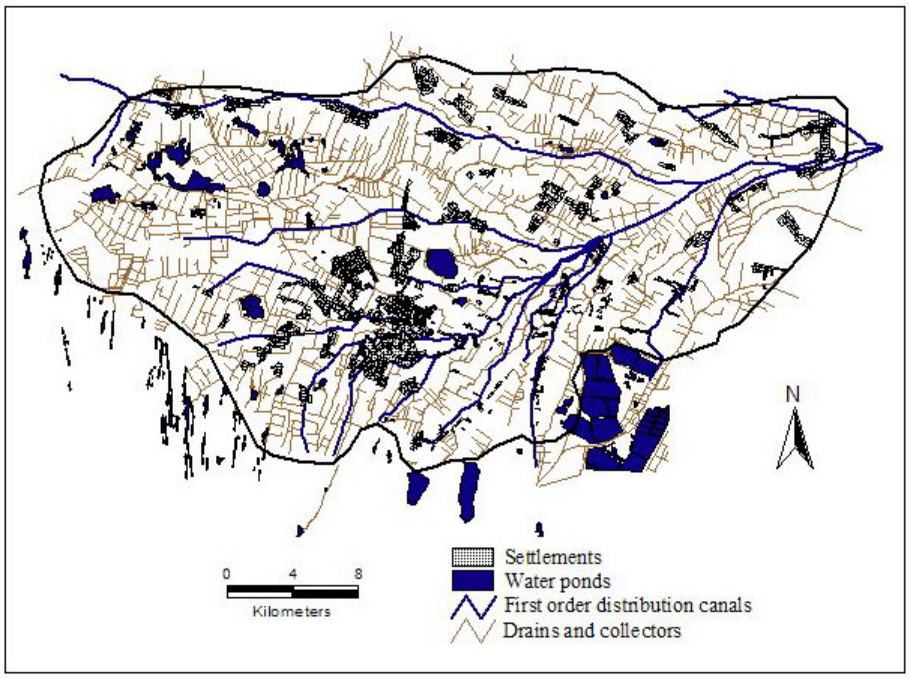

Figure 4.4 District map of drains and first-order irrigation canal

Soils

The soils in the district are classified by the local classification system as desert zone meadow-oasis soils and by the World Reference Base (WRB) as Cambisols, which

comprise about 11 % of the soils in Uzbekistan (Abdullaev, 2002). These soils have a moderate humus content (0.8-1.4%) compared to the rest of the soils in this desert zone. They are developed on the quaternary alluvial fan deposits from old river channels. These bedded deposits can be distinguished by river bed, near-bed, and lake facies with the thickness of the alluvial deposits ranging from a few centimeters to over 20 m (Tursunov and Abdullaev, 1987). This differentiation is explained by sands distally replaced by silty and clayey sediments away from the river beds.

Lake facies are developed as a result of the river or irrigation activities, where depressions serve as a sink for water accumulation, and these areas are represented by relatively homogeneous loamy and clayey soils. Since the land in the area has been constantly under development, many studies distinguish soils by irrigation duration, basically referring to them as soils of ancient or recent development. Hydrogeological conditions play an important role in soil formation, as does the evolution of the soil profile. Processes of salinization/desalinization are related to the presence of shallow groundwater levels (Tursunov and Abdullaev, 1987). Thus, although the processes of soil formation in the district and the region as a whole are understood, their intricate occurrence patterns can be complex.

Available previous soil survey

The Uzbek Institute of Soil Science and Agrochemistry was authorized by the government to carry out a soil inventory and assessment of land resources of the country. The last soil survey of the Khorezm region was carried out during the years 1996-1997. A data set that aggregates measurements from 54 soil profiles was made available by the institute. Due to a number of missing values in the dataset, the total number of observations may vary for specific variables. The soil texture for the district level was estimated using this dataset, where values of the top horizons were selected. The top horizon was selected in order to match the depth for the survey conducted at the farm scale. Three values for the texture variable were missing; the average depth of the 51 top horizons was 31 cm (including one shallow horizon of 18 cm and a deep horizon of 55 cm depth).

4.3.2 ANN model for soil salinity prediction

ANN simulation model characteristics

The PC-based software package NeuroSolutions 4.31 (NeuroDimension, Inc.) was used in this study to simulate the artificial neural network operation. The package is integrated into the spreadsheet. The steps outlined in Maier and Dandy (1999) and Bowden et al. (2002) are used as a guide in this work. The default values suggested in the software are used unless stated otherwise.

The dataset of 441 cases was split into 3 subsets, namely training, cross-validation, and testing. Here, the training subset is used to train the network, the cross-validation subset is used to avoid the neural network overtraining (training is terminated when mean squared error (MSE) of the cross-validation subset increases), and the testing subset is used to test the built model on unseen data. Bowden et al. (2002) showed that data division into subsets is important for the model performance, and three subsets should represent the same population, thus similar populations in all subsets were aimed at. Some variables could not be divided into equal subsets, which can be attributed to singular or rare events present in the dataset. However, in general the statistics of the subset are in good agreement and appear to come from the same population. The statistical summaries of the subsets are given in Table 4.2.

A multi-layer perceptron (MLP) with one hidden layer with 12 *processing elements* was built with the help of a wizard provided in the software. Details are provided in the software manual, but it is generally advised to experiment with different numbers and types of the composing components. Nonlinearity was introduced by the *hyperbolic tangent function (TanhAxon)*, and the *momentum* learning rule was used for updating the weights. Predictions of the neural network were assessed with the correlation coefficient R between measured and estimated values, means squared error (MSE), and mean absolute error, MAE. The number of *epochs* (iterations) was set to 10000 and the network was trained 10 times. The best weights during these runs were applied for the testing subset. The best weights were then applied to the *production* (dataset to be predicted) subset.

Table 4.2 Statistics of subsets of variables used for training, cross-validation, and testing of the neural network

Subset	Variable	Unit	Mean	Standard Deviation	Minimum	Maximum	CV (%)	N
Inputs								
Training	clay	%	8	5	0.1	22	60	265
Cross-validation	clay	%	9	5	0.1	20	59	66
Testing	clay	%	9	5	0.1	21	63	110
Training	silt	%	33	20	0.3	87	62	265
Cross-validation	silt	%	31	19	2	73	62	66
Testing	silt	%	32	20	2	74	62	110
Training	drainage distance	meters	192	141	3	567	73	265
Cross-validation	drainage distance	meters	209	141	5	456	67	66
Testing	drainage distance	meters	191	147	3	500	77	110
Training	log RVI (ratio vegetation index))		0.07	0.11	-0.13	0.39	150	265
Cross-validation	log RVI		0.08	0.11	-0.12	0.29	140	66
Testing	log RVI		0.07	0.1	-0.11	0.33	148	110
Training	band3		76	18	44	131	24	265
Cross-validation	band3		79	21	47	136	28	66
Testing	band3		77	19	36	126	24	110
Training	band5		78	19	26	136	25	265
Cross-validation	band5		83	23	52	142	28	66
Testing	band5		78	21	27	141	27	110
Training	upslope area*		8.37	1.39	5.65	13.66	17	265
Cross-validation	upslope area*		8.45	1.35	5.85	12.06	16	66
Testing	upslope area*		8.67	1.56	5.78	15.46	18	110
Training	wetness index		12.38	1.87	8.92	20.57	15	265
Cross-validation	wetness index		12.30	1.66	9.44	18.16	14	66
Testing	wetness index		12.74	2.27	9.29	21.76	18	110
Training	slope	degrees	0.1	0.08	0	0.54	79	265
Cross-validation	slope	degrees	0.1	0.06	0.004	0.35	65	66
Testing	slope	degrees	0.1	0.08	0.003	0.35	82	110
Training	aspect	degrees	188	97	2	360	52	265
Cross-validation	aspect	degrees	183	98	23	354	53	66

Subset	Variable	Unit	Mean	Standard Deviation	Minimum	Maximum	CV (%)	N
Testing	aspect	degrees	187	91	11	360	49	110
Training	slope length	100 m^{-1}	0.03	0.03	0.00	0.25	111	265
Cross-validation	slope length	100 m^{-1}	0.03	0.03	0.0004	0.13	97	66
Testing	slope length	100 m^{-1}	0.03	0.03	0.0002	0.15	115	110
Training	divergence-convergence		-0.26	11.71	-38.39	53.82	-4572	265
Cross-validation	divergence-convergence		2.69	10.22	-25.10	44.02	380	66
Testing	divergence-convergence		-3.28	12.95	-55.31	34.06	-395	110
Training	curvature		-0.000001	0.00005	-0.0001	0.0003	-10337	265
Cross-validation	curvature		0.000005	0.00004	-0.0001	0.0001	804	66
Testing	curvature		-0.000005	0.00004	-0.0002	0.0002	-866	110
Training	planimetric curvature		0.000001	0.00003	-0.00008	0.0002	3235	265
Cross-validation	planimetric curvature		0.000004	0.00002	-0.00006	0.00007	642	66
Testing	planimetric curvature		-0.000003	0.00002	-0.000068	0.0001	-579	110
Training	profile curvature		-0.000001	0.00003	-0.0001	0.0001	-2308	265
Cross-validation	profile curvature		0.000002	0.00002	-0.000048	0.00007	1168	66
Testing	profile curvature		-0.000001	0.00002	-0.000094	0.000097	-2291	110
Training	surface curvature (curv7)		-0.0004	0.005	-0.03	0.02	-1556	265
Cross-validation	surface curvature (curv7)		0.001	0.005	-0.01	0.02	382	66
Testing	surface curvature (curv7)		-0.0006	0.006	-0.02	0.02	-1006	110
Training	terrain characterization index		-0.003	0.02	-0.08	0.07	-750	265
Cross-validation	terrain characterization index		0.003	0.02	-0.05	0.05	572	66
Testing	terrain characterization index		-0.004	0.02	-0.12	0.05	-566	110
Training	solar radiation	kW m^{-2}	7.14	0.003	7.12	7.15	0	265
Cross-validation	solar radiation	kW m^{-2}	7.142	0.003	7.1367	7.1474	0	66
Testing	solar radiation	kW m^{-2}	7.14	0.003	7.13	7.15	0	110
Training	elevation	meters	92	1	90	94	1	265
Cross-validation	elevation	meters	92	1	90	94	1	66
Testing	elevation	meters	92	1	90	94	1	110
Training	groundwater salinity	dS m^{-1}	3.13	0.79	1.9	5	25	265
Cross-validation	groundwater salinity	dS m^{-1}	2.94	0.65	1.97	4.8	22	66
Testing	groundwater salinity	dS m^{-1}	2.95	0.7	1.97	4.68	24	110
Training	groundwater table depth	cm	125	14	93	155	11	265

Subset	Variable	Unit	Mean	Standard Deviation	Minimum	Maximum	CV (%)	N
Cross-validation	groundwater table depth	cm	128	15	93	155	12	66
Testing	groundwater table depth	cm	126	13	92	155	10	110
Outputs								
Training	CM vertical	dS m^{-1}	0.51	0.33	0.03	1.92	65	265
Cross-validation	CM vertical	dS m^{-1}	0.46	0.33	0.03	1.26	73	66
Testing	CM vertical	dS m^{-1}	0.51	0.33	0.01	1.4	64	110
Training	log CMv		-0.42	0.37	-1.52	0.28	-88	265
Cross-validation	log CMv		-0.5	0.42	-1.52	0.1	-83	66
Testing	log CMv		-0.44	0.43	-2	0.15	-98	110

**values were log10 transformed*

ANN environmental variables

The model was first tried with a selected set of variables extracted from the remote sensing and terrain analysis with the notion that using one variable per principal component (Table 3.3) as a proxy would yield results similar to those using component scores, since the selected attributes have the highest component score within each principal component. The performance of the neural network models was not stable, and the test results of the built models were not satisfactory. Apart from the instable results of the model, the fact that different attributes play a different role at different spatial scales led to the decision to include all terrain attributes in the model together with selected remote sensing variables.

In addition, sensitivity analysis was performed to identify those input variables that had more of an effect on estimating soil salinity. The software has a built-in option for this analysis. It provides a measure of the relative importance among the inputs of the neural model and illustrates how the model output varies in response to input variation. The first input is varied between its mean +/- 1 standard deviations while all other inputs are fixed at their respective means. The network output is computed for 50 steps above and below the mean. This process is repeated for each input. A report is generated that summarizes the variation of each output with respect to the variation in each input.

4.3.3 Validation survey of the district scale map

The district was represented by 2000 points randomly generated by the Arcview sample generator extension (DNR, 2004). This number of points was arbitrarily chosen to cover the whole area of the district, and it was expected to provide a sufficient number of points in case the existing layers did not cover some parts of the district area, such as was the case with the soil survey information obtained from the Soil Research Institute. This soil survey did not cover the northwest part of the district area, and therefore the number of points was reduced to 1755. The trained model was subsequently applied to these remaining 1755 points with the same attributes used in training, obtained and calculated from remote sensing, terrain, distance to collector, texture, and groundwater data. The resulting map from predicted values of soil salinity was interpolated using the kriging method in ArcMap.

The validation of this generated salinity map was done by ground-truthing at 315 randomly selected locations. Sampling sites were randomly selected but with a view to logistics and accessibility and were required to cover the whole area of the delimited district area. Their coordinates were uploaded into a handheld GPS in advance and navigation by GPS was used to locate the sampling points on the ground. The timeframe for the validation survey was scheduled to be in summer, running for about 4-5 weeks and starting in July. This was exactly one year after the previous farm-scale survey. The same set of measurements was collected at each sampling location as described in section 2.2.3. Vertical electromagnetic conductivity measurement (CMv, dS m^{-1}) was used in the analysis, because the results of the analyses using CMh (which is measured horizontally) were similar in most cases. To remind the reader, CMv and CMh measuring depths are roughly 1.5 and 0.75 m, respectively. CMh analysis was used to assess the difference in measurements taken in summer and autumn.

Comparison of measured and estimated soil salinity values can be carried out in several ways. As the nature of the relationships between soil salinity and controlling attributes used in the model is complex, a straightforward standard correlation is not expected. At first, to see how soil salinity was distributed, their values were reclassified into 7 ranges (arbitrarily chosen depending on the map outcome for each measured and estimated map values) corresponding to <0.75, 0.75-0.85, 0.85-0.95, 0.95-1.05, 1.05-1.15, 1.15-1.25, and >1.25 dS m^{-1}. The subsequent procedure is to compare these reclassified data sources, in this case by the widely used Cohen's Kappa statistics. Kappa statistics give a measure of agreement between estimated and measured categories, represented in this case as percent of accurate matches.

Since dividing values into classes hinders comparison of continuous data, for example if a grid represented by (x, y) has a measured value of 0.74 dS m^{-1} and is assigned to one class, and the same grid (x, y) has a predicted value of 0.76 dS m^{-1} and is assigned to another class, the comparison of a grid's associated classes will not have a true meaning, because the difference between measured and estimated value is only in the order of ~3 %, and the estimate should be considered correct.

Therefore, the accuracy of a predicted map is estimated by *percentage error*, i.e., the area correctly predicted within 10, 20, and 30 % deviation from the measured values. These analyses were performed with raster maps, which were derived from

interpolation of the 1755 points used to estimate soil salinity. The presentation of the results is augmented by graphical analysis, showing the spatially delineated areas.

Additionally, map and measured CMv inspection was carried out to identify possible sources of errors between measured and estimated soil salinity values. For this purpose, the validation survey conducted during July 2003 was repeated in autumn (October 2003) after the end of the vegetation season (mainly cotton). The measurements taken during autumn covered most of the area with the exclusion of areas covered mostly by sands.

4.4 Results

4.4.1 Neural network model performance

In this section, the performance of the neural network that was trained on the data collected at the farm scale is described. The intention of the modeling was to use the trained model to upscale soil salinity to the district level in the next step.

Among the several neural networks tested, the multi-layer perceptron (MLP) with *momentum* as a learning rule showed the best results. Statistics of the performance are summarized in Table 4.3. The scatter plots of the measured and estimated salinity values for each subset are shown in Figure 4.5. The results indicate that the model performed well taking into account high correlation coefficients and low mean square error (MSE) between measured and estimated soil salinity values. MSE values are comparable between subsets, which indicates that the division into subsets yielded relatively uniform samples in these subsets. Table 4.3 shows slightly higher correlation coefficients for cross-validation and testing data sets, compared to the lower r for the training subset, which might be due to the presence of some extreme values contained in the training subset.

Table 4.3 Neural network results of training; estimates between measured and estimated soil salinity values

Data set	Correlation coefficient, R	MSE	MAE
Training subset	0.82	0.04	0.14
Cross-validation subset	0.85	0.03	0.14
Testing subset	0.83	0.03	0.14

As initially intended, the model was trained with selected environmental factors resulting from the principal components analysis. Two remote sensing and terrain attributes from each component were selected to build the model. However, the performance of the neural network was not always stable, and the results were not satisfactory (for example, different results with each training and low correlation with estimated soil salinity results). The inclusion of more input variables tended to improve the model performance. A similar tendency in the prediction of soil attributes was observed by Park and Vlek (2002). Additionally, Leij et al. (2004) demonstrated that for the different dependent soil-hydraulic properties, different sets of input variables gave more accurate results. Therefore, it seemed a reasonable decision to include all terrain attributes.

The improved performance of the model is also illustrated by the sensitivity analysis. This analysis showed that the response of the model was highly dependent on the following terrain variables, listed in decreasing order of importance: curvature (curv), plan curvature (planc), profile curvature (profc), and solar radiation (solar). It appears that soil salinity is largely influenced by the micro-topography, convexity or concavity, which tends to influence surface water concentration.

It is also worth noting that, apart from profile curvature (profc), the variable curvature (curv), plan curvature (planc), and solar radiation (solar) were not included in the initial model building process, since they did not have the highest component scores within the associated principal component (see section 3.4). Excluding these variables might be one of the reasons for the unstable model performance.

The strong influence of the terrain attributes indicated by the sensitivity analysis is encouraging, because it supports the concept of the previously mentioned soil-landscape relationship (see section 4.2) in this study. The explanation of spatial distribution of soil salinity by terrain information, i.e., shape, is attractive for empirical model building, because it does not require detailed soil surveys over the landscape. The major drawback of such models, however, is that they are scale specific or range (of values it had been trained on) specific, which is also noticeable in the scatter plots (Figure 4.5).

One striking feature in the scatter plots of all three subsets is the deviation in the lower and higher range between measured and estimated values. The network seems

to overestimate values in the low salinity range and to underestimate those in the higher range. Two obvious reasons could be: (1) the data fed into the model do not have enough samples representing the extreme ranges in order for the model to make good approximations or (2) model training, i.e., the approximation arrived at by the given way of model training (see section 4.3.2) was not sufficient, and further training with varying settings should be attempted. The chosen network result was the best among indefinite types and number of models tried, and this model produced stable and comparable results on a number of occasions when the model was started from scratch. Therefore, it was decided to keep this model 'as is' rather than to resort to transformations to improve the result, with the reservation that the produced model could be improved (i.e., with more hidden layers).

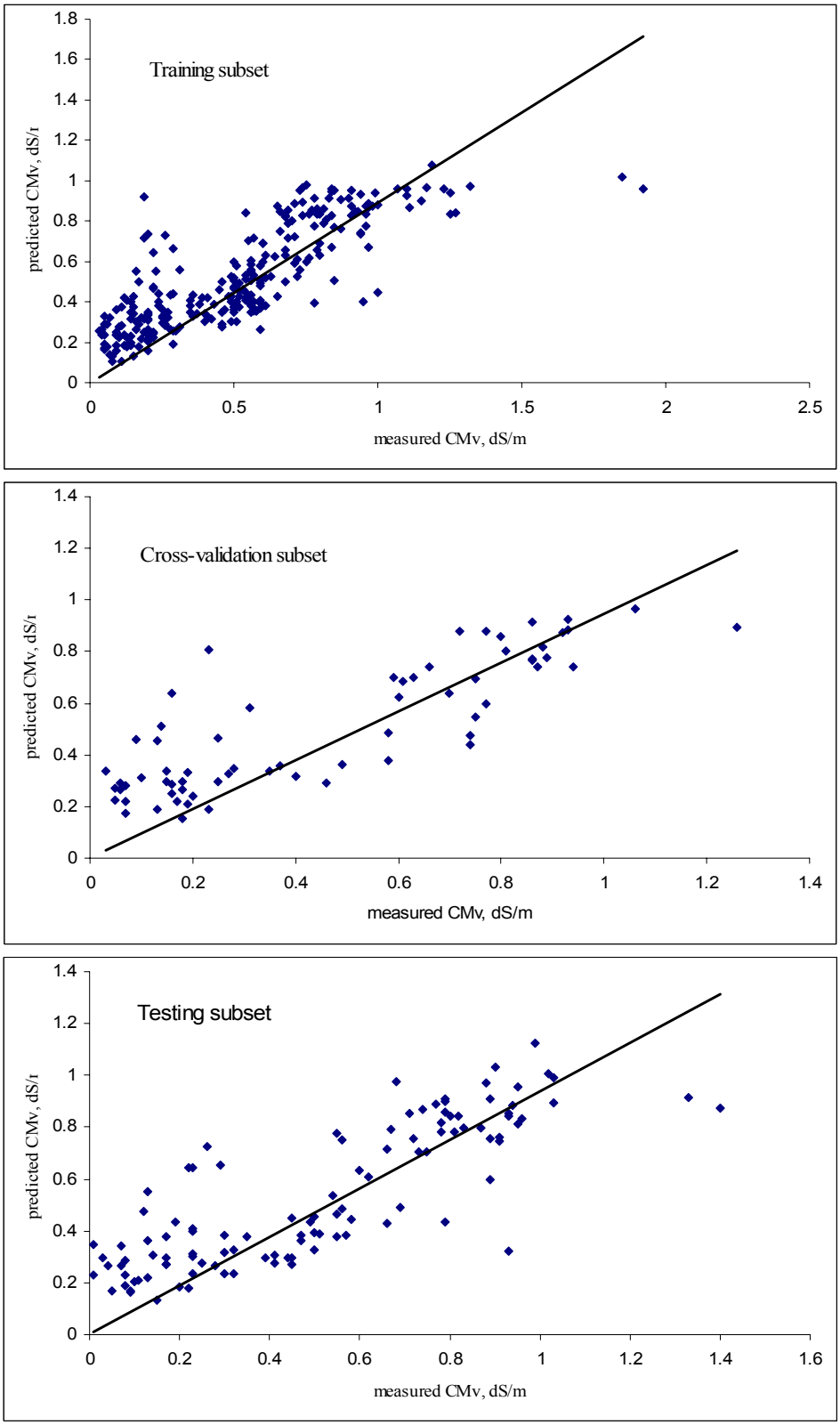

Figure 4.5 Scatter plots between measured and estimated CMv for different subsets in the neural network

4.4.2 Validation results

This part deals with the validation (or ground truthing) of the soil salinity at the district level which was estimated using the neural network model trained on the data collected from the farm survey.

The statistics of the measured and estimated soil salinity data sets are summarized in Table 4.4 (measurements from the autumn survey are included and will be referred to later). The interpolated maps of measured and estimated soil salinity are presented in Figure 4.6. The maps do not significantly resemble one another, and the Kappa statistics indicate that only 24 % of the grid values matched. Yet, there are areas in the measured salinity map that match estimated areas. The mean values for both maps are close to each other (0.94 and 1.04); MSE and RMSE are 0.05 and 0.25, respectively, and speculation of contrasting arguments with statistical tests can be made depending on which one needs to be proved.

Table 4.4 Statistics of trained, measured, and estimated values of soil salinity

		Trained	Estimated	Measured (summer)	Measured (autumn)
		----------dS m^{-1}----------			
Mean		0.51	0.94	1.04	0.61
Median		0.51	0.93	1.02	0.60
Mode		0.20	1.86	0.96	0.62
Std. deviation		0.33	0.11	0.19	0.20
Skewness		0.79	2.21	0.67	1.39
Minimum		0.03	0.41	0.59	0.23
Maximum		1.92	1.86	1.78	2.08
Percentiles	25	0.20	0.88	0.91	0.48
	50	0.51	0.93	1.02	0.6
	75	0.73	0.99	1.14	0.71

Despite the fact that the average value of the training subset was considerably lower than that of the estimated set (Table 4.4), the model performed well, i.e., it predicted the average salinity for the district close to the average value of the measured set. Perhaps this was due to the inclusion of the lowest and highest ranges into the training subset during the model training. This ability of the neural network to model complicated functions demonstrates that they are excellent approximators.

The choice of the approach for comparing results, i.e., the way to assign values to fixed classes only, could be questioned. Depending on the approach, the outcome can change. Close analyses of the results revealed that the majority of the data pairs were within the 0-20 % error interval. Therefore, when interpreting similarity between measured and estimated salinity maps, the significance of the error-based comparison should be given more weight. The following section discusses sources of error and, based on that the results, attempts to delineate areas where soil salinity is considered to be matching.

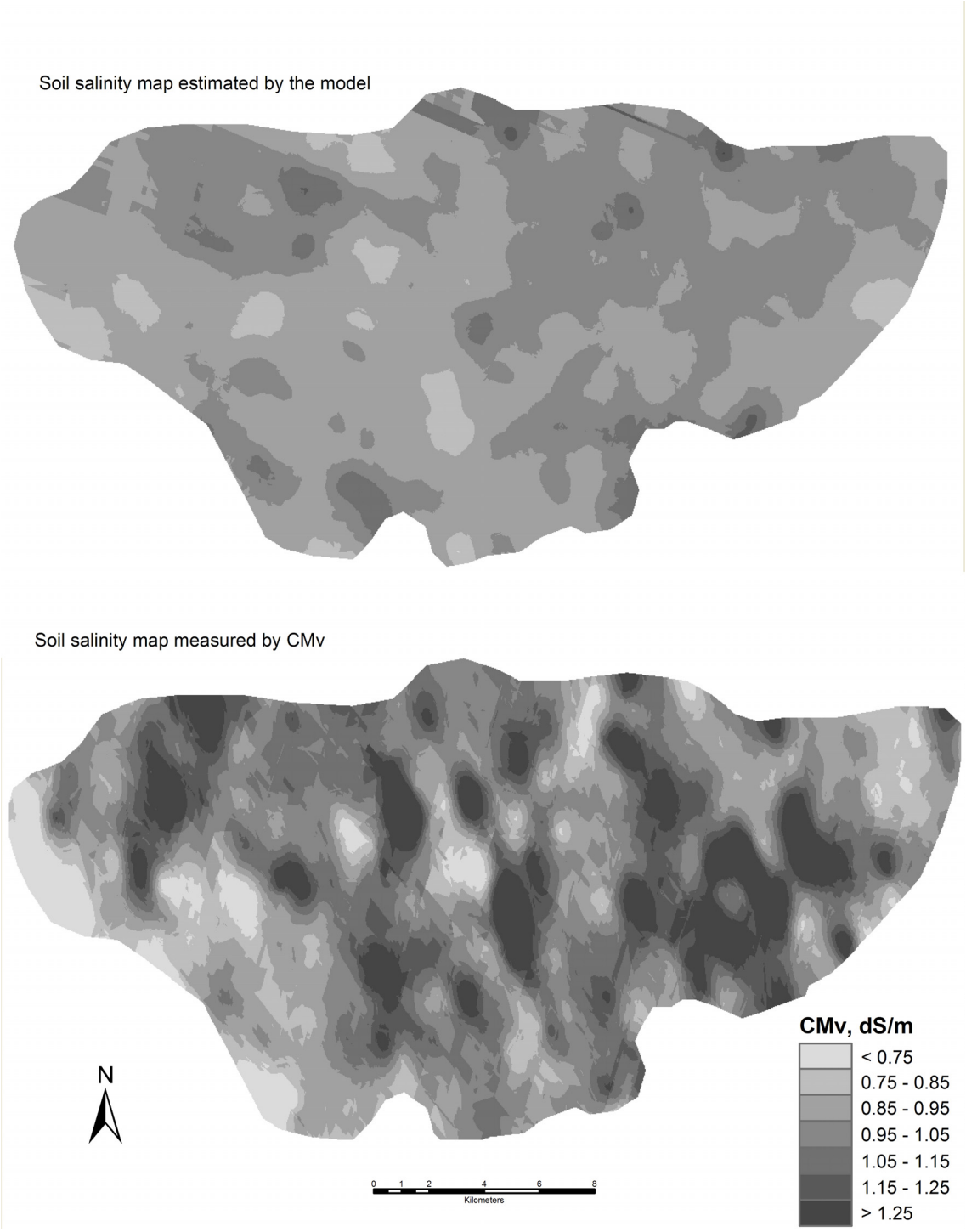

Figure 4.6 Measured and estimated maps of soil salinity at district scale

4.4.3 Sources of error and their possible influence on model outcome

Although when numerically compared, the mean estimated and measured soil salinity values at the district scale appear to be similar, this does not apply to a visual comparison of estimated and measured soil salinity maps. However, the deviation of the estimated soil salinity is not large, and this section explores possible sources of errors to consider when comparing two maps.

The difference between the measured and estimated average values suggests an underestimation by 10 %. One possible cause is that the training subset comes from a sample population with a lower mean value than the estimated data set. Another reason is differences between soil salinity relationships in convex, concave, and flat areas. Since the sensitivity analysis of the neural network identified curvatures to be the most influential, the errors in the model can be attributed to cycling patterns of correlation effects between soil salinity and considered variables, and the resulting behavior of the estimated salinity. Kachanoski et al. (1985) observed that A-horizon variability was due to spatial variability of surface curvature, but the relationship was complicated due to the presence of positive and negative correlations over different frequency ranges.

Graphically, one can plot measured and estimated salinity values sorted in ascending order by the percentage error. The overall number of points that are within a 10, 20, 30, and 40 % percentage error and the general trend of the measured and estimated soil salinity values can thus be better appreciated (Figure 4.7). First, it can be seen that in general the estimated values are below the measured values, suggesting soil salinity was indeed underestimated by the model. Second, approximately 70 % of the estimated values have an error below 20 %.

Apart from the error inherent in the constructed model, the differences between measured and estimated values can be manifold and add bias to the results. The most obvious ones to be considered are (i) the instrument error, (ii) the error added by interpolation and data handling between software, and (iii) the error introduced by natural soil salinity variation or management practices.

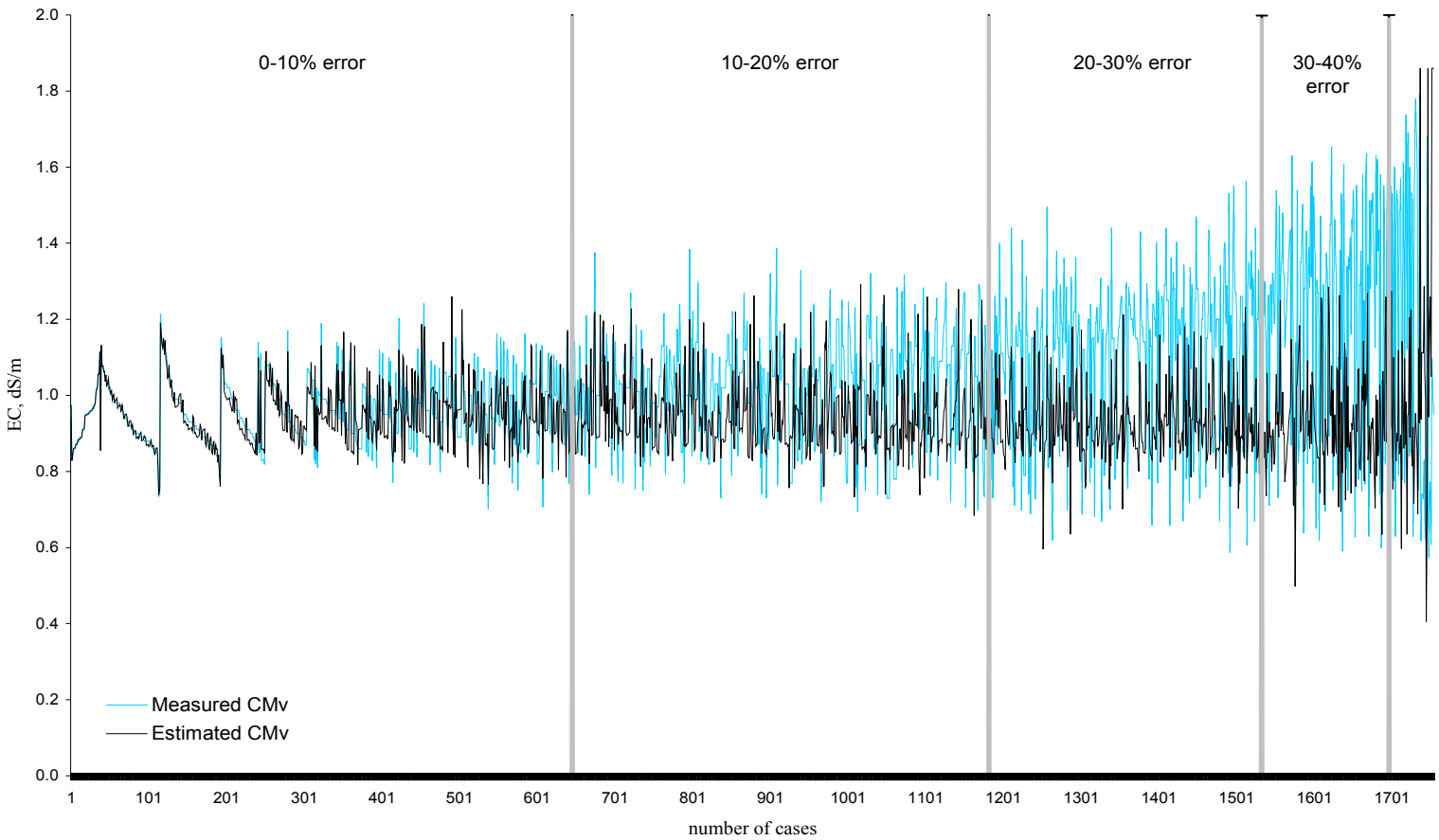

Figure 4.7 Relationship between measured and estimated CMv sorted by percentage error in ascending order. Error zones indicate how many samples are within displayed error zones (e.g., approximately 634 samples (36 % out of total data points) are within 10 % error of the measured values)

As with any measuring device, the instrument used (CM-138) has measurement precision errors although the manual of the CM-138 does not provide the percentage error or any other type of tolerance intervals of the readings. However, under ideal conditions, errors are assumed to be acceptable and not more than 10 % (GF-Instruments, *personal communication*, 2004). However, temperature can have a strong influence, because most of these types of devices are designed to work within a 10-40°C temperature range. Recently, Robinson et al. (2004) observed as much as 20 % 'drift' in the measurement of bulk soil electrical conductivity using the EM-38 due to high ambient temperatures and correction factors used by the equipment. High temperatures (35-45°C) during both sampling surveys in the summers of 2002 and 2003 were normal in this study. The possible overheating of the black-and-orange color tube covers of the device and the extent of heat influence is unknown.

Furthermore, the 'delicate' use of the device is difficult where distant points over large areas have to be sampled. Device failure for different reasons occurred several times, and a second device had to be used on some occasions. This might have also caused some errors, because the comparison of the two available devices showed some deviation in the readings for the same spot in the fields. To avoid this type of variation, the device used from the beginning was used as much as possible. Deviations in readings might also have been caused by poor cable contact resulting from inevitable rough handling.

Interpolation is another source of error. Soil salinity interpolation has been extensively studied, and errors resulting from interpolation itself can be quite substantial. For example, the RMSE for different kriging methods examined by Triantafilis et al. (2001) varied from 1 to 1.3 in conditions very similar to this study area. Errors do not necessarily add up with each interpolation, but the current result is one of the realizations of data used among many other possible realizations.

Influence of soil and irrigation management

The map of estimated soil salinity distribution shows the expected average distribution based on environmental relationships only. The ANN model built can be considered as independent of management, and the only influencing factors are those used in the

model. However, perhaps the differences between estimated and measured maps are related to soil and irrigation management.

To illustrate why soil salinity is affected by management or seasonal variation, the relationship between CMv and CMh in summer and autumn can be used. The simple scatter plots of the CM-138 measurements display the differences in the datasets collected for training at the farm level (Figure 4.8a) and for testing at the district level (Figure 4.8b and c). As mentioned earlier, the soil salinity values at farm level were mainly in the lower range, and the slope of the linear relationship between CMv and CMh passes close to zero. This means that CMv and CMh values are equal, and soil salinity within the soil profile is distributed uniformly in the top 1.5 m of the soil.

At the district level however, measurements taken in the summer show a different picture, and the regression line passing through the points has an intercept of 0.5. This suggests that some of the salinity is located below 75 cm, possibly due to leaching. Summer soil salinity values are elevated compared to autumn data. Statistical tests (paired *t* test and Wilcoxon signed rank test) between electromagnetic measurements taken in summer and autumn showed they are significantly different. Higher salinity values during the summer might be caused by high evapotranspiration due to high temperatures and vegetation growth, bringing groundwater salinity up into the soil and leading to saline irrigation water.

Groundwater fluctuations maybe the cause of poor relationship between the CMv taken in the summer and autumn. Irrigation in autumn is largely non-existent in this region, and the salinity tends to stabilize within the soil profile. A highly dispersed scatter plot (Figure 4.8c) indicates that there is no straightforward relationship between summer and autumn data. Such information is not captured in the ANN model and hard to incorporate without knowledge of the local rotation system. Therefore, further research should address the field water management schemes and their crop rotation schemes. With this information it would be possible to include crops planned to be grown on a particular field and expected water input based on crop water requirements.

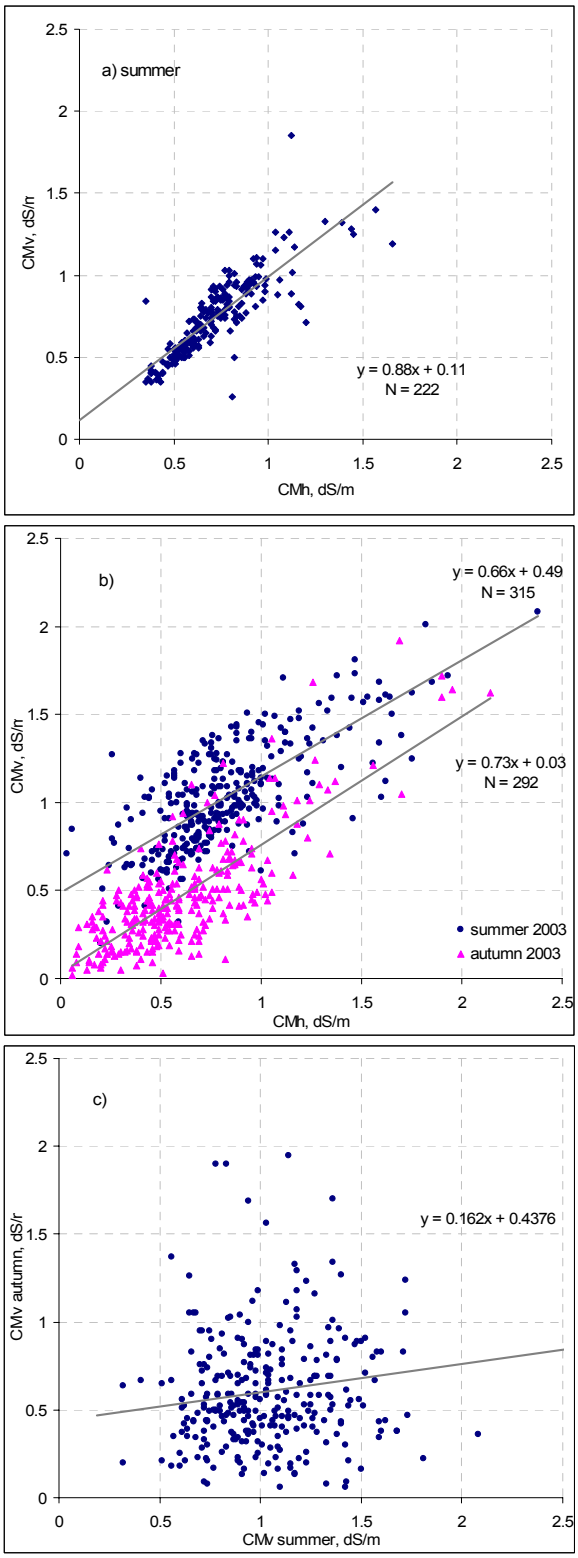

Figure 4.8 CM-138 scatter plots
a) summer scatter plot between CMv and CMh
b) comparison between CMv and CMh scatter plots taken in summer and autumn 2003
c) relationship between CMv taken in summer and autumn

Soil texture map differences

As the model outcome essentially depends on the accuracy of the input data, it was possible to assess the accuracy of the soil texture data that was fed into the ANN model. Soil-electromagnetic conductivity is greatly influenced by the soil texture (more details in section 3.4.3. The utilization of available soil survey data is important, but at the same time the accuracy of the available data is often unknown. For upscaling purposes, this study used the data described in section 4.3.1. During the validation survey, topsoil samples were acquired for a cross-check. Figure 4.9 presents the district map of the topsoil clay content derived from the available soil profile information in 1996 obtained from the Soil Research Institute. It can be seen that there are some areas of the district boundary where the clay content could not be extrapolated. The clay content from the 1996 survey is clearly different from that of the 2003 validation survey. Differences can be mainly attributed to the lack of data points in the 1996 survey and the 'bumpy' pattern resulting from interpolation. On the other hand, the map from the validation survey shows smoothly changing patterns, which is largely due to sufficient data points for interpolation. It is unlikely that there was a change in the clay content between 1996 and 2003 to the extent that these maps would differ as they do.

The differences between the clay content data that was fed into the ANN model and the real clay content distribution demonstrated in Figure 4.9 is considerable, yet the ANN model was able to approximate from such data. Such deviations between the data fed into the model and the actual distribution of the data might be also the case with other attributes (i.e. distance to drains, topography). However, the accuracy of other attributes was not possible to examine and to include in the sampling campaign. Therefore, the comparison based on percentage error is thought to conform to such specifics in this study. In the next section, it is demonstrated what percent of the total estimated area can match the measured area if this method is used.

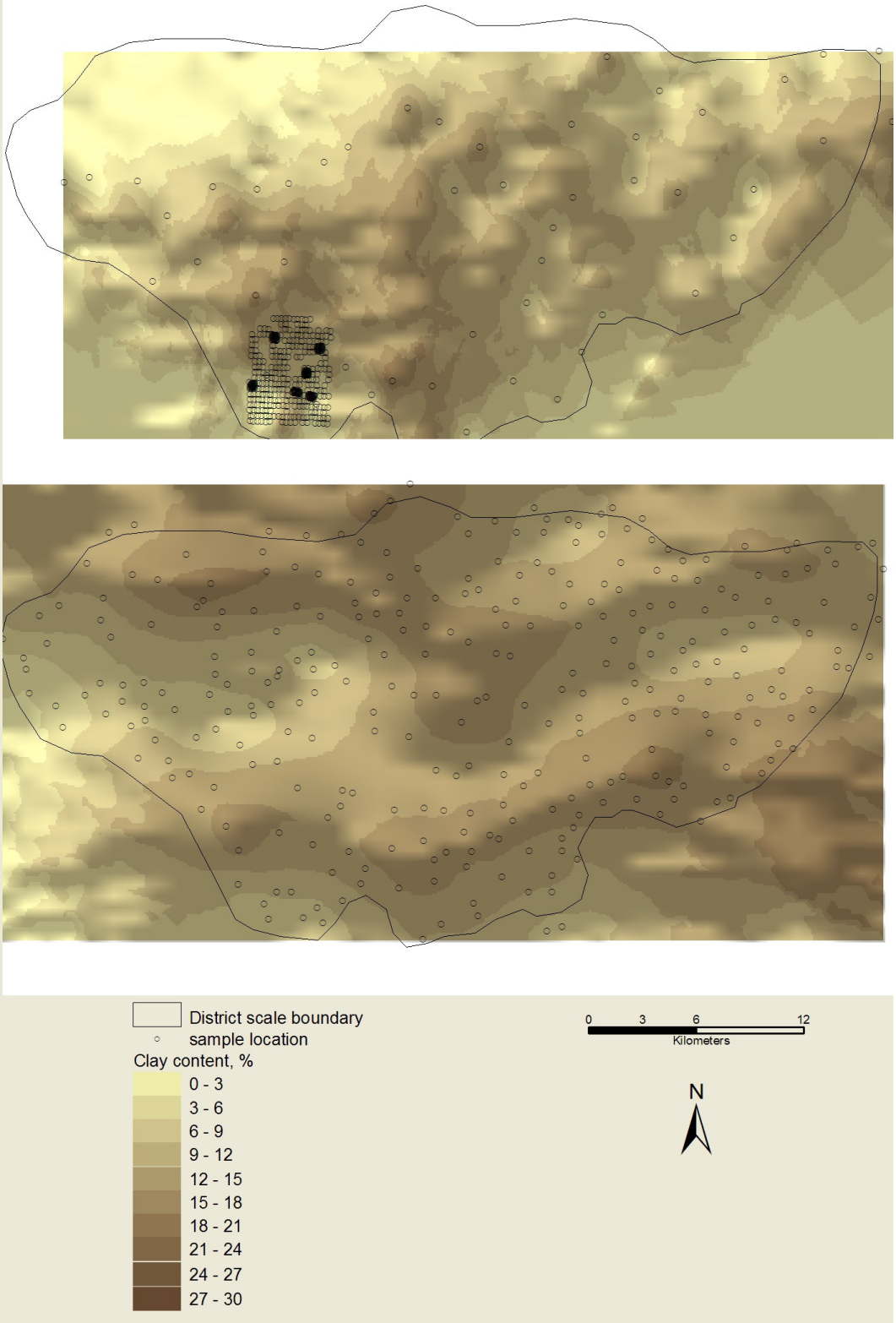

Figure 4.9 District topsoil clay content derived from Soil Research Institute soil survey data 1996 (upper map) and clay content obtained during validation survey in 2003 (lower map)

4.4.4 Error-induced spatial soil salinity distribution

Graphical representation of the deviation between estimated and measured soil salinity values presented in Figure 4.7 indicates that 70 % of the values are within 20 % error. If we consider the extent of errors described in the previous section (4.4.3) in this model is considered, a percentage error of <20 % can be assumed acceptable. Based on this assumption, the areal extent of the matching soil salinity that is considered within the acceptable percentage error range can be mapped. Figure 4.10 presents such maps and a 30 % error map for comparison. Naturally, the matching areas increase as the error tolerance is increased, but the 20 % error map suggests that the ANN assesses soil salinity in a large proportion of the region with acceptable accuracy.

4.4.5 Regression analyses between measured and estimated soil salinity

The correlation coefficient between salinity measured in summer and estimated values is significant but very low (r=0.06). Figure 4.11 presents regression analyses for the data points with 10 and 20 % error and for the total number of data points (n = 1755). Interestingly, when forced through zero, the regression lines in all graphs have a high R^2 (>0.9) with slopes equal to 1. Such a high R^2 is logical outcome since electrical conductivity of 0 dS m^{-1} measured by CMv should equal an estimated 0 dS m^{-1}.

If standard regression analyses are considered without forcing the regression line through zero, a high R^2 (0.6) for the 10% error data points is obtained (Figure 4.11a). From the total number of data points (n = 1755), 634 observation points (36%) are within 10 % error. Naturally, the R^2 drops to 0.2 when the error margin is relaxed to 20 %, but in this case almost 70 % of observations fall within the 20 % error (Figure 4.11b).

Similarly, when the total data points are fitted with least square error method the regression has the lowest R^2 (0.003) with the slope approaching zero and the intercept close to 1 (Figure 4.11). The intercept is essentially the mean estimated CMv.

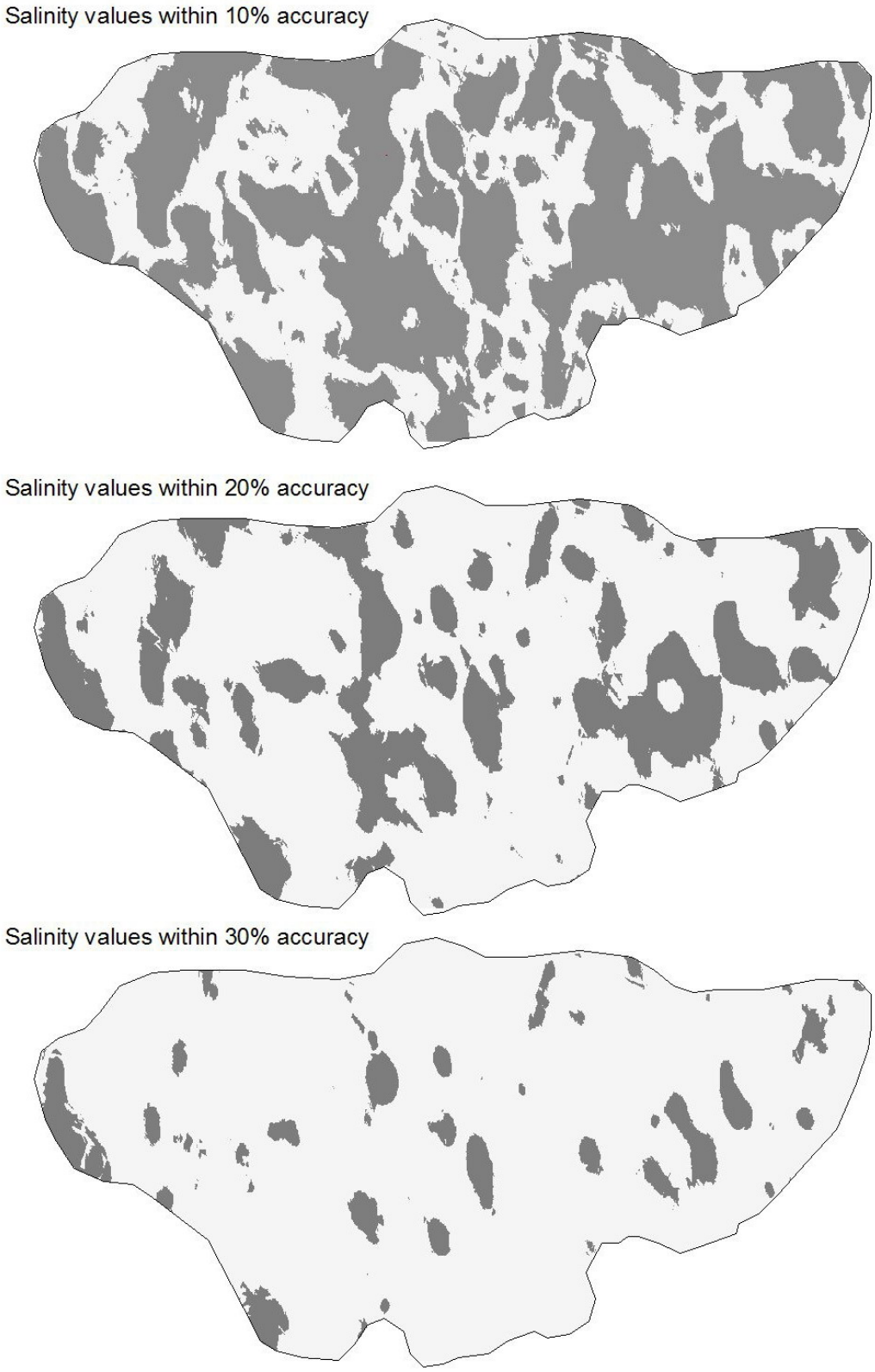

Figure 4.10 Soil salinity at district scale estimated by neural network model. White areas show where estimated values of soil salinity are within 10, 20, and 30 % error range relative to the measured values

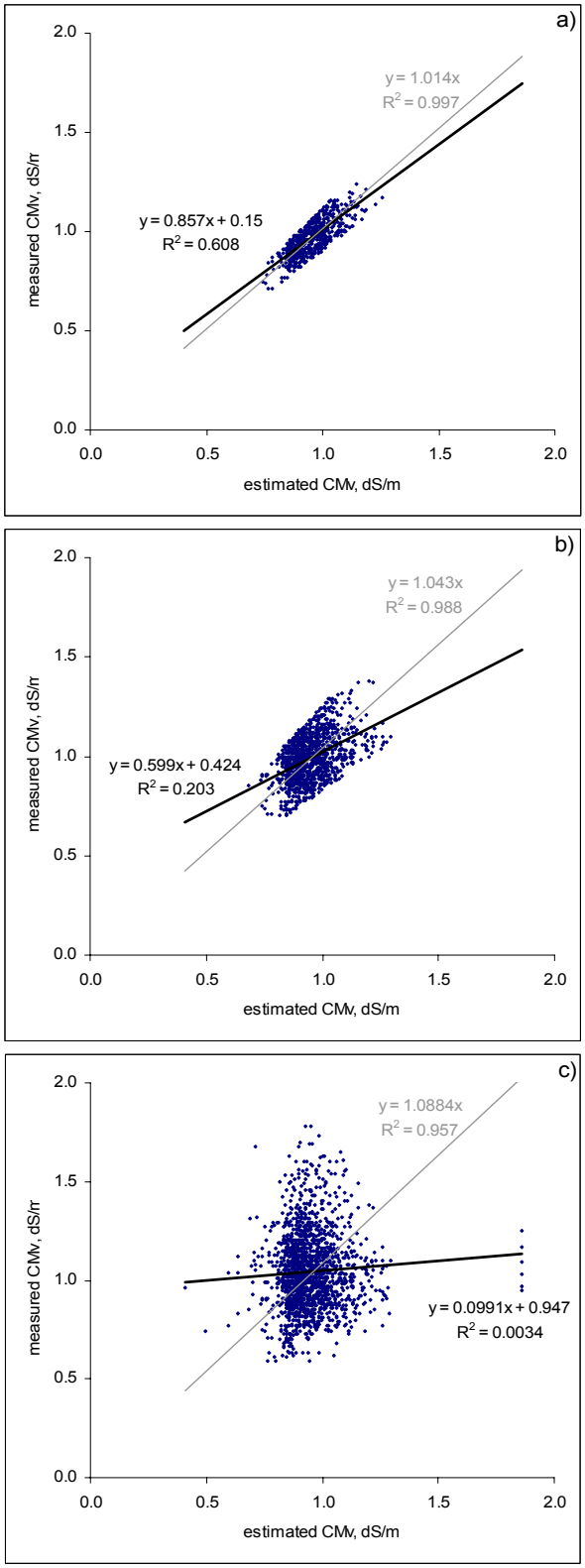

Figure 4.11 Regression analyses between measured and estimated soil salinity
a) with data points within 10% error
b) within 20% error
c) total number of points

4.5 Discussion and conclusions

4.5.1 Modeling approach

The much referenced CLORPT model for soil formation is simplified with the main factors quantified here in the model being relief (terrain attributes) and parent material (soil texture is used as a proxy) and including the man-made drainage network (distance from drains), groundwater information (groundwater depth and salinity), and remote sensing data (band signal values and indices). The modeling procedure has to deal with two main issues, namely change of scales and the complex relationship among the variables involved.

The hierarchical approach described in section 4.2 is implemented and Figure 4.12 visualizes the aggregation of data to estimate soil salinity at the farm scale. The scale is given on the left of the graph to obtain or describe the data given on the axes drawn on the plane. For example, to estimate soil salinity at the farm scale, the soil information from point data (soil structure scale) obtained using K3 (simple model) was used to relate its value to other data that were available on a larger scale. Remote sensing was available at the regional scale, and the remote sensing indices were obtained using K3 and K5 approaches, which apply simple comprehensive models and detailed specialistic models to calculate vegetation indices. Similarly, terrain information was available at a small county scale, and the terrain indices were calculated using K3 and K4 approaches, which is mixture of simple and complex comprehensive models. The use of the comprehensive models is encouraged, since they can be applied to other models and are therefore easy to implement by others.

The neural network training proved to be a very important step forward in modeling, since the approximations generated by the different architectures of the network can be very diverse. Some attempts to try different structures of the network (generalized feed-forward, and multilayer perceptron with varying types of internal functions) did not improve the model, with the default backpropagation method being the most suitable.

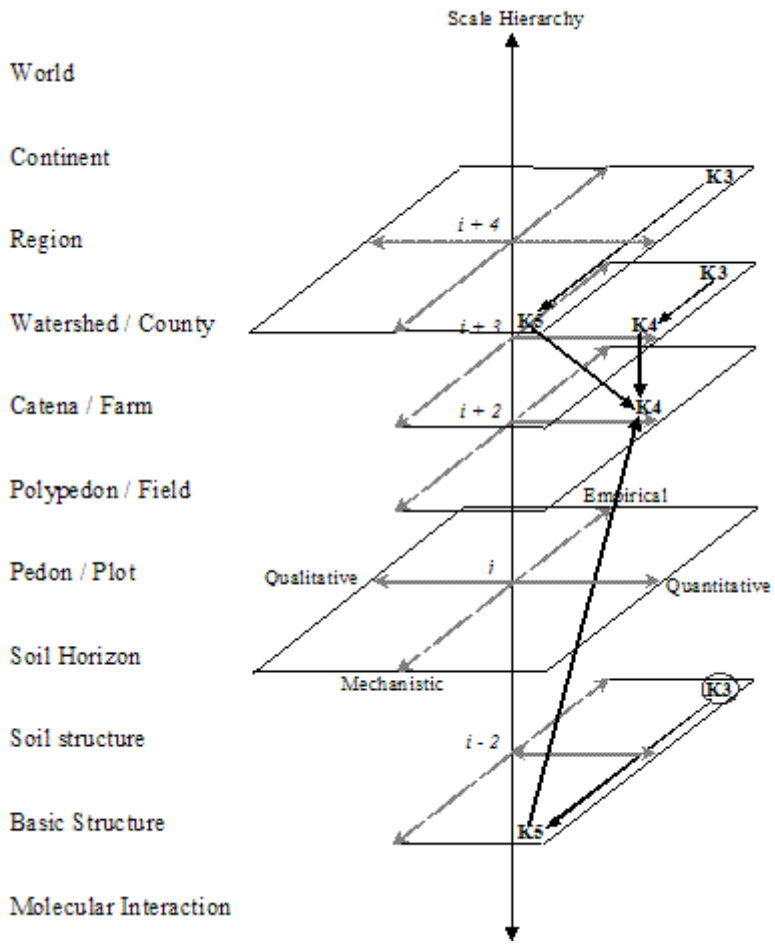

Figure 4.12 Hierarchical approach for Khiva district soil salinity model at farm scale and knowledge levels (after Hoosbeek and Bouma, 1998)

4.5.2 Interpretation of estimated soil salinity

The use of environmental attributes and the soil salinity relationship to upscale spatial distribution of soil salinity from farm to district scale essentially estimated identical mean soil salinity values (0.94 dS m^{-1} vs. 1.04 dS m^{-1}). The neural network model used to spatially estimate soil salinity values at selected grid points was able to approximate soil salinity despite the possible errors in intermediate processing of data. This shows that the model based on the included variables can effectively estimate soil salinity. The major advantage of this approach is that available data is used.

As technology and models are developing rapidly, many methodologies emphasize the flexibility and adeptness to modifications. The model can be fine-tuned by incorporation of other related factors, for instance, geodesic data such as the extent of impermeable layers. By incorporating new correlating factors, further complication

of the model should be avoided. By fine-tuning a "window of opportunity" (Bouma and Hoosbeek, 1996) is rather created for improvement of the model by replacing with either a better explaining factor or a more easily obtainable factor.

Hoosbeek and Bouma (1998) emphasize that to be able to define suitable indicators for certain soil and land qualities, it is important to have insight into the characteristics of an indicator. For soil salinity, variables were deliberately removed and necessary ones selected according to the strict limitations outlined in section 4.2. As previously attempted by many others, the Jenny model was simplified by the assumption that other parameters are constant and was based largely on terrain and texture, and on additional parameters from remote sensing and drainage networks. The emphasis was on terrain attributes because, as Moore et al. (1993) and McKenzie et al. (2000) proved, these provide a much better basis for visualizing other data layers and allow generating sets of data that have value in their own right (e.g., slope and solar radiation), even when the exact form of the relationship is unknown.

It appears that soil salinity variability is mainly due to the spatial variability of curvatures (curv, planc, profc), which tend to control water movement and solar radiation which in turn control the evaporation rate. The neural network was more sensitive to these micro-scale controlling factors than to those that are obvious at a macro-scale. As Yair (1990) observed, soil or sediment-covered areas having relatively high absorption capacities will experience reduced runoff, shallow infiltration and decreased water availability for leaching, and this leads over time to salt accumulation at shallow depths. Similarly, in the study region, finer textured soils have a generally higher salt content than coarser textured soils, and landform attributes become important as local controlling factors. In fact, most of the areas where estimated and measured soil salinity values do not match are the areas where the soil texture data from the SRI data and the validation survey show great dissimilarities, or where the clay content could not be properly extrapolated.

The use of a non-spatial neural network approach presents a problem in the comparison of results (Laffan, 1999). Non-spatial tools do not provide truly spatial error measures. Therefore, it is important to assess their spatial performance in order to identify local areas of acceptable prediction. Laffan (1999) addressed this problem and has recently proposed geographically weighted regression with increasing circle size to

detect errors at different scales. There is a need for elaborating assessment techniques to determine the reason for the occasional spurious effects on generated maps.

This study demonstrates that the ANN approach is suitable for upscaling purposes, and results are valid within a given error range of soil salinity. Timing of measurement is important, and as the study was conducted during the summer period, further application of this model to other districts of the Khorezm region and validation should involve the summer period only. Application of the results to other seasons must be carefully assessed before making any generalizations about soil salinity distribution using the model presented here.

The influence of soil and irrigation management on soil salinity is a dynamic feature and should be included in models of soil salinity distribution. Capturing and parameterizing this dynamic feature into a model could potentially improve model performance. Therefore, further research is needed to quantify these parameters, which can then be included in the further development of such models.

5 GENERAL DISCUSSION AND CONCLUSIONS

An increasing number of studies have shown that soil properties can be estimated in the field by devices or correlated with environmental attributes. Soil salinity and other soil parameters were determined on a regular grid (150 x 200 m), with 6 nested grid sampling points with a finer grid size (40 m x 40 m), over the study area of 3 x 4 km in the Khorezm region, Uzbekistan. The main purpose of the study was to model soil salinity distribution at the farm scale and to upscale to a district scale based on environmental attributes with an emphasis on readily available and cheaply obtainable data. The sampling design ensured inclusion of the environmental variables and soil salinity variation, and in contrast to conventional survey methods, the sampling is repeatable.

The following discussion elaborates the approach and its success in achieving the objective. To structure the discussion, soil salinity estimations are examined from the viewpoint of scales, starting from point scale and proceeding to district scale.

5.1 Quick soil salinity estimates at small scales

The availability of measuring devices based on electrical conductivity offers the possibility for quick estimation of soil salinity directly in the field. Laboratory methods of salinity assessments based on TDS measurements can be replaced by conductivity measurements in the soil paste (EC_p) because with these, the TDS could be estimated with sufficient accuracy (R^2=0.76). The locally made instrument (2XP) appears suitable. Alternatively, the concentration of individual salt ions could be estimated by measuring the electrical conductivity of the paste, because EC_p and salt ions were highly correlated. The only practical disadvantage of the EC_p-based estimation of soil salinity is the need for distilled water to prepare the soil paste.

Field estimates of soil salinity can also be done using EC-based sensors that measure bulk soil electrical conductivity (EC_a). All the examined devices based on this approach (2P, 4P, and CM-138) showed similarly low correlations when regressed against EC_a calculated using the Rhoades model ($R^2 < 0.5$). However, the experimental results showed that the result of any EC_a measurement greatly depends on the soil moisture content (see section 2.4). The low accuracy of the regression of EC_a-based

measurements against TDS can be the result of low soil moisture content. Therefore, to avoid such discrepancies, EC_a-based salinity assessments should be conducted when soil moisture is adequate, for instance, when soils are at field capacity. Under such conditions, the instrument-based soil salinity assessment appeared to be a viable alternative to conventional laboratory-based TDS measurements, because it provided quick estimation in the field.

The assessment of different electrical conductivity devices, although a minor part of this study, is of great practical value. They are easy to use and offer opportunities for easy agricultural extension, because simple EC devices can be locally assembled. For example, the comparison of the locally made 2-probe with the commercial 4-probe electrical conductivity meter in the bucket experiment showed that the readings of both probes were similar.

One needs to recognize that the suitability of these devices is different for different scales. Using the handheld devices, soil samples can be as small as 20 grams to estimate salinity (EC in the soil paste) by 2XP. This is convenient for micro-scale assessment of soil salinity.

Under field conditions salinity at a certain depth is sometimes of interest, for example, when the crop is established and soil salinity of the subsoil 20-40 cm layer is unknown. In such cases, the 2P and 4P devices are most suitable because they provide both point and depth measurements. Point measurements are also useful for assessing soil salinity over small plots.

CM-138 measurements, on the other hand, offer a vital advantage over other instrument-based salinity assessments, because they provide information on the average salt content in the soil profile. Readings of the CM-138 are less affected by topsoil disturbances, allow much quicker, intensive, and on-the-go measurements. Electromagnetic-based devices are the most advantageous, especially at larger scales, when there is a need to map fields or larger areas.

5.2 Farm and district scale soil-salinity estimates using geostatistics

While CM-138 measurements offer easier and faster mapping of soil salinity over large areas than other methods, the coverage of a whole farm or even district is still cumbersome. The presence of a spatial structure (see Figure 3.6) in soil salinity data

suggests that the sampling can be done with larger distances between sampling points (around 300-400 m), so that accurate interpolation can still be successfully accomplished. In the absence of previous salinity surveys, the spatial variability of environmental explanatory variables that are easier to measure, such as soil texture or topography, can be used to infer soil salinity. As a result, the area can be mapped with less effort without losing high accuracy and resolution.

Emphasis was placed here on readily and cheaply obtainable data, because there is a wealth of data already accumulated by local organizations that could be used for this analysis. Many studies using soil-landscape analysis showed that terrain alone can be a very good predictor of the spatial distribution of soil properties. This study demonstrated that at the farm scale there are relationships between selected environmental attributes and soil salinity, measured by the CM-138 in particular. The relationships were demonstrated by the ANOVA analysis for soil texture and salinity, and correlation analysis between the remaining environmental attributes and soil salinity.

While it is generally recognized that soil texture and topographic position exert a large influence on soil salinity, the tools that prove this relationship can be inadequate or 'blind'. In this study, the standard statistical approach (ANOVA and correlation coefficient) showed the influence of soil texture and remote sensing indices on soil salinity, but was unable to confirm the influence of topography.

On the other hand, the neural network sensitivity analysis was sensitive to terrain attributes and to curvatures in particular. These kinds of discrepancies among statistical methods might be common, and therefore a proper understanding of the strengths of the applied methods is essential. However, an explanation of the reasons for the behavior of the statistical methods is beyond the scope of this discussion. The apparent conclusion from this example is as follows: The standard statistical method was able to "see" the influence of soil texture and remote sensing indices because they exert their influence on a large scale. For example, the soil salinity gradually decreased from heavier soils to sandy soils. This also affected remote sensing indices as sandier soils had less vegetation. A neural network, in contrast, was able to distinguish the influence of the curvatures because curvatures determine the shape of the location, be it concave or convex. The effect perhaps can be seen best in Figure 3.8, where a majority

of the convex areas have low salinity values. Thus, the standard statistical method was able to discern farm or district level relationship of soil salinity with environmental attributes, whereas the neural networks were better at the local scales. The multi-scale influence of the environmental attributes is in accordance with field experience.

Taking modeling a step further, the neural network model was built to estimate soil salinity (measured by the CM-138) at farm scale based solely on environmental attributes. The use of this model to estimate soil salinity for the district produced soil salinity estimates comparable to the measured values.

District-scale soil salinity estimation was successful, but the visual comparison of estimated and measured soil salinity maps was not immediately convincing. The ability of the neural network model to estimate salinity was acceptable only with an error tolerance of ~20%, as detailed in section 4.4.3. The 20% tolerance error image (Figure 4.10) of the estimated soil salinity is acceptable because it is as near-realistic a map as one can get with the existing data. Mismatches greater than 20 % appeared to be clustered on this map, suggesting that some factor playing an important role in determining salinity was not properly captured in the model.

5.3 Implications of the study and scope for further research

This study concentrated on the application of scientific concepts and methods from established (i.e., soil science) and emerging fields (i.e., pedometrics) to provide alternatives to conventional surveys of soil salinity. This is of paramount importance in the study region because at the moment, real salinity data are rarely used in decision making both at the farm and the policy-making levels, due to difficulties in their determination. A maximum effort was made to make the methods practically applicable by various local organizations engaged in soil salinity monitoring, mapping, or reclamation.

Soil salinity estimates can be based both on electrical conductivity of the paste (EC_p) with high accuracy and on the salinity assessments in bulk soil (EC_a) provided that the soil moisture conditions are close to field capacity. The use of EC-based devices together with geostatistical analysis substantially reduces the number of soil samples required by traditional methods to accurately and intensively map the spatial distribution of soil salinity over large areas. Such capability will of course lower the

costs of such assessments and therefore allow monitoring at a higher frequency of spatial scales and shorter time intervals.

Environmental correlation of attributes derived from terrain and from remote sensing, i.e., soil texture, groundwater level and salinity, and distance to the drains, can be used to estimate the spatial distribution of soil salinity in the districts. These environmental attributes are available and often exist as high resolution maps in the region. For higher accuracy, additional parameters may need to be taken into account.

There are also limitations to this approach, which can be divided into three major categories. The first is the transfer of paper-based maps and data into digital form. This is the case when data and maps are available only as hard copies for several reasons, such as late computerization or lack of an established infrastructure for digital processing. The second is the need for specialized skills in order to process data and understand the concepts on which this study is based. The third is the uncertainty of the available data, in spite of the fact that local organizations often possess huge databases (on paper) of data which are routinely collected. However, in some cases they can be termed as 'data rich, information poor' for various reasons (ill-conceived design, insufficient density of observation points, etc.) which result in limited applicability.

Although this study attempted to include the land management factor (here in the form of crop maps), the relation of this factor with soil salinity or other variables was not satisfactory. Yet, as has already been mentioned earlier in the text, it is believed that the soil salinity model will greatly benefit from the inclusion of a management factor. There is, however, a need for easy parameterization techniques for the management factor so that it can be quickly included into the model.

5.4 Conclusions

The chosen approach was shown to be adequate for the task of estimating soil salinity considering that (i) the sampling design ensured inclusion of the important and easy-to-use environmental variables and sufficiently accounted for soil salinity variation, (ii) in contrast to conventional survey methods, the sampling is repeatable, and (iii) environmental correlation based mainly on terrain-derived parameters and soil texture is the general method for the spatial estimation of soil salinity. However, tools to account for such correlation need to be properly selected, because standard correlation or

regression tools do not necessarily discard unrelated variables only but often could not 'see' the expected relationship between variables (for example, low correlation of terrain attributes; however, neural network sensitivity analysis on the other hand was able to distinguish terrain attributes).

Quick techniques for soil salinity determination based on electrical conductivity were assessed and proved to be satisfactory in all cases. Measurement based on soil paste (EC_p) was highly accurate (R^2=0.76), whereas EC_a measurements at a point scale in the field were of low accuracy (R^2<0.5). However, field assessment of soil salinity was considerably enhanced by the use of the CM-138 because large areas can be quickly assessed, which, in spite of lower accuracy is desirable. Accurate readings of the EC_a (using the 2P, the 4P, or the CM-138) are possible provided that the soil moisture conditions are close to field capacity. Such estimates would mainly refer to the field scale and the conditions close to the optimum soil moisture content would be a few days after the irrigation event.

Characterization of the spatial variability of soil salinity at farm scale revealed that:

i) Topsoil (30 cm) salinity was highly variable even at short distances (40 m) compared to average soil salinity down to 0.75 m and 1.5 m depth, measured by CM-138. Therefore, environmental correlation of soil salinity should be conducted with a soil salinity measured with larger volume and at greater depth because, as CM-138 measurements showed, these offer a more stable spatial structure.
ii) Overall distribution of soil salinity was influenced by soil texture and topography, while at the local scale, terrain attributes such as curvature, plan and profile curvatures, and solar radiation were the more important factors. The inclusion of these soil salinity controlling variables in modeling is fundamental, and efforts must be directed towards obtaining reliable and accurate databases to acquire these variables.

The following variables used in the modeling of the spatial distribution of soil salinity enhanced the model performance:

iii) Factors obtained from remote sensing had significant correlation coefficients with both salinity of topsoil and measured by the CM-138. Since band signals and calculated indices are mainly an indication of vegetation, the correlation suggests salinity affects crop growth significantly and can be used as an indicator. In contrast, crop type did not influence soil salinity significantly. Possible reasons were that only two (cotton and wheat) main crops are grown in the study area, and differences between fields under cotton or wheat were not significant. This suggests that vegetation per se can be used as an indicator but inclusion of the crop types in the prediction model needs further research.

iv) Distance to drains is an important factor, especially for the bulk soil salinity (measured by CM-138) of the profile. It was lower for the topsoil (TDS and Cl analyses), which might be due to higher spatial variation of the topsoil salinity.

v) Groundwater table depth and salinity had high correlation with soil salinity; however, the direction of the influence could not be explained. The lack of groundwater observation wells for the study area, low reliability of the data, and errors in interpolating the data might account for this anomaly.

Soil salinity estimation combining remote sensing data with the available soil database in Khorezm:

vi) The environmental correlation model built for the farm scale estimated soil salinity using a neural network with a correlation coefficient between estimated and measured soil salinity of 0.83. Explicit prediction of salinity was satisfactory taking into account contrasting scales of soil salinity variation and environmental data derived from varying scales and unknown accuracy.

vii) The use of environmental attributes and soil salinity relationship to upscale spatial distribution of soil salinity from farm to district scale essentially estimated identical mean soil salinity values (0.94 dS m^{-1} vs. 1.04 dS m^{-1}). The ANN map coincided better with the measured salinity map as the error tolerance was relaxed to 20 %. Mismatches with more than 20 % error appear clustered

and suggest that some additional factors may need to be included in the neural network model, possibly by training over a wider area.

In general, the results indicate that significant correlations between quantifiable terrain attributes and soil salinity exist. This study shows that such a relationship was successfully used to estimate soil salinity at farm scale solely based on quantified environmental variables. The upscaling proved to be satisfactory, but further research is needed before wider application of the model. Considering that the neural network model is an empirical model, further training of the model is required where locations and conditions are different to those on the farm where the model was generated.

Furthermore, the model depends on the strength of the relationship between environmental variables and soil salinity. Therefore, the environmental variables must be available for the study area in high resolution or easily measurable.

6 REFERENCES

Abdullaev U. 2002. Uzbekistan country report. [Online]. Available at http://www.fao.org/ag/agl/swlwpnr/reports/y_nr/z_uz/uz.htm#referen (accessed January 13, 2004; verified June 21, 2004). FAO Land and Water Development Division.

Aladin N.V. 1998. Some palaeolimnological reconstruction and history of the Aral Sea basin and its catchment area. In: Ecological research and monitoring of the Aral Sea Deltas: A basis for restoration. UNESCO Aral Sea Project 1992-1996 Final Scientific Reports, pp. 3-12.

Anonymous. 1963. Methods of agrochemical, agrophysical, and microbiological analyses in irrigated cotton areas, Tashkent.

Anonymous. 1977. Metody agrokhimicheskih analyzov pochv i rasteniy Sredney Azii (Agrochemical methods of analysis of soils and plants in Central Asia), 5th ed, Tashkent, pp. 173-180.

Archie G.E. 1942. The electrical resistivity log as an aid to determining some reservoir characteristics. Transactions of the AIME 146:54-63.

Bell J.C., Grigal D.F. and Bates P.C. 2000. A soil-terrain model for estimating spatial patterns of soil organic carbon. In: Wilson J. P. and Gallant J. C. (eds) Terrain Analysis: Principles and Applications. John Wiley and Sons, New York, pp. 295-310.

Bouma J., and Hoosbeek M.R. 1996. The contribution and importance of soil scientists in interdisciplinary studies dealing with land. In: Wagenet R. J. and Bouma J. (eds) The role of soil science in interdisciplinary research, Vol. 45. Soil Science Soc Amer Spec Publ, pp. 1-15.

Bouma J., Finke P.A., Hoosbeek M.R. and Breeuwsma A. 1998. Soil and water quality at different scales: concepts, challenges, conclusions and recommendations. Nutrient Cycling in Agroecosystems 50:5-11.

Bowden G.J., Maier H.R. and Dandy G.C. 2002. Optimal division of data for neural network models in water resources applications. Water Resources Research 38:1-11.

Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1984. Classification and regression trees Wadsworth, Belmont, California.

Burt T.P. and Park S.J. 1999. The distribution of solute processes on an acid hill-slope and the delivery of solutes to a stream: I. Exchangeable bases. Earth Surface Processes and Landforms 24:781-797.

Cherkassky V. and Mulier F. 1998. Learning from data: concepts, theory, and methods John Wiley and Sons, New York.

Chernishev A.K. 2004. X-Express conductometer. Instructions on the use, calibration, and measurements, Agromeliotaraqqiyot, Tashkent, 24pp.

Chub V.E. and Myagkov S.V. 1999. A model of the flood transformation in the Amudarya River, In: Ellis B. (ed) Impacts of Urban Growth on Surface Water and Groundwater Quality. Proceedings of IUGG 99 Symposium HS5, Vol. 259. IAHS, Birmingham.

Corwin D.L. and Rhoades J.D. 1982. An improved technique for determining soil electrical conductivity-depth relations from above-ground electromagnetic measurements. Soil Science Society of America Journal 46:517-520.

Corwin D.L. and Hendrickx J.M.H. 2002. Indirect measurement of solute concentration. Electrical resistivity - Wenner array, In: Dane J. H. and Topp G. C. (eds) Methods of soil analysis, Vol. 4. SSSA Book Series 5, Madison, WI, pp. 1282-1287.

Corwin D.L. and Lesch S.M. 2003. Application of Soil Electrical Conductivity to Precision Agriculture: Theory, Principles, and Guidelines. Agron J 95:455-471.

Curmi P., Durand P., Gascuel-Odoux C., Merot P., Walter C. and Taha A. 1998. Hydromorphic soils, hydrology and water quality: spatial distribution and functional modelling at different scales. Nutrient Cycling in Agroecosystems 50:127-142.

Dalton F.N., Herkelath W.N., Rawlins D.S. and Rhoades J.D. 1984. Time domain reflectometry: Simultaneous measurement of soil water content and electrical conductivity with a single probe. Science 224:989-990.

Department for Natural Resources. 2004. DNR random sample generator. Release 2.2. Department for Natural Resources, Minnesota.

Eklund P.W., Kirkby S.D. and Salim A. 1998. Data mining and soil salinity analysis. International Journal of Geographical Information Science 12:247-268.

Evans F.H. and Caccetta P.A. 2000. Broad-scale spatial prediction of areas at risk from dryland salinity. Cartography 29:33-40.

FAO 2003. Fertilizer use by crop in Uzbekistan. FAO Land and Water Development Division, Rome.

Florinsky I.V., Eilers R.G. and Lelyk G.W. 2000. Prediction of soil salinity risk by digital terrain modelling in the Canadian prairies. Canadian Journal of Soil Science 80:455-463.

Gessler P.E., McKenzie N.J. and Hutchinson M. 1996. Progress in soil-landscape modelling and spatial prediction of soil attributes for environmental models Proceedings of Third International Conference/Workshop on Intergrating GIS and Environmental Modeling. National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA, Santa Fe, NM.

Glantz M. 2002. Water, climate, and development issues in the Amudarya Basin. Report of informal planning meeting held 18-19 June 2002 in Philadelphia, Pennsylvania. Environmental and Societal Impacts Group, NCAR, Boulder, CO.

Goossens R., de Vliegher B.M., Ghabour T.K., and Gad A. 1996. Monitoring, simulation and prediction of soil salinity and waterlogging risk using remote sensing and GIS on provincial level (Ismailia, Egypt). In: Spiteri A. (ed) Remote Sensing '96. Integrated applications for risk assessment and disaster prevention for the Mediterranean. 16th EARSeL Symposium, Malta.

Goossens R., van Ranst E., Ghabour T.K. and El Badawi M. 1998. The use of remote sensing and GIS to detect gypsiferous soils in the Ismailia Province, Egypt. In: Nieuwenhuis G. J. A. et al. (eds) Operational Remote Sensing for Sustainable Development. Proceedings of the 18th EARSeL Symposium, Enschede, Netherlands.

Guitjens J.C. and Hanson B.R. 1990. Spatial and temporal variabilities in salinity. In: Tanji K. K. (ed) Agricultural Salinity Assessment and Management, Vol. 71. American Society of Civil Engineers, New York, pp. 327-348.

Hall L.M., Brainard J.R., Bowman R.S., and Hendrickx J.M.H. Determination of solute distributions in the vadose using downhole electromagnetic induction (submitted).

Hanson B.R. and Grattan S.R. 1990. Field sampling of soil, water, and plants. In: Tanji K. K. (ed) Agricultural Salinity Assessment and Management, Vol. 71. American Society of Civil Engineers, New York, pp. 186-200.

Hassan H.M., Warrick A.W. and Amoozegar-Fard A. 1983. Sampling volume effects on determining salt in a soil profile. Soil Science Society of America Journal 47:1265-1267.

Hendrickx J.M.H., Baerends B., Raza Z.I., Sadig M. and Chaudhry A.M. 1992. Soil salinity assessment by electromagnetic induction of irrigated land. Soil Science Society of America Journal 56:1933-1941.

Hendrickx J.M.H., Das B., Corwin D.L., Wraith J.M. and Kachanoski R.G. 2002. Relationship between soil water solute concentration and apparent soil electrical conductivity, In: Dane J. H. and Topp G. C. (eds) Methods of soil analysis, Vol. 4. SSSA Book Series 5, Madison, WI, pp. 1275-1282.

Heuvelink G.B.M. 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. Nutrient Cycling in Agroecosystems 50:255-264.

Heuvelink G.B.M. and Pebesma E.J. 1999. Spatial aggregation and soil process modelling. Geoderma 89:47-65.

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.

Hillel D. 2000. Salinity management for sustainable irrigation. Integrating science, environment, and economics, World Bank.

Hoosbeek M.R. and Bryant R.B. 1992. Towards the quantitative modeling of pedogenesis - a review. Geoderma 55:183-210.

Hoosbeek M.R. and Bouma J. 1998. Obtaining soil and land quality indicators using research chains and geostatistical methods. Nutrient Cycling in Agroecosystems 50:35-50.

Ibrakhimov M. 2004. Spatial and temporal dynamics of groundwater table and salinity in Khorezm (Aral Sea Basin), Uzbekistan. Ph.D. Thesis. Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany.

Jansen M.J.W. 1998. Prediction error through modelling concepts and uncertainty from basic data. Nutrient Cycling in Agroecosystems 50:247-253.

Johnston K., ver Hoef J.M., Krivoruchko K. and Lucas N. 2001. Using ArcGIS Geostatistical Analyst, ESRI Inc., USA.

Johnston M.A., Savage M.J., Moolman J.H., and du Plessis H.M. 1997. Evaluation of calibration methods for interpreting soil salinity from electromagnetic induction measurements. Soil Science Society of America Journal 61:1627-1633.

Joreskog K.G. 1999. How large can a standardized coefficient be? [Online]. Available at http://www.ssicentral.com/lisrel/column2.htm (verified December 21, 2004). Scientific Software International.

Jurinak J.J. 1990. The chemistry of salt-affected soils and waters: In: Tanji K. K. (ed) Agricultural Salinity Assessment and Management, Vol. 71. American Society of Civil Engineers, New York, pp. 42-63.

Kachanoski R.G., Rolston D.E., and de Jong E. 1985. Spatial and spectral relationships of soil properties and microtopography: I. Density and thickness of *A* horizon. Soil Science Society of America Journal 49:804-812.

Kachinsky N.A. 1958. Mehanicheskiy i microagregatniy sostav pochvi. Metodi ego izucheniya AN SSSR, Moscow.

Karavanova E.I. 2001. Spectral reflectance as a tool to study soils in semi-arid regions, In: Stott D. E. et al. (eds) Sustaining the global farm. Selected papers from the 10th International Soil Conservation Organization Meeting held May 24-29, 1999 at Purdue University, West Lafayette, IN, pp. 217-222.

Karavanova E.I., Shrestha D.P., and Orlov D.S. 2001. Application of remote sensing techniques for the study of soil salinity in semi-arid Uzbekistan, In: Bridges E. M. et al. (eds) Response to Land Degradation. Oxford & IBH Publishing Co. Pvt. Ltd, pp. 261-273.

Kaurichev I.S. (ed.) 1989. Pochvovedenie, Agropromizdat, Moscow, 719pp.

Kust G.S. 1997. The main tendencies in the dynamics of soil salinity within delta areas of the Amu Darya and Syr Darya rivers in conditions of active modern desertification. Eurasian Soil Science 30.

Laffan S.W. 1999. Spatially assessing model error using geographically weighted regression GeoComputation 99. Proceedings of the 4th International Conference on GeoComputation, Virginia.

Landon J.R. (ed) 1984. Booker tropical soil manual. Booker Agriculture International Limited, Bath, 441pp.

Leij F.J., Romano N., Palladino M., Schaap M.G. and Coppola A. 2004. Topographic attributes to predict soil hydraulic properties along a hillslope transect. Water Resources Research 40:W02407.

Lesch S.M., Rhoades J.D., Lund L.J. and Corwin D.L. 1992. Mapping soil salinity using calibrated electromagnetic measurements. Soil Science Society of America Journal 56:540-548.

Li C.B. 1997. Electric conductance, In: Yu T. R. (ed) Chemistry of variable charge soils. Oxford University Press, New York, pp. 272-307.

Maas E.V. and Hoffman G.J. 1977. Crop salt tolerance - current assessment. Irrigation and Drainage Division, ASCE 103:115-134.

Maier H.R. and Dandy G.C. 1999. Empirical comparison of various methods for training feed-forward neural networks for salinity forecasting. Water Resources Research 35:2591-2596.

Martius C., Silva E.G., Garcia M.V.B. and de Morais J.W. 2004. Environmental parameters control soil-and litter-inhabiting termites in rain forest and agroforestry systems in central Amazonia (unpublished).

McBratney A.B., Odeh I.O.A., Bishop T.F.A., Dunbar M.S. and Shatar T.M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97:293-327.

McKenzie N.J. and Ryan P.J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89:67-94.

McKenzie N.J., Gessler P.E., Ryan P.J. and O'Connell D.A. 2000. The role of terrain analysis in soil mapping. In: Wilson J. P. and Gallant J. C. (eds) Terrain Analysis: Principles and Applications. John Wiley and Sons, New York, pp. 245-265.

McNeill J.D. 1980. Electromagnetic terrain conductivity measurement at low induction numbers Techincal Note TN-6. Geonics Limited, Mississauga, ON, Canada.

Micklin P.P. 1988. Dessication of the Aral Sea: A water management disaster in the Soviet Union. Science 241:1170-1176.

Milne G. 1935. Some suggested units of classification and mapping, particularly for East African soils. Soil Research 4:183-198.

Moore I.D., Grayson R.B. and Ladson A.R. 1993. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. In: Beven K. J. and Moore I. D. (eds) Terrain analysis and distributed modelling in hydrology. Wiley, Chichester, pp. 7-34.

Moore I.D., Gessler P.E., Nielsen G.A. and Peterson G.A. 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57:443-452.

Mualem, Y., and S.P. Friedman. 1991. Theoretical prediction of electrical conductivity in saturated and unsaturated soil. Water Resources Research 27:2771-2777.

Mulla D.J. 1988. Estimating spatial patterns in water content, matric suction, and hydraulic conductivity. Soil Science Society of America Journal 52:1547-1553.

Mulla D.J. and McBratney A.B. 2000. Soil spatial variability. In: Sumner M. E. (ed) Handbook of soil science. CRC Press LLC, Boca Raton.

Myagkov S.V. 1995. A model of water and salt exchange between a river and groundwater. In: IAHS (ed) Proceedings of a Boulder Symposium Tracer Technologies for Hydrological Systems, Vol. 229. IAHS Publications.

Nielsen D.R., Tillotson P.M. and Vieira S.R. 1983. Analyzing field-measured soil-water properties. Agricultural Water Management 6:93-109.

Nurmanov A.N. 1966. Salinization and waterlogging of land in the Amu-Darya River Delta, Nukus, Karakalpakstan.

Odeh I.O.A., McBratney A.B. and Chittleborough D.J. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197-214.

Odeh I.O.A., Todd A.J., Triantafilis J. and McBratney A.B. 1998. Status and trends of soil salinity at different scales: the case for the irrigated cotton growing region of eastern Australia. Nutrient Cycling in Agroecosystems 50:99-107.

Pariente S. 2001. Soluble salts dynamics in the soil under different climatic conditions. Catena 43:307-321.

Park S.J. and Burt T.P. 2002. Identification and Characterization of Pedogeomorphological Processes on a Hillslope. Soil Science Society of America Journal 66:1897-1910.

Park S.J. and Vlek P.L.G. 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. Geoderma 109:117-140.

Park S.J. and van de Giesen N. 2004. Delineation of soil-landscape units along hillslopes to identify the spatial domains of hydrological processes. Journal of Hydrology 295:28-46.

Park S.J., McSweeney K. and Lowery B. 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103:249-272.

Park S.J., Ruecker G.R., Agyare W.A., Akramhanov A. and Vlek P.L.G. 2004. Sensitivity of spatial resolution and flow routing algorithm to predict the spatial distribution of soil properties (Geoderma:accepted).

Patel R.M., Prasher S.O., Goel P.K. and Bassi R. 2002. Soil salinity prediction using artificial neural networks. Journal of the American Water Resources Association 38:91-100.

Persson M., Sivakumar B., Berndtsson R., Jacobsen O.H. and Schjonning P. 2002. Predicting the Dielectric Constant-Water Content Relationship Using Artificial Neural Networks. Soil Sci Soc Am J 66:1424-1429.

Petersen G.W., Bell J.C., McSweeney K., Nielsen G.A. and Robert P.C. 1995. Geographic Information Systems in agronomy. Advances in Agronomy 55:67-111.

Popov V.G., Sektimenko V.E. and Tursunov A.A. 1992. Izmenenie pochevennogo pokrova sovremennoy delti Amudaryi pod vliyaniem antropogennogo opustynivaniya Fan, Tashkent.

Principe J.C., Euliano N.R. and Lefebvre C.W. 2000. Neural and adaptive systems: fundamentals through simulations John Wiley & Sons, New York.

Rhoades J.D. and Ingvalson R.D. 1971. Determining salinity in field soils with soil resistance measurements. Soil Science Society of America Journal 35:54-60.

Rhoades J.D. and van Schilfgaarde J. 1976. An electrical conductivity probe for determining soil salinity. Soil Science Society of America Journal 40:647-651.

Rhoades J.D. and Corwin D.L. 1981. Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. Soil Science Society of America Journal 45:255-260.

Rhoades J.D., Raats P.A.C. and Prather R.J. 1976. Effects of liquid-phase electrical conductivity, water content, and surface conductivity on bulk soil electrical conductivity. Soil Science Society of America Journal 40:651-655.

Rhoades J.D., Chanduvi F. and Lesch S.M. 1999. Soil salinity assessment: methods and interpretation of electrical conductivity measurements Irrigation and Drainage Paper, Vol. 57. FAO, Rome.

Rhoades J.D., Waggoner B.L., Shouse P.J. and Alves W.J. 1989. Determining soil salinity from soil and soil-paste electrical conductivities: Sensitivity analysis of models. Soil Science Society of America Journal 53:1368-1374.

Rhoades J.D., Manteghi N.A., Shouse P.J. and Alves W.J. 1989. Estimating soil salinity from saturated soil-paste electrical conductivity. Soil Science Society of America Journal 53:428-433.

Rhoades J.D., Shouse P.J., Alves W.J., Manteghi N.A. and Lesch S.M. 1990. Determining soil salinity from soil electrical conductivity using different models and estimates. Soil Science Society of America Journal 54:46-54.

Richards L.A. (ed) 1954. Diagnosis and improvement of saline and alkali soils. Vol. 60. USDA, Washington, pp. 1-160.

Robbins C.W. and Wiegand C.L. 1990. Field and laboratory measurements. In: Tanji K. K. (ed) Agricultural Salinity Assessment and Management. Vol. 71. American Society of Civil Engineers, New York, pp. 201-219.

Roberts D.W., Dowling T.I. and Walker J. 1997. FLAG: A fuzzy landscape analysis GIS method for dryland salinity assessment. Technical Report No 8/97. CSIRO Land and Water, Canberra, Australia.

Robinson D.A., Lebron I., Lesch S.M. and Shouse P.J. 2004. Minimizing drift in electrical conductivity measurements in high temperature environments using the EM-38. Soil Science Society of America Journal 68:339-345.

Ruhe R.V. 1956. Geomorphic surfaces and the nature of soils. Soil Science 82:441-455.

Saifulin R., Russ S., Fazylova M., Fakhrutdinova N. and Petrenko Y. 1998. Management of water resources in Uzbekistan and way of rising its efficiency. United Satets Agency for International Development, Tashkent.

Saito H. and Goovaerts P. 2000. Geostatistical interpolation of positively skewed and censored data in a dioxin contaminated site. Environmental Science and Technology 34:4228-4235.

Salama R.B., Otto C.J. and Fitzpatrick R.W. 1999. Contributions of groundwater conditions to soil and water salinization. Hydrogeology Journal 7:46-64.

Sarle W.S. 1997. Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets [Online]. Available at ftp://ftp.sas.com /pub/neural/FAQ.html (verified February 15, 2005).

Scull P., Franklin J., Chadwick O.A. and McArthur D. 2003. Predictive soil mapping: a review. Progress in Physical Geography 27:171-197.

Searle R. and Baillie J. 1997. Prediction of landscape salinity hazard using a geographic information system [Online]. Available at http://www.nrm.qld. gov.au/era/projects/sham.html (accessed January 27, 1998; verified November 2, 2004). Department of Natural Resources, State of Queensland

Seelig B.D. and Richardson J.L. 1994. A sodic soil toposequence related to focused water flow. Soil Science Society of America Journal 58:156-163.

Shepherd K.D. and Walsh M.G. 2001. Light reflectance provides rapid assessment of soil quality. Natural resource problems, priorities and policies programme. Working Paper Series 2001-1.

Shirokova Y., Forkutsa I. and Sharafutdinova N. 2000. Use of electrical conductivity instead of soluble salts for soil salinity monitoring in Central Asia. Irrigation and Drainage Systems 14:199-205.

Sudduth K.A., Kitchen N.R., Bollero G.A., Bullock D.G. and W.J. Wiebold. 2003. Comparison of Electromagnetic Induction and Direct Sensing of Soil Electrical Conductivity. Agron J 95:472-482.

Tanji K.K. 1990. Nature and extent of agricultural salinity. In: Tanji K. K. (ed) Agricultural Salinity Assessment and Management, Vol. 71. American Society of Civil Engineers, New York, pp. 1-17.

Thermo. 2004. Orion conductivity theory [Online]. Available at http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_11377.pdf (verified 10.09.2004). Thermo Electron Corporation.

Tomer M.D. and James D.E. 2004. Do Soil Surveys and Terrain Analyses Identify Similar Priority Sites for Conservation? Soil Science Society of America Journal 68:1905-1915.

Triantafilis J., Laslett G.M. and McBratney A.B. 2000. Calibrating an electromagnetic induction instrument to measure salinity in soil under irrigated cotton. Soil Science Society of America Journal 64:1009-1017.

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.

Tursunov L.T. and Abdullaev S. 1987. Pochvenno-fizicheskaya harakteristika nizov'ev Amudaryi FAN, Tashkent.

Valenza A., Grillot J.C. and Dazy J. 2000. Influence of groundwater on the degradation of irrigated soils in a semi-arid region, the inner delta of the Niger River, Mali. Hydrogeology Journal 8:417-429.

van Dijk A., den Haan R., van den Hoven P. and Bos R. 1999. Can WARMAP save the Aral Sea? Land and Water International 93(2).

Walter C., McBratney A.B., Douaoui A. and Minasny B. 2001. Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Australian Journal of Soil Research 39:259-272.

Wang D., Wilson C. and Shannon M.C. 2002. Interpretation of salinity and irrigation effects on soybean canopy reflectance in visible and near-infrared spectrum domain. International Journal of Remote Sensing 23:811-824.

Warrick A.W., Myers D.E. and Nielsen D.R. 1986. Geostatistical methods applied in soil science. In: Klute A. (ed) Methods of soil analysis, Part 1 - Physical and mineralogical methods, Vol. 9. American Society of Agronomy, Madison, WI, pp. 53-82.

Wilson J.P. and Gallant J.C. 2000. Digital terrain analysis. In: Wilson J. P. and Gallant J. C. (eds) Terrain Analysis: Principles and Applications. John Wiley and Sons, New York, pp. 1-27.

Wraith J.M. 2002. Time domain reflectometry. In: Dane J. H. and Topp G. C. (eds) Methods of soil analysis, Vol. 4. SSSA Book Series 5, Madison, WI, pp. 1289-1296.

Wu J., Norvell W.A., Hopkins D.G. and Welch R.M. 2002. Spatial variability of grain cadmium and soil characteristics in a durum wheat field. Soil Science Society of America Journal 66:268-275.

Wysocki D.A., Schoeneberger P.J. and LaGarry H.E. 2000. Geomorphology of soil landscapes. In: Sumner M. E. (ed) Handbook of soil science. CRC Press LLC, Boca Raton.

Yair A. 1990. The role of topography and surface cover upon soil formation along hillslopes in arid climates. Geomorphology 3:287-299.

Yates S.R. and Warrick A.W. 2002. Geostatistics. In: Dane J. H. and Topp G. C. (eds) Methods of soil analysis, Vol. 4. SSSA Book Series 5, Madison, WI, pp. 81-118.

Yemefack M., Rossiter D.G. and Njomgang R. 2005. Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon. Geoderma 125:117-143.

Zevenbergen L.W. and Thorne C.R. 1987. Quantitative analysis of land surface topography. Earth Surface Processes and Landforms 12:47-56.

Zhollibekov B. 1995. Izmenenie pochvennogo pokrova i landshaftov yuzhnogo Priaralya v sviazi s antropogennim vozdeistviem. Bilim.

Zhu A.X. 2000. Mapping soil landscape as spatial continua: The neural network apporach. Water Resources Research 36:663-677.

Zhu A.X., Hudson B., Burt J., Lubich K. and Simonson D. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65:1463-1472.

7 APPENDIX

How to set up the Neural Network Model to estimate soil salinity distribution

Neural network model building and training was done using NeuroSolutions 4.21 from a software company NeuroDimensions (www.nd.com). The version 4.21 was approximately 21 Mb in size. The demo version can be downloaded from the company website, which regularly releases updated versions of the software. Neural network model building consists of several components. The following is the detailed protocol of steps.

Data preparation and format

NeuroSolutions for Excel is a Microsoft Excel add-in that allows performing all procedures within Excel. To start this add-in, the Excel icon should be launched from within NeuroSolutions. All the following described operations are made from Excel, the data set should, therefore, be in that format. Data preparation should follow the standard steps of cleaning, checking, removing missing values, etc., before starting with neural network building.

NeuroSolutions requires variables in columns and cases in rows. The data set can contain more variables and number of cases in rows than required for network building and training, because only tagged variable columns and rows will be used by the network.

Preprocessing and tagging data

Preprocessing data is mainly randomization of rows before training the model.

1. Preprocess data – Randomize rows.
2. Tag data:
   a. Columns as *input* (CLAY_K, SILT_K, DCOLL, LRVI, BAND5, BAND3, LNUA, SL_DEGRE, AS_DEGRE, PROFC, WT, LS, DC, CURV, PLANC, CURV7, TCI, SOLAR, ALT, GWT10, GWS10,) and *desired* (CMV).
   b. Rows as *training, cross validation, testing,* and *production*. *Training* subset of data is used to train the network, *cross validation* subset is used at the same time during training to control the network from overtraining. The *testing* subset is

used to check the trained network on the subset that is completely new to the network. The trained network is applied to a *production* subset for prediction where input variables are known and the desired variable is unknown. (Note: Data set division into subsets for training, cross validation, and testing plays an important role in the network performance.)

Neural network building

3. Building the neural network using the built-in NeuroBuilder. The neural network is built on-the-fly, and depends on data sets with already tagged columns and rows. The NeuroBuilder assigns parameters according to the data set size used for the model, i.e. number of processing elements (PE) in the hidden layer. The number of processing elements of the network depends on the size of input and desired variables.
4. Select from menu NeuroSolutions in Excel – Create/open network – New
5. Follow the steps in NeuroBuilder selecting Multilayer Perceptrons (MLPs) with default parameters (1 hidden layer) and the option of termination of training when Mean Squared Error (MSE) of cross validation subset starts to increase ("Supervised Learning Control" window). To view the MSE graph during training, in the last window "Probe Configuration" tick the training set and CV set in sections named Error and Performance Measures.

Training, testing, sensitivity of the network

6. Train network – Train N times.
7. For the salinity training model, the number of epochs was set to 10000 and number of runs to 10. This will generate a number of sheets with a summary sheet named "Train# Report", which contains the graph that depicts the MSE trend of training and cross-validation subsets during training. The table below the graph presents minimum and final MSE, and the closer the values to zero the better training results.
8. Test network – Test
9. This step tests whether the trained network is loading the best weights determined during training. Please note: Set report type to *regression*. The generated report sheet is named "Test# Report", with the graph of observed and modeled values

(testing data set comes from the subset of rows that were tagged as 'testing' described in Section 2b above) and summary table. The last column in the table presents the correlation coefficient *r* between observed and modeled values.

10. It is necessary to repeat training and testing procedures several times to understand how the network behaves and to achieve better results.
11. Test network – Sensitivity about the mean
12. Sensitivity about the mean determines which of the input variables has more effect on network outputs.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the people whom I would like to thank for their support. I express my deepest gratitude to Prof. Vlek, my first supervisor, for his guidance, advice, and the his trust in me to conduct this study. Sincere thanks to Dr. Soojin Park, my tutor, for his encouragement, constructive analysis and inputs and instruction in modeling approaches. I am very grateful to my tutor Dr. Christopher Martius for his constant support from the start of the PhD, his day-to-day supervision and detailed review of drafts of this thesis. Dr. John Lamers was very helpful throughout the study, especially during the field phase, giving me enough confidence to successfully complete the planned work.

I am grateful to Prof. Skowronek, my second supervisor, for his thorough review of the manuscript. Dr. Rolf Sommer and Prof. Hendrickx carefully read draft chapters of this work and provided much valuable feedback. It was especially interesting to read the comments from all reviewers on my sometimes confusing statements. Thanks to Colin Davenport and Ms. Jend for the quick and careful editing and to Miryam Schneider for her patience in the final preparation of this thesis for publication.

Sincere thanks are also due to my former supervisor at the Tashkent Institute of Irrigation and Melioration (former TIIAME), Dr. Pulatov, for recommending me for the doctoral study at ZEF, to Prof. Ramazon Kuziev, my local supervisor, for the time and assistance he provided during the research period in Uzbekistan, and to Dr. Shirokova, who shared ideas and was always ready to update and advise on the salinity research.

This study could not have been conducted without the cooperation of the staff and team of students from Urgench State University, who withstood the extremely hot weather during the field campaigns. The ZEF Khorezm Project office personnel were excellent in managerial issues and the drivers helped as much as they could. I greatly appreciate the support and kind help from all these people.

Living in Bonn and Urgench has given me the chance of meeting different people of different nationalities and languages, which were never a hindrance to friendship. Thanks to Irina and Nodir for their excellent support, especially during the final writing stages of this thesis, to Darya for the energy she brought into the company. Also, there were good opportunities to learn from one another: Many thanks to Hayot

for sharing his knowledge, to Gerd, Christopher, and Gulbakhor for their advice regarding remote sensing and GIS, and thanks to Asia and Alexander – I know a little bit about forestry now. Exposure to the real German culture would not have been possible without Marc and Anne, thanks for showing me Christmas and carnival; Susanne, thanks for the warm welcome. I always enjoyed the company of my office mates, ZEF colleagues and friends, Bart, Kerstin, Nathaniel, Bori, Tommasso, Khin, Kavita, Alice, and the Uzbekistan gang (Bahtiyor, Akram, Khusniddin, Hurshid, Akmal, Doniyor, etc.).

It is very difficult to mention all who contributed to the success of this work and not mentioning your name is not an act of ingratitude. It suffices to say thank you very much to all of you.

Last but not least, I would like to thank and dedicate this work to my family, to my mother, wife, brothers and sisters, all who supported me through the ups and downs of the research process, and to my children Surayo, Shirin, and Sevara for bearing with my often long absence during my study period away from home.